Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces *

Stefano Lisini †
Dipartimento di Scienze e Tecnologie Avanzate
Università degli Studi del Piemonte Orientale

Abstract

We study existence and approximation of non-negative solutions of partial differential equations of the type

$$\partial_t u - \operatorname{div}(A(\nabla(f(u)) + u\nabla V)) = 0 \qquad \text{in } (0, +\infty) \times \mathbb{R}^n, \tag{0.1}$$

where A is a symmetric matrix-valued function of the spatial variable satisfying a uniform ellipticity condition, $f : [0, +\infty) \to [0, +\infty)$ is a suitable non decreasing function, $V : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Introducing the energy functional $\phi(u) = \int_{\mathbb{R}^n} F(u(x))\, dx + \int_{\mathbb{R}^n} V(x)u(x)\, dx$, where F is a convex function linked to f by $f(u) = uF'(u) - F(u)$, we show that u is the "gradient flow" of ϕ with respect to the 2-Wasserstein distance between probability measures on the space $\mathbb{R}^n$, endowed with the Riemannian distance induced by A^{-1}. In the case of uniform convexity of V, long time asymptotic behaviour and decay rate to the stationary state for solutions of equation (0.1) are studied. A contraction property in Wasserstein distance for solutions of equation (0.1) is also studied in a particular case.

Keywords: nonlinear diffusion equations, parabolic equations, variable coefficient parabolic equations, gradient flows, Wasserstein distance, asymptotic behaviour
Mathematical subject classification (2000): 35K55, 35K15, 35B40

1 Introduction

The aim of this paper is to study existence, approximation and asymptotic behaviour of non-negative solutions of evolution equations of the type

$$\partial_t u(t, x) - \operatorname{div}(A(x)(\nabla(f(u(t, x))) + u(t, x)\nabla V(x))) = 0 \qquad \text{in } (0, +\infty) \times \mathbb{R}^n, \tag{1.1}$$

with initial datum u_0 satisfying

$$u_0 \in L^1(\mathbb{R}^n), \quad u_0 \geq 0, \quad \|u_0\|_{L^1(\mathbb{R}^n)} = 1, \quad \int_{\mathbb{R}^n} |x|^2 u_0(x)\, dx < +\infty. \tag{1.2}$$

*This work has been partially supported by the IMATI (CNR Pavia, Italy)
†e-mail: stefano.lisini@unipv.it

Here $A : \mathbb{R}^n \to \mathbb{M}^{n\times n}$ is a Borel measurable, symmetric matrix valued function satisfying a uniform ellipticity condition,

$$\lambda|\xi|^2 \le \langle A(x)\xi, \xi\rangle \le \Lambda|\xi|^2, \qquad \forall x \in \mathbb{R}^n, \quad \forall \xi \in \mathbb{R}^n, \qquad \lambda > 0, \tag{1.3}$$

$V : \mathbb{R}^n \to (-\infty, +\infty]$ is a function satisfying

$$V \text{ is convex,} \quad \text{lower semi continuous,} \quad \text{bounded from below,} \tag{1.4}$$

$f : [0, +\infty) \to [0, +\infty)$ is a non decreasing function given by

$$f(u) := F'(u)u - F(u),$$

where $F : [0, +\infty) \to \mathbb{R}$ satisfies

$$F \in C^1(0, +\infty), \quad F \text{ is convex,} \quad \text{continuous at } 0, \quad F(0) = 0, \tag{1.5}$$

the condition of superlinear growth at infinity and a condition on behaviour near zero:

$$\lim_{z\to+\infty} \frac{F(z)}{z} = +\infty, \qquad\qquad \lim_{z\downarrow 0} \frac{F(z)}{z^\alpha} > -\infty, \quad \text{for some } \alpha > \frac{n}{n+2}, \tag{1.6}$$

the McCann convexity condition (see [25]):

$$\text{the map} \quad z \mapsto z^n F(z^{-n}) \qquad \text{is convex}^1, \tag{1.7}$$

and a technical assumption of doubling: there exists a constant $C > 0$ such that

$$F(z + w) \le C(1 + F(z) + F(w)) \qquad \forall z, w \in [0, +\infty). \tag{1.8}$$

The existence and approximation of solutions of (1.1) with initial datum (1.2) is obtained interpreting (1.1) as "gradient flow", with respect to a suitable Wasserstein distance, of the energy functional (sum of internal and potential energy functionals)

$$\phi(u) := \int_{\mathbb{R}^n} F(u(x))\, dx + \int_{\mathbb{R}^n} V(x)u(x)\, dx \tag{1.9}$$

defined on the set

$$D(\phi) := \{u \in L^1(\mathbb{R}^n) : u \ge 0, \|u\|_{L^1(\mathbb{R}^n)} = 1, \int_{\mathbb{R}^n} |x|^2 u(x)\, dx < +\infty, \phi(u) < +\infty\}. \tag{1.10}$$

The choice of the domain of ϕ is justified by the fact that the equation (1.1) is parabolic of second order and in divergence form. When the initial datum satisfies (1.2) it is natural to look for solutions $u \ge 0$ with $\|u(t, \cdot)\|_{L^1(\mathbb{R}^n)} = 1$. In other words $u(t, \cdot)$ is a probability density.

[1] The McCann condition include also that the mapping $z \mapsto z^n F(z^{-n})$ is non-increasing. This property follows from $F(0) = 0$ and F convex.

The equation (1.1) describes nonlinear (linear in the particular case $f(u) = u$) diffusion with drift in non-homogeneous and anisotropic material (see for instance [13] for models of diffusion).

The simplest example of such equation is the heat-porous medium equation with variable coefficients (see [31] for a complete up-to-date reference on porous medium equation)

$$\partial_t u - \operatorname{div}(A\nabla u^m) = 0, \tag{1.11}$$

which corresponds to the choice

$$F(z) = \begin{cases} z \log z & \text{for } m = 1 \\ \frac{1}{m-1} z^m & \text{for } m > 1, \end{cases} \qquad V = 0. \tag{1.12}$$

In the case of $A(x) \equiv I$, where I denotes the identity matrix, the Wasserstein gradient flow structure of the equation (1.1) was pointed out for the first time in [22] in the case of linear diffusion, and in [28] in the case of porous medium equation. This point of view was further developed in [5] and [11].

Gradient flow in Riemannian manifolds. In order to clarify the gradient flow interpretation of equation (1.1) we recall the definition of gradient flow in a smooth Riemannian manifold.

Given a Riemannian manifold $\mathscr{M}$, with metric tensor g, we denote by $\langle \cdot, \cdot \rangle_x$ and $|\cdot|_x$ the scalar product and its associated norm on the tangent space $T_x\mathscr{M}$. Given a smooth functional $\phi : \mathscr{M} \to \mathbb{R}$, the gradient of ϕ in $\mathscr{M}$, denoted by $\nabla_g \phi$, is the tangent vector field defined by $\langle \nabla_g \phi(x), \boldsymbol{v}(x) \rangle_x = \operatorname{diff} \phi_{|x} \boldsymbol{v}(x)$, for every vector field $\boldsymbol{v}$.

The gradient flow of ϕ in $\mathscr{M}$ is the dynamical system whose trajectories are solutions of the differential equation

$$\dot{y}(t) = -\nabla_g \phi(y(t)) \qquad \text{in } T_{y(t)}\mathscr{M}. \tag{1.13}$$

Along the gradient flow trajectories the functional ϕ decreases "as fast as possible", according with the metric structure. Indeed, given a smooth curve $y : [0, +\infty) \to \mathscr{M}$, by the chain rule and Cauchy-Schwartz inequality we have

$$\begin{aligned} \frac{d}{dt}\phi(y(t)) &= \operatorname{diff} \phi_{|y(t)} \dot{y}(t) = \langle \nabla_g \phi(y(t)), \dot{y}(t) \rangle_{y(t)} \\ &\geq -|\nabla_g \phi(y(t))|_{y(t)} |\dot{y}(t)|_{y(t)} \geq -\frac{1}{2}|\nabla_g \phi(y(t))|^2_{y(t)} - \frac{1}{2}|\dot{y}(t)|^2_{y(t)}. \end{aligned} \tag{1.14}$$

Since in every real vector space with scalar product we have that

$$\langle \boldsymbol{w}, \boldsymbol{v} \rangle = -\frac{1}{2}|\boldsymbol{w}|^2 - \frac{1}{2}|\boldsymbol{v}|^2 \quad \Longleftrightarrow \quad \boldsymbol{w} = -\boldsymbol{v}, \tag{1.15}$$

equality holds in (1.14) if and only if y solves (1.13). Consequently the gradient flow trajectories of ϕ in $\mathscr{M}$ can be characterized by the energy identity

$$\frac{d}{dt}\phi(y(t)) = -\frac{1}{2}|\nabla_g \phi(y(t))|^2_{y(t)} - \frac{1}{2}|\dot{y}(t)|^2_{y(t)}. \tag{1.16}$$

With suitable notions of modulus of the gradient and modulus of the derivative in metric spaces, the analogous of the identity (1.16) can be taken as definition of gradient flow trajectory in metric spaces. This generalization was performed by De Giorgi's school in [18], [19] with the theory of curves of maximal slope (see [5] where the theory is reformulated).

The formal interpretation of Otto. In [28] Otto proposed a formal interpretation of the space of probability measures on $\mathbb{R}^n$ as a sort of infinite dimensional Riemannian manifold. The "tangent space" at the "point" $\mu = u\mathscr{L}^n$ can be identified with a subspace of $L^2(\mu;\mathbb{R}^n)$ with the standard scalar product

$$\langle \boldsymbol{v}, \boldsymbol{w}\rangle_{L^2(\mu;\mathbb{R}^n)} := \int_{\mathbb{R}^n} \langle \boldsymbol{v}, \boldsymbol{w}\rangle u\, dx, \tag{1.17}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard scalar product on $\mathbb{R}^n$.
Given a smooth curve $t \mapsto u_t$ of probability densities, the associated tangent vector $\boldsymbol{v}_t$ can be characterized as the vector field satisfying the continuity equation

$$\partial_t u + \operatorname{div}(\boldsymbol{v}u) = 0 \tag{1.18}$$

and minimizing

$$\|\boldsymbol{v}_t\|^2_{L^2(\mu_t;\mathbb{R}^n)} = \int_{\mathbb{R}^n} |\boldsymbol{v}_t(x)|^2 u_t(x)\, dx,$$

where $\mu_t = u_t\mathscr{L}^n$. Differentiating the energy functional (1.9) along a smooth curve $t \mapsto u_t$ with tangent vector $\boldsymbol{v}_t$ we formally obtain the chain rule

$$\begin{aligned}\frac{d}{dt}\phi(u_t) &= \int_{\mathbb{R}^n} \big(F'(u_t) + V\big)\partial_t u_t\, dx = -\int_{\mathbb{R}^n} \big(F'(u_t) + V\big) \operatorname{div}(\boldsymbol{v}_t u_t)\, dx \\ &= \int_{\mathbb{R}^n} \langle \nabla F'(u_t) + \nabla V, \boldsymbol{v}_t\rangle u_t\, dx = \left\langle \frac{\nabla f(u_t)}{u_t} + \nabla V, \boldsymbol{v}_t \right\rangle_{L^2(\mu_t;\mathbb{R}^n)},\end{aligned} \tag{1.19}$$

where the last equality is obtained by observing that

$$u\nabla F'(u) = \nabla\big(uF'(u) - F(u)\big) = \nabla f(u). \tag{1.20}$$

Since

$$\begin{aligned}\left\langle \frac{\nabla f(u_t)}{u_t} + \nabla V, \boldsymbol{v}_t \right\rangle_{L^2(\mu_t;\mathbb{R}^n)} &\geq -\left\| \frac{\nabla f(u_t)}{u_t} + \nabla V \right\|_{L^2(\mu_t;\mathbb{R}^n)} \|\boldsymbol{v}_t\|_{L^2(\mu_t;\mathbb{R}^n)} \\ &\geq -\frac{1}{2}\left\| \frac{\nabla f(u_t)}{u_t} + \nabla V \right\|^2_{L^2(\mu_t;\mathbb{R}^n)} - \frac{1}{2}\|\boldsymbol{v}_t\|^2_{L^2(\mu_t;\mathbb{R}^n)},\end{aligned} \tag{1.21}$$

we can say, in formal analogy to (1.16), that u is a trajectory of the gradient flow of ϕ with respect to the formal Otto's structure if

$$\frac{d}{dt}\phi(u_t) = -\frac{1}{2}\left\| \frac{\nabla f(u_t)}{u_t} + \nabla V \right\|^2_{L^2(\mu_t;\mathbb{R}^n)} - \frac{1}{2}\|\boldsymbol{v}_t\|^2_{L^2(\mu_t;\mathbb{R}^n)}. \tag{1.22}$$

Taking into account the chain rule (1.19) and (1.15), we can deduce that u satisfies (1.22) if and only if

$$\boldsymbol{v}_t = -\left(\frac{\nabla f(u_t)}{u_t} + \nabla V\right) \tag{1.23}$$

(the equality in (1.23) is understood in $L^2(\mu_t;\mathbb{R}^n)$). Recalling that u satisfies (1.18), from (1.23) we can deduce that a gradient flow trajectory u solves the partial differential equation (1.1) with $A \equiv I$.

Formally, the natural distance between two probability densities u^0, u^1 induced by the Otto's structure is

$$d^2(u^0,u^1) = \inf\left\{\int_0^1 \|\boldsymbol{v}_t\|^2_{L^2(\mu_t;\mathbb{R}^n)}\,dt : \partial_t u_t + \operatorname{div}(\boldsymbol{v}_t u_t) = 0, u_0 = u^0, u_1 = u^1\right\}. \tag{1.24}$$

As showed in [7] and [5], the distance (1.24) coincides with the already known Wasserstein distance between probability measures in $\mathbb{R}^n$.

In general the equation (1.1) with $A \not\equiv I$ has the structure

$$\partial_t u + \operatorname{div}(\boldsymbol{v}u) = 0, \tag{1.25}$$

$$\boldsymbol{v}_t = -A\left(\frac{\nabla f(u_t)}{u_t} + \nabla V\right). \tag{1.26}$$

In order to obtain an inequality like (1.21) with the vector (1.26) instead of (1.23) it is sufficient to change the scalar product (1.17), on the "tangent space" of the space of probability densities at the "point" $\mu = u\mathscr{L}^n$, with the new scalar product, induced by the matrix $G := A^{-1}$,

$$\langle \boldsymbol{v}, \boldsymbol{w}\rangle_{L^2_G(\mu;\mathbb{R}^n)} := \int_{\mathbb{R}^n} \langle G\boldsymbol{v}, \boldsymbol{w}\rangle u\,dx. \tag{1.27}$$

The chain rule (1.19) can be rewritten as

$$\frac{d}{dt}\phi(u_t) = \left\langle \frac{A\nabla f(u_t)}{u_t} + A\nabla V, \boldsymbol{v}_t\right\rangle_{L^2_G(\mu_t;\mathbb{R}^n)}, \tag{1.28}$$

and, denoting as usual $\|\boldsymbol{v}\|^2_{L^2_G(\mu;\mathbb{R}^n)} := \langle \boldsymbol{v}, \boldsymbol{v}\rangle_{L^2_G(\mu;\mathbb{R}^n)}$, the inequality (1.21) reads

$$\left\langle \frac{A\nabla f(u_t)}{u_t} + A\nabla V, \boldsymbol{v}_t\right\rangle_{L^2_G(\mu_t;\mathbb{R}^n)} \geq -\frac{1}{2}\left\|\frac{A\nabla f(u_t)}{u_t} + A\nabla V\right\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} - \frac{1}{2}\|\boldsymbol{v}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}. \tag{1.29}$$

Analogously to (1.22) we can say that u is a trajectory of the gradient flow of ϕ with respect to this modified formal Otto's structure if

$$\frac{d}{dt}\phi(u_t) = -\frac{1}{2}\left\|\frac{A\nabla f(u_t)}{u_t} + A\nabla V\right\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} - \frac{1}{2}\|\boldsymbol{v}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}. \tag{1.30}$$

Also in this case, from (1.28), (1.30) and (1.15) we obtain that for a trajectory u of the gradient flow of ϕ with respect this new structure, (1.26) holds. Finally from (1.25) we obtain that u solves (1.1).

The natural distance between two probability densities u^0, u^1, associated to this modified Otto's structure, is

$$d^2_G(u^0,u^1) = \inf\left\{\int_0^1 \|\boldsymbol{v}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}\,dt : \partial_t u_t + \operatorname{div}(\boldsymbol{v}_t u_t) = 0, u_0 = u^0, u_1 = u^1\right\}. \tag{1.31}$$

We will show in Corollary 2.5, in the same spirit of Benamou-Brenier and [5], that the distance d_G in (1.31) coincide with a particular Wasserstein distance in $\mathbb{R}^n$ that we will define in the next paragraph and we will denote by W_G.

Wasserstein distance. We first introduce the Borel measurable symmetric metric tensor $G(x) := A^{-1}(x)$ satisfying (thanks to (1.3)) the uniform ellipticity condition:

$$\Lambda^{-1}|\xi|^2 \leq \langle G(x)\xi, \xi\rangle \leq \lambda^{-1}|\xi|^2, \qquad \forall x \in \mathbb{R}^n, \quad \forall \xi \in \mathbb{R}^n. \tag{1.32}$$

The Riemannian distance on $\mathbb{R}^n$ induced by G is then defined by

$$d(x,y) = \inf\left\{\int_0^1 \sqrt{\langle G(\gamma(t))\dot\gamma(t), \dot\gamma(t)\rangle}\, dt : \gamma \in AC([0,1];\mathbb{R}^n),\ \gamma(0) = x,\ \gamma(1) = y\right\}, \tag{1.33}$$

where $AC([0,1];\mathbb{R}^n)$ denotes the set of absolutely continuous curves in $\mathbb{R}^n$ parametrized in the interval $[0,1]$. We denote by $\mathbb{R}^n_G$ the metric space $\mathbb{R}^n$ endowed with the distance d.

The Wasserstein distance between two Borel probability measures μ, ν on $\mathbb{R}^n_G$ with finite second moment, is defined by

$$W_G(\mu,\nu) := \left(\min\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} d^2(x,y)\, d\gamma(x,y) : \gamma \in \Gamma(\mu,\nu)\right\}\right)^{\frac{1}{2}}, \tag{1.34}$$

where $\Gamma(\mu,\nu)$, called the set of admissible plans between μ and ν, is the set of all Borel probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with first marginal μ and second marginal ν, i.e.

$$\Gamma(\mu,\nu) := \{\gamma \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi^1_{\#}\gamma = \mu, \pi^2_{\#}\gamma = \nu\}, \tag{1.35}$$

where $\pi^1(x,y) := x$ and $\pi^2(x,y) := y$ are, respectively, the projections on the first and the second component and $_{\#}$ denotes the push-forward operator on measures (see Section 2.2).

We denote by $\mathscr{P}_2(\mathbb{R}^n_G)$ the complete, separable metric space of Borel probability measures with finite 2-moment, endowed with the distance W_G (for an introduction to Wasserstein distance see, e.g., [32]). We say that a curve $\mu : [0,1] \to \mathscr{P}_2(\mathbb{R}^n_G)$, $t \mapsto \mu_t$ is a constant speed geodesic of the metric space $\mathscr{P}_2(\mathbb{R}^n_G)$ if

$$W_G(\mu_s, \mu_t) = |t-s| W_G(\mu_0, \mu_1), \qquad \forall s,t \in [0,1].$$

In the sequel we often identify the measures which are absolutely continuous with respect to the Lebesgue measure $\mathscr{L}^n$ with their densities.

The approximation scheme for gradient flows. The standard approach to show existence of solutions to equations having a gradient flow structure, is the variational formulation of the time discretization implicit Euler scheme. This method, in fact, was generalized to metric spaces in [17] with the theory of minimizing movements, and further developed in [2] and [5]. In our particular case the method can be stated as follows.

Given a time step $\tau > 0$ and an initial datum $u_0 \in D(\phi)$, define a sequence u_τ^k obtained by solving recursively

$$u_\tau^k \text{ minimizes in } D(\phi) \text{ the functional } u \mapsto \frac{1}{2\tau} W_G^2(u, u_\tau^{k-1}) + \phi(u), \qquad k = 1, 2, 3, \dots \tag{1.36}$$

starting from $u_\tau^0 = u_0$. Defining the piecewise constant function

$$\overline{u}_\tau(t) := u_\tau^k \qquad \text{if} \quad t \in ((k-1)\tau, k\tau], \tag{1.37}$$

a limit point (with respect to the narrow convergence in the space of probability measures) of $\overline{u}_\tau$ for $\tau \to 0$, is a candidate to solve the variational formulation of (1.1). We recall that a sequence of Borel probability measures on $\mathbb{R}^n$, $\mu_k \in \mathscr{P}(\mathbb{R}^n)$ narrowly converges to $\mu \in \mathscr{P}(\mathbb{R}^n)$ if

$$\lim_{k \to +\infty} \int_{\mathbb{R}^n} \varphi(x)\, d\mu_k(x) = \int_{\mathbb{R}^n} \varphi(x)\, d\mu(x) \qquad \forall \varphi \in C_b(\mathbb{R}^n), \tag{1.38}$$

where $C_b(\mathbb{R}^n)$ is the space of continuous bounded functions.

When $A \equiv I$, the convergence of $\overline{u}_\tau$ to a solution of (1.1) is proved in [22] in the case of linear diffusion and [5] in the general case. In recent years, for other class of evolution equations, this kind of problem has attracted a lot of attention. Among many papers dealing with similar problems, particularly significant are [1], [26], [27], [8], [9]. The case of nonlinear diffusion equations with time-dependent coefficients is considered in [29]. A particular case of one dimensional variable coefficient Fokker-Planck equation is considered in [23].

We point out that the proof of [22], in the case of linear diffusion and a smooth potential V, is based on a sort of "first variation" of the functional $u \mapsto \frac{1}{2\tau} W_I^2(u, u_\tau^{k-1}) + \phi(u)$. It would not be too difficult to extend this technique when A is of class C^2. Here, motivated by the desire to work with lower regularity assumptions on A, we adopt a purely metric approach, which works when $G = A^{-1}$ satisfies the following lower semi-continuity property:

$$\text{the map} \quad x \mapsto \langle G(x)\xi, \xi \rangle \qquad \text{is lower semi-continuous} \quad \forall \xi \in \mathbb{R}^n \tag{1.39}$$

(see also Remark 1.2 for other conditions). Our approach is based on the theory of minimizing movements and the theory of curves of maximal slope in metric spaces taking as a reference [5].

Statement of the main result. With the usual identification of the absolutely continuous measures with their densities, the energy functional ϕ can be thought as a functional defined on the metric space $\mathscr{P}_2(\mathbb{R}^n_G)$ in this way, $\phi : \mathscr{P}_2(\mathbb{R}^n_G) \to (-\infty, +\infty]$

$$\phi(\mu) = \begin{cases} \int_{\mathbb{R}^n} F(u(x))\, dx + \int_{\mathbb{R}^n} V(x) u(x)\, dx & \text{if } \mu = u\mathscr{L}^n \in \mathscr{P}_2^r(\mathbb{R}^n_G) \\ +\infty & \text{otherwise,} \end{cases} \tag{1.40}$$

where $\mathscr{P}_2^r(\mathbb{R}^n_G)$ denotes the subset of $\mathscr{P}_2(\mathbb{R}^n_G)$ consisting of absolutely continuous measures with respect to the Lebesgue measure $\mathscr{L}^n$. With this convention, the effective domain of ϕ, i.e. the set $\{\mu \in \mathscr{P}_2(\mathbb{R}^n_G) : \phi(\mu) < +\infty\}$ coincides with (1.10).

In order to motivate the following definition we observe that for a smooth solution u of the equation (1.1), by the same calculation leading to (1.19), we obtain that the rate of decay of the energy ϕ along the solution u is

$$\frac{d}{dt}\phi(\mu_t) = -\left\| \frac{A\nabla f(u_t)}{u_t} + A\nabla V \right\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}. \tag{1.41}$$

It is then natural to define the functional

$$g(\mu) := \begin{cases} \left\| \frac{A\nabla f(u)}{u} + A\nabla V \right\|_{L^2_G(\mu;\mathbb{R}^n)} & \text{if } \mu = u\mathscr{L}^n \in D(g) \\ +\infty & \text{otherwise} \end{cases} \tag{1.42}$$

where the domain of g is defined by

$$D(g) := \{\mu = u\mathscr{L}^n \in D(\phi) : f(u) \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n), \frac{\nabla f(u)}{u} + \nabla V \in L^2(\mu;\mathbb{R}^n)\}. \tag{1.43}$$

We observe that the definition of g makes sense. Indeed the set where $u = 0$ has null μ measure and the internal part of the domain of V, $\Omega := \mathrm{Int}(D(V))$ is not empty when ϕ is proper. The gradient of V, ∇V, is defined $\mathscr{L}^n$-a.e. in Ω (indeed the convexity of V implies that V is locally Lipschitz in Ω and consequently $\mathscr{L}^n$-a.e. differentiable in Ω (see, e.g., [21])). Moreover for every $\mu \in D(\phi)$ the support of μ has to be contained in $\overline{\Omega}$, and then all the integrals on the whole $\mathbb{R}^n$ with respect to the measure μ are in effect integrals on Ω, where ∇V is defined.

We recall that a curve $\mu : I \to \mathscr{P}_2(\mathbb{R}^n_G)$, where I is an interval, is called absolutely continuous if there exists $m \in L^1(I)$ such that $W_G(\mu_t, \mu_s) \leq \int_s^t m(r)\,dr$ for every $s, t \in I$, $s < t$. For any absolutely continuous curve $\mu : I \to \mathscr{P}_2(\mathbb{R}^n_G)$, there exists the metric derivative (see [5]) defined by

$$|\mu'|(t) := \lim_{h\to 0} \frac{W_G(\mu_{t+h}, \mu_t)}{|h|} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{1.44}$$

We denote by $AC^2_{\mathrm{loc}}([0,+\infty); \mathscr{P}_2(\mathbb{R}^n_G))$ the space of locally absolutely continuous curves $\mu : [0,+\infty) \to \mathscr{P}_2(\mathbb{R}^n_G)$, (i.e. μ is absolutely continuous in any bounded subinterval of $[0,+\infty)$) such that $|\mu'|$ belongs to $L^2_{\mathrm{loc}}([0,+\infty))$.

Theorem 1.1 (Existence and convergence)**.** *Assume that A satisfies (1.3) and (1.39), F satisfies (1.5), (1.6), (1.7), (1.8) and V satisfies (1.4).*
Given $\mu_0 = u_0\mathscr{L}^n \in D(\phi)$, a minimizer u^k_τ in (1.36) exists. Defining the probability measures $M^k_\tau := u^k_\tau\mathscr{L}^n$ and the piecewise constant function $\overline{M}_\tau : [0,+\infty) \to \mathscr{P}_2(\mathbb{R}^n_G)$ by

$$\overline{M}_\tau(t) := M^k_\tau \quad \text{if} \quad t \in ((k-1)\tau, k\tau], \tag{1.45}$$

then for every sequence $\tau_n \to 0$ there exists a subsequence (still denoted by τ_n) and a curve $\mu \in AC^2_{\mathrm{loc}}([0,+\infty); \mathscr{P}_2(\mathbb{R}^n_G))$ such that

$$\overline{M}_{\tau_n}(t) \text{ narrowly converges to } \mu_t \qquad \forall t \in [0,+\infty). \tag{1.46}$$

For every $t \in [0,+\infty)$ the measure μ_t is absolutely continuous with respect to the Lebesgue measure, and the function u defined by $\mu_t = u_t \mathscr{L}^n$ satisfies

$$f(u) \in L^1_{\rm loc}((0,+\infty); W^{1,1}_{\rm loc}(\mathbb{R}^n)), \qquad t \mapsto g(\mu_t) \in L^2_{\rm loc}([0,+\infty)), \tag{1.47}$$

the function $t \mapsto \phi(\mu_t)$ is locally absolutely continuous and the energy identity holds

$$\frac{d}{dt}\phi(\mu_t) = -\frac{1}{2}g(\mu_t)^2 - \frac{1}{2}|\mu'|^2(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in [0,+\infty), \tag{1.48}$$

and u is a weak solution of the variable coefficients nonlinear diffusion equation (1.1), in the sense that

$$\frac{d}{dt}\int_{\mathbb{R}^n} \varphi(x)u_t(x)\,dx = \int_{\mathbb{R}^n} \langle A(x)\nabla(f(u_t(x))) + A(x)\nabla V(x)u_t(x), \nabla\varphi(x)\rangle\,dx \qquad \forall \varphi \in C^\infty_c(\mathbb{R}^n), \tag{1.49}$$

where the equality is understood in the sense of distributions in $(0,+\infty)$.
Moreover the following convergence results hold:

$$\lim_{n\to\infty} \phi(\overline{M}_{\tau_n}(t)) = \phi(\mu_t) \qquad \forall t \in [0,+\infty), \tag{1.50}$$

$$\lim_{n\to\infty} g(\overline{M}_{\tau_n}(t)) = g(\mu_t) \qquad \text{in } L^2_{\rm loc}([0,+\infty)), \tag{1.51}$$

$$\lim_{n\to\infty} |M'_{\tau_n}| = |\mu'| \qquad \text{in } L^2_{\rm loc}([0,+\infty)), \tag{1.52}$$

where $|M'_\tau|$ is the piecewise constant function defined by

$$|M'_\tau|(t) := \frac{W_G(M^n_\tau, M^{n-1}_\tau)}{\tau} \qquad \text{if } t \in ((n-1)\tau, n\tau).$$

Strategy of the proof. The general strategy used here is that of make rigorous the formal assertions given in the paragraph of the Otto's interpretation. This strategy was used in the case $A \equiv I$ in [5]. In our case, new difficulties arise because the Wasserstein distance W_G is induced by a non-flat metric, and the functional ϕ is in general not convex along geodesics in $\mathscr{P}_2(\mathbb{R}^n_G)$. On the other hand, under our assumptions on F and V, the metric theory in [5] ensures the well-posedness of the scheme (1.36) and yields the existence of a curve of probability measures $\mu : [0,+\infty) \to \mathscr{P}_2(\mathbb{R}^n_G)$, $t \mapsto \mu_t$, belonging to the space $AC^2_{\rm loc}([0,+\infty); \mathscr{P}_2(\mathbb{R}^n_G))$, which is a so-called curve of maximal slope for ϕ in $\mathscr{P}_2(\mathbb{R}^n_G)$. Roughly speaking, a curve of maximal slope is an absolutely continuous curve satisfying a metric version of the energy identity (1.30), more precisely

$$\frac{d}{dt}\phi(\mu_t) = -\frac{1}{2}|\mu'|^2(t) - \frac{1}{2}|\partial^-\phi|^2_G(\mu_t), \tag{1.53}$$

where $|\partial^-\phi|_G$ is an abstract object, defined in (3.10), which generalizes the modulus of the gradient and $|\mu'|$ is the metric derivative defined in (1.44). In order to recover the energy identity (1.30) from the metric energy identity (1.53) and to show that u solves (1.1), we will take the following three steps.

The first step, performed in Section 2 and interesting by itself, consists in the study of the existence and uniqueness of a vector field $\boldsymbol{v}$ associated to a curve $\mu \in AC^2([0,T];\mathscr{P}_2(\mathbb{R}^n_G))$ such that the continuity equation holds

$$\partial_t \mu_t + \mathrm{div}(\boldsymbol{v}_t \mu_t) = 0 \tag{1.54}$$

and the following equality holds

$$|\mu'|(t) = \|\boldsymbol{v}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} \,. \tag{1.55}$$

The vector field $\boldsymbol{v}$ associated to the curve μ plays the role of the tangent vector field to the curve μ. The content of Section 2 is a non trivial extension of the analogous theory in the space $\mathscr{P}_2(\mathbb{R}^n_I)$ developed in Chapter 8 of [5] .

In the second step we establish the following equality along the curve μ_t

$$|\partial^- \phi|_G(\mu_t) = \left\| \frac{A\nabla f(u_t)}{u_t} + A\nabla V \right\|_{L^2_G(\mu_t;\mathbb{R}^n)} , \tag{1.56}$$

which is obtained in (3.27) of Section 3.

The third step is to prove the chain rule (1.28) in order to write

$$\frac{d}{dt}\phi(\mu_t) = \int_{\mathbb{R}^n} \left\langle \frac{\nabla f(u_t)}{u_t} + \nabla V, \boldsymbol{v}_t \right\rangle u_t\, dx = \left\langle \frac{A\nabla f(u_t)}{u_t} + A\nabla V, \boldsymbol{v}_t \right\rangle_{L^2_G(\mu_t;\mathbb{R}^n)} , \tag{1.57}$$

and this follows without difficulties from the analogous result in the space $\mathscr{P}_2(\mathbb{R}^n_I)$ as shown in Lemma 3.4.

Thanks to (1.55) and (1.56), the energy identity (1.30) follows from (1.53). From the chain rule (1.57), the energy identity (1.30) and (1.54), we obtain that u is a weak solution of (1.1).

Remark 1.2. Theorem 1.1 provides also the existence of a weak solution of equation (1.1) if the matrix A is $\mathscr{L}^n$-a.e. equal to a matrix $\tilde{A}$ whose inverse $\tilde{G} := \tilde{A}^{-1}$ satisfies (1.39) (it is an immediate consequence of the definition of weak solution (1.49)). A condition on A, ensuring that there exists $\tilde{A}$ as before, is that the discontinuity set $S := \{x \in \mathbb{R}^n : A \text{ is not contiuous at } x\}$ is closed and $\mathscr{L}^n(S) = 0$. Indeed we can define

$$\tilde{A}(x) := \begin{cases} A(x) & \text{if } x \in \mathbb{R}^n \setminus S \\ \Lambda I & \text{if } x \in S, \end{cases}$$

which satisfies the required properties.

Asymptotic behaviour. When the potential V is uniformly convex, under the assumption (1.7) on F, the functional ϕ is uniformly convex along geodesics in $\mathscr{P}_2(\mathbb{R}^n_I)$ (see [25], where this notion of convexity was introduced, and [5]). Consequently, there exists a unique minimum point $\mu_\infty = u_\infty \mathscr{L}^n$ of ϕ which turns out to be a stationary state of equation (1.1), also in the case of variable coefficients. In the case $A \equiv I$, the asymptotic behaviour of solutions of equation (1.1) was studied in [12] (see also [5] for the general metric case). As we will show in Remark 1.6, ϕ is, in general, not convex along geodesics of $\mathscr{P}_2(\mathbb{R}^n_G)$ but, thanks to the ellipticity condition (1.3), the equation (1.1) with variable coefficients has the same asymptotic behaviour of the equation with $A \equiv I$.

We summarize the properties of asymptotic behaviour in the following Theorem, proved in Section 3.

Theorem 1.3 (Asymptotic behaviour)**.** *In addition to the assumptions of Theorem 1.1 we assume that the potential V is α-convex for some $\alpha > 0$, i.e.*

$$V(tx + (1-t)y) \le tV(x) + (1-t)V(y) - \frac{1}{2}\alpha t(1-t)|x-y|^2, \qquad \forall x, y \in \mathbb{R}^n, \quad \forall t \in [0,1].$$

Then there exists a unique minimizer $\mu_\infty = u_\infty \mathscr{L}^n$ of the functional ϕ (μ_∞ is called stationary state) and we have

$$g(\mu)^2 \ge 2\lambda\alpha(\phi(\mu) - \phi(\mu_\infty)) \qquad \forall \mu \in D(\phi), \tag{1.58}$$

where λ is the ellipticity constant of the matrix A in (1.3). Moreover, for every $\mu_0 \in D(\phi)$, denoting by μ_t the corresponding solution given by Theorem 1.1, we have

$$\phi(\mu_t) - \phi(\mu_\infty) \le e^{-2\lambda\alpha t}(\phi(\mu_0) - \phi(\mu_\infty)) \qquad \forall t \in (0, +\infty) \tag{1.59}$$

and

$$W_G(\mu_t, \mu_\infty) \le e^{-\lambda\alpha t}\sqrt{\frac{2}{\lambda\alpha}(\phi(\mu_0) - \phi(\mu_\infty))} \qquad \forall t \in (0, +\infty). \tag{1.60}$$

Remark 1.4. The convergence in generalized entropy (1.59) yields, in the most relevant cases, the strong $L^1(\mathbb{R}^n)$ convergence with rates depending, in general, on the nonlinearity of F and the properties of μ_∞. In the case of linear diffusion $F(u) = u \log u$, by means of the Csiszár-Kullback inequality

$$\|u_t - u_\infty\|^2_{L^1(\mathbb{R}^n)} \le 2(\phi(\mu_t) - \phi(\mu_\infty)),$$

we obtain

$$\|u_t - u_\infty\|_{L^1(\mathbb{R}^n)} \le e^{-\lambda\alpha t}\sqrt{2(\phi(\mu_0) - \phi(\mu_\infty))} \qquad \forall t \in (0, +\infty). \tag{1.61}$$

In the case of porous medium type diffusion $F(u) = u^m$, Theorem 31 of [10] (see also [28]) can be applied in the case $m \le 2$ (see the case (d) after the Theorem) and yields

$$\|u_t - u_\infty\|_{L^1(\mathbb{R}^n)} \le C(\phi(\mu_t) - \phi(\mu_\infty))^{1/2}, \tag{1.62}$$

whereas Theorem 32 of [10] can be applied in the case $m \ge 2$ and yields

$$\|u_t - u_\infty\|_{L^1(\mathbb{R}^n)} \le C(\phi(\mu_t) - \phi(\mu_\infty))^{1/m}. \tag{1.63}$$

In the general nonlinear case, under some assumptions relating the nonlinear function F and μ_∞ that must be checked in every particular case, Theorem 25 in [10] yields the inequality

$$\|u_t - u_\infty\|_{L^1(\mathbb{R}^n)} \le U(\phi(\mu_t) - \phi(\mu_\infty)) \tag{1.64}$$

where the function $U : [0, +\infty) \to [0, +\infty)$ is an increasing function continuous at 0, depending on the nonlinearity F and, in a non explicit way, on the properties of u_∞.

Contractivity. In recent years, the issue of the Wasserstein distance contractivity of the solutions of several classes of evolution equations has attracted a lot of attention. In general the α-contractivity of the gradient flow of a functional ϕ in a geodesic metric space is frequently a consequence of the α-convexity of the functional ϕ along geodesics of the metric space (see [11] in the space $\mathscr{P}_2(\mathbb{R}^n_I)$, [33] in $\mathscr{P}_2(\mathscr{M})$ where $\mathscr{M}$ is a Riemannian manifold, and [5] in the abstract metric context). As we will observe in Remark 1.6 below, ϕ is, in general, not convex along geodesics in $\mathscr{P}_2(\mathbb{R}^n_G)$ and the condition characterizing the α-convexity of ϕ in the linear case in $\mathscr{P}_2(\mathbb{R}^n_G)$ is (1.69). This condition implies the α-contractivity of the gradient flow of ϕ with respect to W_G (see [33] and [30]). The condition (1.69), which requires the C^2 regularity of A and V, could be hard to check and difficult to write in terms of the coefficients of A and their partial derivatives. In the following Theorem we present a condition that could be easier to check than (1.69) and implies α-contractivity with respect to the Wasserstein distance W_I.

Theorem 1.5. *Assume that $A = aI$ with $a \in C^1(\mathbb{R}^n)$ and the condition (1.3) holds, and let $F(u) = u \log u$.*
Let $\mu^1_0 = u^1_0 \mathscr{L}^n \in D(\phi)$, $\mu^2_0 = u^2_0 \mathscr{L}^n \in D(\phi)$ and $\mu^1_t = u^1_t \mathscr{L}^n$, $\mu^2_t = u^2_t \mathscr{L}^n$ the solutions given by Theorem 1.1. If there exists $\alpha \in \mathbb{R}$ such that

$$\begin{aligned} &\langle \nabla a(x) - \nabla a(y), x - y\rangle - \langle a(x)\nabla V(x) - a(y)\nabla V(y), x - y\rangle \\ &+ n|\sqrt{a(x)} - \sqrt{a(y)}|^2 \leq -\alpha |x-y|^2 \qquad \forall\, x, y \in \mathbb{R}, \end{aligned} \tag{1.65}$$

then

$$W_I(\mu^1_t, \mu^2_t) \leq e^{-\alpha t} W_I(\mu^1_0, \mu^2_0) \qquad \forall t \in (0, +\infty). \tag{1.66}$$

We observe that, by the equivalence of the Wasserstein distances

$$\sqrt{\Lambda^{-1}} W_I(\mu,\nu) \leq W_G(\mu,\nu) \leq \sqrt{\lambda^{-1}} W_I(\mu,\nu),$$

from (1.66) we obtain immediately that

$$W_G(\mu^1_t, \mu^2_t) \leq \sqrt{\Lambda/\lambda}\, e^{-\alpha t} W_G(\mu^1_0, \mu^2_0) \qquad \forall t \in (0, +\infty). \tag{1.67}$$

Clearly the most interesting case in Theorem 1.5 is $\alpha > 0$, where $W_G(\mu^1_t, \mu^2_t) \to 0$ with exponential decay as $t \to +\infty$.

Remark 1.6 (Convexity of the entropy and Ricci curvature)**.** The space $\mathbb{R}^n_G$ with the metric tensor G is a Riemannian manifold of class C^k when the dependence of A on $x \in \mathbb{R}^n$ is of class C^{k-1}. The intrinsic measure on $\mathbb{R}^n_G$ is the Riemannian volume measure $\gamma := (\det G)^{1/2}\mathscr{L}^n$. Given a Borel probability measure μ on $\mathbb{R}^n$, absolutely continuous with respect to $\mathscr{L}^n$ (or equivalently with respect to γ), we denote by $\mu = \rho\gamma$ and $\mu = u\mathscr{L}^n$ its densities. Clearly we have that $u = (\det G)^{1/2}\rho$. The relative entropy of μ with respect to γ is defined by

$$H_\gamma(\mu) := \int_{\mathbb{R}^n} \rho \log \rho \, d\gamma,$$

and it can be rewritten as

$$H_\gamma(\mu) = \int_{\mathbb{R}^n} u \log u \, dx - \frac{1}{2}\int_{\mathbb{R}^n} \log(\det G) u \, dx.$$

By the result of [30], defining $V_G(x) := \frac{1}{2}\log(\det G(x))$, the functional

$$\phi(\mu) := \int_{\mathbb{R}^n} u \log u \, dx + \int_{\mathbb{R}^n} V \, d\mu = H_\gamma(\mu) + \int_{\mathbb{R}^n} (V_G + V) \, d\mu,$$

is α-convex along geodesics in $\mathscr{P}_2(\mathbb{R}^n_G)$, i.e.

$$\begin{gathered}\phi(\mu_t) \leq (1-t)\phi(\mu_0) + t\phi(\mu_1) - \frac{1}{2}\alpha t(1-t) W_G^2(\mu_0,\mu_1), \qquad \forall \mu_0,\mu_1 \in D(\phi), \\ \forall \mu : [0,1] \to \mathscr{P}_2(\mathbb{R}^n_G) \text{ constant speed geodesic connecting } \mu_0 \text{ to } \mu_1, \quad \forall t \in [0,1],\end{gathered} \tag{1.68}$$

if and only if

$$\mathrm{Ric}_x + \mathrm{Hess}_x(V_G + V) \geq \alpha. \tag{1.69}$$

In (1.69) Ric_x denotes the quadratic form associated to the Ricci tensor in the Riemannian manifold $\mathbb{R}^n_G$ and $\mathrm{Hess}_x V$ is the Hessian quadratic form of V (with respect to the Riemannian structure).

In general the expression of condition (1.69) in coordinates is complicated and hard to check. For instance we explicit condition (1.69) in the case $n = 2$ and $G(x) = g(x)I$. The components of the Ricci tensor are $R_{ij} = R^s_{ijs}$ (the sum is understood when the index are repeated) where $R^l_{ijk} = \Gamma^m_{ik}\Gamma^l_{jm} - \Gamma^m_{jk}\Gamma^l_{im} + \partial_{x_j}\Gamma^l_{ik} - \partial_{x_i}\Gamma^l_{jk}$ are the components of the Riemann curvature tensor and $\Gamma^k_{ij} = \frac{1}{2}(\delta_{ik}\partial_{x_j}\log g + \delta_{jk}\partial_{x_i}\log g - \delta_{ij}\partial_{x_k}\log g)$ are the Christoffel symbols. The components of the Hess V are $H_{ij} = \partial^2_{x_ix_j}V - \Gamma^k_{ij}\partial_{x_k}V$. Computing the Christoffel symbols we obtain $\Gamma^1_{11} = \Gamma^2_{12} = \Gamma^2_{21} = -\Gamma^1_{22} = \frac{1}{2}\partial_{x_1}\log g$ and $\Gamma^2_{22} = \Gamma^1_{12} = \Gamma^1_{21} = -\Gamma^2_{11} = \frac{1}{2}\partial_{x_2}\log g$. Substituting in the components of the Ricci tensor we obtain $R_{11} = R_{22} = -\frac{1}{2}\Delta \log g = -\frac{1}{2}(\partial^2_{x_1x_1}\log g + \partial^2_{x_2x_2}\log g)$ and $R_{12} = R_{21} = 0$. The condition (1.69) can be written as

$$\begin{aligned}&-\frac{1}{2}\Delta \log g I + H(\log g) - \frac{1}{2}\nabla \log g \otimes \nabla \log g + \frac{1}{2}\nabla^\perp \log g \otimes \nabla^\perp \log g \\ &+ H(V) - \frac{1}{2}\nabla \log g \otimes \nabla V + \frac{1}{2}\nabla^\perp \log g \otimes \nabla^\perp V - \alpha g I \geq 0\end{aligned}$$

in the sense of positive definite matrix, where $H(\varphi)$ is the Hessian matrix with respect to the euclidean metric and $\nabla^\perp\varphi = (\partial_{x_2}\varphi, -\partial_{x_1}\varphi)$. This condition is exactly the Bakry-Emery condition for logarithmic Sobolev inequalities (see condition (A1) in [6]), precisely

$$\frac{1}{2}\Delta a I - \frac{1}{2}\langle \nabla a, \nabla V\rangle I - H(a) + aH(V) + \frac{1}{2}\nabla a \otimes \nabla V + \frac{1}{2}\nabla V \otimes \nabla a - \alpha I \geq 0.$$

The equivalence between the two conditions can be proved by using the relations $\partial_{x_i} a = -\frac{1}{g}\partial_{x_i}\log g$, $\partial^2_{x_ix_j} a = \frac{1}{g}(\partial_{x_i}\log g \partial_{x_j}\log g - \partial^2_{x_ix_j}\log g)$ and $\nabla\varphi \otimes \nabla\psi - \nabla^\perp\varphi \otimes \nabla^\perp\psi = \nabla\varphi \otimes \nabla\psi + \nabla\psi \otimes \nabla\varphi - \langle \nabla\varphi, \nabla\psi\rangle I$.

Contents of the rest of the paper

The proofs of Theorems 1.1, 1.3 and 1.5 are given in Section 3, together with some lemmata and the definition of curve of maximal slope. In the following Section 2 we study the continuity equation and its link with Wasserstein distance in $\mathbb{R}^n$ with the non smooth Riemannian metric G.

2 Continuity equation in $\mathbb{R}^n$ with a non smooth Riemannian metric

2.1 Metric derivative of absolutely continuous curves in $\mathbb{R}^n_G$

We recall that $\mathbb{R}^n_G$ denotes the metric space $\mathbb{R}^n$ endowed with the distance d defined in (1.33). Given an interval I, a curve $u : I \to \mathbb{R}^n_G$ is called absolutely continuous if there exists $m \in L^1(I)$ such that $d(u(t), u(s)) \leq \int_s^t m(r)\, dr$ for every $s, t \in I$, $s < t$. We denote by $AC(I; \mathbb{R}^n_G)$ the class of absolutely continuous curves from the interval I on $\mathbb{R}^n_G$. Clearly the condition (1.32) implies that the distance d is equivalent to the euclidean one. Then it follows that $AC(I; \mathbb{R}^n_G) = AC(I; \mathbb{R}^n)$. We recall that for every absolutely continuous curve $u \in AC(I; \mathbb{R}^n_G)$ there exists the metric derivative (see [5]) defined and denoted by

$$|u'|(t) := \lim_{h \to 0} \frac{d(u(t+h), u(t))}{|h|} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.1}$$

The following Proposition shows that, under a suitable lower semi continuity assumption on G, the metric derivative of absolutely continuous curves in $\mathbb{R}^n_G$ coincides with the norm (on the tangent space of the Riemannian manifold $\mathbb{R}^n_G$) of the pointwise derivative.

Proposition 2.1. *We assume that the application*

$$x \mapsto \langle G(x)\xi, \xi\rangle \qquad \text{is lower semi continuous} \qquad \forall \xi \in \mathbb{R}^n. \tag{2.2}$$

If $u \in AC(I; \mathbb{R}^n_G)$, where I is an interval, then $\dot u(t) := \lim_{h \to 0} \frac{u(t+h) - u(t)}{h}$ exists for $\mathscr{L}^1$-a.e. $t \in I$ and

$$|u'|(t) = \sqrt{\langle G(u(t))\dot u(t), \dot u(t)\rangle} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.3}$$

Proof. The existence of $\dot u(t)$ and $|u'|(t)$ for $\mathscr{L}^1$-a.e. $t \in I$ is well known.

In order to prove (2.3) we choose $t \in I$ such that $|u'|(t)$ and $\dot u(t)$ exist. Using the curve $s \in [0, 1] \mapsto u(t + sh)$, which connects $u(t)$ to $u(t + h)$, we can estimate

$$\begin{aligned} d(u(t+h), u(t)) &\leq \int_0^1 \sqrt{\langle G(u(t+sh))\dot u(t+sh)h, \dot u(t+sh)h\rangle}\, ds \\ &= \left| \int_t^{t+h} \sqrt{\langle G(u(\tau))\dot u(\tau), \dot u(\tau)\rangle}\, d\tau \right|. \end{aligned}$$

Then it follows that

$$|u'|(t) \leq \sqrt{\langle G(u(t))\dot u(t), \dot u(t)\rangle}$$

for any Lebesgue point of $\tau \mapsto \sqrt{\langle G(u(\tau))\dot u(\tau), \dot u(\tau)\rangle}$.

Conversely, by the assumption of lower semi continuity (2.2), for every $\varepsilon > 0$ and every $\xi_0 \in S^{n-1} := \{\xi \in \mathbb{R}^n : |\xi| = 1\}$, there exists $\delta_\varepsilon(\xi_0) > 0$ such that

$$\langle G(x)\xi_0, \xi_0\rangle \geq \langle G(u(t))\xi_0, \xi_0\rangle - \varepsilon/2 \qquad \forall x \in B_{\delta_\varepsilon(\xi_0)}(u(t)), \tag{2.4}$$

where $B_\delta(y) := \{z \in \mathbb{R}^n : |z - y| < \delta\}$.
By the bilinearity and the symmetry of the map $(\xi, \bar{\xi}) \mapsto \langle G(y)\xi, \bar{\xi}\rangle$ and the Cauchy-Schwartz inequality we have that

$$\begin{aligned} |\langle G(y)\xi_1, \xi_1\rangle - \langle G(y)\xi_2, \xi_2\rangle| &= |\langle G(y)(\xi_1 + \xi_2), \xi_1 - \xi_2\rangle| \\ &\leq \langle G(y)(\xi_1 + \xi_2), \xi_1 + \xi_2\rangle^{1/2} \langle G(y)(\xi_1 - \xi_2), \xi_1 - \xi_2\rangle^{1/2} \end{aligned}$$

and using the ellipticity condition (1.32) we obtain that

$$|\langle G(y)\xi_1, \xi_1\rangle - \langle G(y)\xi_2, \xi_2\rangle| \leq 2\lambda^{-1}|\xi_1 - \xi_2|, \qquad \forall y \in \mathbb{R}^n, \quad \forall \xi_1, \xi_2 \in S^{n-1}. \tag{2.5}$$

From (2.4) and (2.5) it follows that

$$\langle G(x)\xi, \xi\rangle \geq \langle G(u(t))\xi, \xi\rangle - \varepsilon \qquad \forall x \in B_{\delta_\varepsilon(\xi_0)}(u(t)), \qquad \forall \xi \in B_{\varepsilon\lambda/8}(\xi_0) \cap S^{n-1}. \tag{2.6}$$

Choosing a finite set $\{\xi_0^i \in S^{n-1} : i = 0, 1, \dots N_\varepsilon\}$ such that the family $\{B_{\varepsilon\lambda/8}(\xi_0^i) \cap S^{n-1}\}_{i=0,1,\dots N_\varepsilon}$ covers S^{n-1}, and setting $\delta_\varepsilon := \min\{\delta_\varepsilon(\xi_0^i) : i = 0, 1, \dots N_\varepsilon\}$, from (2.6) we obtain

$$\langle G(x)\xi, \xi\rangle \geq \langle G(u(t))\xi, \xi\rangle - \varepsilon \qquad \forall x \in B_{\delta_\varepsilon}(u(t)), \qquad \forall \xi \in S^{n-1}.$$

Finally the bilinearity of the map $\xi \mapsto \langle G\xi, \xi\rangle$, shows that

$$\langle G(x)\xi, \xi\rangle \geq \langle G(u(t))\xi, \xi\rangle - \varepsilon|\xi|^2 \qquad \forall x \in B_{\delta_\varepsilon}(u(t)), \qquad \forall \xi \in \mathbb{R}^n. \tag{2.7}$$

By the continuity of u there exists $h_\varepsilon > 0$ such that for every h, $|h| < h_\varepsilon$ we have $u(t + h) \in B_{\delta_\varepsilon}(u(t))$. By the ellipticity assumption (1.32), the symmetric matrix $G(u(t)) - \varepsilon I$ is positive definite when $\varepsilon < \Lambda^{-1}$. Since the Riemannian distance induced by G in $B_{\delta_\varepsilon}(u(t))$ coincides with d and $G(u(t)) - \varepsilon I$ is a constant metric tensor in $B_{\delta_\varepsilon}(u(t))$ (the geodesics in this last case are the segments), (2.7) yields

$$d(u(t + h), u(t)) \geq \sqrt{\langle (G(u(t)) - \varepsilon I)(u(t + h) - u(t)), u(t + h) - u(t)\rangle}.$$

Dividing by $|h|$ and passing to the limit for $h \to 0$ we obtain

$$|u'|(t) \geq \sqrt{\langle (G(u(t)) - \varepsilon I)\dot{u}(t), \dot{u}(t)\rangle}.$$

Being ε arbitrary we conclude. □

Remark 2.2. We observe that the property of lower semi continuity (2.2) has been assumed only to prove the inequality

$$|u'|(t) \geq \sqrt{\langle G(u(t))\dot{u}(t), \dot{u}(t)\rangle}.$$

The validity of the equality (2.3) is strictly linked to the possibility of reconstruct the metric G by derivation of the distance d. More precisely the metric G can be reconstructed by derivation of the distance d if

$$\sqrt{\langle G(x)\xi, \xi\rangle} = \lim_{t \to 0^+} \frac{d(x + t\xi, x)}{t} \qquad \forall \xi \in \mathbb{R}^n, \quad \forall x \in \mathbb{R}^n. \tag{2.8}$$

If the coefficients of G are only Borel, there are examples where this property is not satisfied for x in a set of positive Lebesgue measure (see [16]). For instance in $\mathbb{R}^2$, if we choose $G(x) = g(x)I$, $g(x) := \chi_E(x) + 2\chi_{\mathbb{R}^2\setminus E}(x)$ where $E := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \in \mathbb{Q} \text{ or } x_2 \in \mathbb{Q}\}$ and χ_A is the characteristic function of the set A, we have that $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$, and the equality (2.8) is not satisfied for every $x \in \mathbb{R}^2$. Also for the intrinsic distance, defined in [14], which is independent of the equivalence class of G, in general this property is not satisfied in a set of positive Lebesgue measure as showed in [15] (example 5).

2.2 Absolutely continuous curves in Wasserstein spaces and continuity equation

Push forward of measures In this paragraph we recall the definition of the push forward operator on measures, frequently used throughout the paper.
We denote by $\mathscr{P}(X)$ the set of Borel probability measures on the separable metric space X. If Y, Z are separable metric spaces, $\mu \in \mathscr{P}(Y)$ and $F : Y \to Z$ is a Borel map, the *push forward of μ through F*, denoted by $F_\#\mu \in \mathscr{P}(Z)$, is defined as follows:

$$F_\#\mu(B) := \mu(F^{-1}(B)) \qquad \forall B \in \mathscr{B}(Z), \tag{2.9}$$

where $\mathscr{B}(Z)$ is the family of Borel subsets of Z. It is not difficult to check that this definition is equivalent to

$$\int_Z \psi(z)\, d(F_\#\mu)(z) = \int_Y \psi(F(y))\, d\mu(y) \qquad \forall \psi \in C_b(Z). \tag{2.10}$$

More generally the previous formula holds even if $\psi : Z \to \mathbb{R}$ is a bounded Borel function or an $F_\#\mu$-integrable function.
We also recall the following composition rule:

$$(G \circ F)_\#\mu = G_\#(F_\#\mu) \quad \forall \mu \in \mathscr{P}(Y), \quad \forall F : Y \to Z, G : Z \to W \text{ Borel maps}. \tag{2.11}$$

2.2.1 Representation of absolutely continuous curves in Wasserstein spaces

Let $I := [0, T]$. In the case of the Wasserstein metric space $\mathscr{P}_2(\mathbb{R}^n_G)$, the curves of $AC^2(I; \mathscr{P}_2(\mathbb{R}^n_G))$, (i.e. the absolutely continuous curves from I to $\mathscr{P}_2(\mathbb{R}^n_G)$ such that their metric derivative, defined in (1.44), $|\mu'|$ belongs to $L^2(I)$) can be represented as superposition of curves of the same kind in the space $\mathbb{R}^n_G$. The superposition is represented by means of a probability measure on the space of continuous curves in $\mathbb{R}^n_G$, concentrated on the subset $AC^2(I; \mathbb{R}^n_G)$. This point of view was studied in [24] for arbitrary metric spaces. We recall here the main result adapted to our purposes.
We denote by Γ the separable, complete metric space of continuous curves from the compact interval I to $\mathbb{R}^n_G$, metrized by the distance of the uniform convergence $d_\infty(u, v) := \sup_{t \in I} d(u(t), v(t))$.
We denote by e_t the evaluation map, defined as follows:

$$e_t : \Gamma \to \mathbb{R}^n_G, \qquad e_t(u) := u(t), \qquad t \in [0, T].$$

Theorem 2.3. *If $\mu \in AC^2(I; \mathscr{P}_2(\mathbb{R}^n_G))$, then there exists $\eta \in \mathscr{P}(\Gamma)$ such that* *(i)* η *is concentrated on* $AC^2(I;\mathbb{R}^n_G)$,

(ii) $(e_t)_\#\eta = \mu_t \qquad \forall t \in I$,

(iii)

$$|\mu'|^2(t) = \int_\Gamma |u'|^2(t)\, d\eta(u) \qquad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in I.$$

2.2.2 Continuity equation and minimal vector field

Let $I = [0,T]$. Given a narrowly continuous (i.e. sequentially continuous with respect to the narrow convergence) curve $\mu : I \to \mathscr{P}_2(\mathbb{R}^n_G)$, $t \mapsto \mu_t$, we can associate to it the probability measure $\bar{\mu} \in \mathscr{P}(I \times \mathbb{R}^n)$ defined by

$$\int_{I\times\mathbb{R}^n} \psi(t,x)\, d\bar{\mu}(t,x) := \frac{1}{T}\int_0^T \int_{\mathbb{R}^n} \psi(t,x)\, d\mu_t(x)\, dt \tag{2.12}$$

for every bounded Borel function $\psi : I \times \mathbb{R}^n \to \mathbb{R}$. If $\boldsymbol{v} : I \times \mathbb{R}^n \to \mathbb{R}^n$ is a time dependent vector field such that $\boldsymbol{v} \in L^2(\bar{\mu};\mathbb{R}^n)$, we say that $(\mu, \boldsymbol{v})$ satisfies the *continuity equation*

$$\partial_t \mu_t + \mathrm{div}(\boldsymbol{v}_t \mu_t) = 0, \tag{2.13}$$

if the relation

$$\frac{d}{dt}\int_{\mathbb{R}^n} \varphi\, d\mu_t = \int_{\mathbb{R}^n} \langle \nabla\varphi, \boldsymbol{v}_t\rangle\, d\mu_t \qquad \forall \varphi \in C_c^\infty(\mathbb{R}^n) \tag{2.14}$$

holds in the sense of distributions in the interval $(0,T)$.

For notational convenience, we define the following set,

$$EC(\mathbb{R}^n) := \big\{(\mu,\boldsymbol{v}) : \mu : I \to \mathscr{P}_2(\mathbb{R}^n) \text{ is narrowly continuous, } \boldsymbol{v} \in L^2(\bar{\mu};\mathbb{R}^n), \\ (\mu,\boldsymbol{v}) \text{ satisfies the continuity equation}\big\}.$$

Given $\boldsymbol{v} \in L^2(\bar{\mu};\mathbb{R}^n)$, we observe that

$$\|\boldsymbol{v}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} := \int_{\mathbb{R}^n} \langle G(x)\boldsymbol{v}_t(x), \boldsymbol{v}_t(x)\rangle\, d\mu_t(x) < +\infty \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I.$$

Theorem 2.4. *Assume that* G *satisfies (1.32) and (2.2).*
If $\mu \in AC^2(I;\mathscr{P}_2(\mathbb{R}^n_G))$ *then there exists a unique vector field* $\tilde{\boldsymbol{v}} : I \times \mathbb{R}^n \to \mathbb{R}^n$ *such that* $(\mu,\tilde{\boldsymbol{v}}) \in EC(\mathbb{R}^n)$ *and*

$$|\mu'|(t) = \|\tilde{\boldsymbol{v}}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} \qquad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in I. \tag{2.15}$$

Moreover $\tilde{\boldsymbol{v}}$ *satisfies*

$$\|\tilde{\boldsymbol{v}}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} \le \|\boldsymbol{v}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} \qquad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in I, \tag{2.16}$$

for every vector field $\boldsymbol{v}$ *such that* $(\mu,\boldsymbol{v}) \in EC(\mathbb{R}^n)$.

Proof. Let $\eta \in \mathscr{P}(\Gamma)$ be a measure satisfying (i), (ii) and (iii) of Theorem 2.3. We set $\bar\eta := \frac{1}{T}\mathscr{L}^1_{|I} \otimes \eta \in \mathscr{P}(I \times \Gamma)$. Defining the evaluation map $e : I \times \Gamma \to I \times \mathbb{R}^n_G$ by $e(t,u) = (t, e_t(u))$, it is immediate to check that $e_\# \bar\eta = \bar\mu$. The disintegration of $\bar\eta$ with respect to e (see e.g. [20] for the disintegration theorem) yields a Borel family of probability measures $\bar\eta_{t,x}$ on Γ concentrated on $\{u \in \Gamma : e_t(u) = x\}$ such that for every $\varphi \in L^1(\bar\eta)$, $\varphi : I \times \Gamma \to \mathbb{R}$, we have

$$u \mapsto \varphi(t,u) \in L^1(\bar\eta_{t,x}) \text{ for } \bar\mu\text{-a.e. } (t,x) \in I \times \mathbb{R}^n, \tag{2.17}$$

$$(t,x) \mapsto \int_\Gamma \varphi(t,u)\, d\bar\eta_{t,x}(u) \in L^1(\bar\mu), \tag{2.18}$$

$$\int_{I\times\Gamma} \varphi(t,u)\, d\bar\eta(t,u) = \int_{I\times\mathbb{R}^n} \int_\Gamma \varphi(t,u)\, d\bar\eta_{t,x}(u)\, d\bar\mu(t,x) \tag{2.19}$$

and the measures $\bar\eta_{t,x}$ are uniquely determined for $\bar\mu$-a.e. $(t,x) \in I \times \mathbb{R}^n$.
We define the vector field $\tilde{\boldsymbol{v}}$ in the following way

$$\tilde{\boldsymbol{v}}_t(x) := \int_\Gamma \dot u(t)\, d\bar\eta_{t,x}(u) \qquad \text{for } \bar\mu\text{-a.e. } (t,x) \in I \times \mathbb{R}^n, \tag{2.20}$$

and we check that $\tilde{\boldsymbol{v}}$ satisfies the required properties.

First of all $\tilde{\boldsymbol{v}}$ is well defined. Indeed the set

$$S := \{(t,u) \in I \times \Gamma : \dot u(t) \text{ exists, } |u'|(t) \text{ exists, } \sqrt{\langle G(u(t))\dot u(t), \dot u(t)\rangle} = |u'|(t)\}$$

is a Borel set and, by Proposition 2.1 and Fubini's Theorem, being $\bar\eta$ concentrated on $I \times AC^2(I, \mathbb{R}^n_G)$, we obtain that $\bar\eta(I \times \Gamma \setminus S) = 0$. Then the map $(t,u) \in I \times \Gamma \mapsto \dot u(t) \in \mathbb{R}^n$ is well defined for $\bar\eta$-a.e. $(t,u) \in I \times \Gamma$. For $\bar\mu$-a.e. $(t,x) \in I \times \mathbb{R}^n$, we have that $\bar\eta_{t,x}(\{u : (t,u) \in I \times \Gamma \setminus S\}) = 0$, and the map $u \mapsto \dot u(t)$ is well defined for $\bar\eta_{t,x}$-a.e. $u \in \Gamma$.

Now we show that $\tilde{\boldsymbol{v}} \in L^2(\bar\mu; \mathbb{R}^n)$. Using Jensen's inequality, (2.19), the property $\bar\eta(S) = 1$ and (iii) of Theorem 2.3 we obtain

$$\begin{aligned}
\frac{1}{T}\int_0^T \|\tilde{\boldsymbol{v}}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}\, dt &= \int_{I\times\mathbb{R}^n} \langle G(x)\tilde{\boldsymbol{v}}_t(x), \tilde{\boldsymbol{v}}_t(x)\rangle\, d\bar\mu(t,x) \\
&= \int_{I\times\mathbb{R}^n} \langle G(x) \int_\Gamma \dot u(t)\, d\bar\eta_{t,x}(u), \int_\Gamma \dot u(t)\, d\bar\eta_{t,x}(u)\rangle\, d\bar\mu(t,x) \\
&\le \int_{I\times\mathbb{R}^n} \int_\Gamma \langle G(x)\dot u(t), \dot u(t)\rangle\, d\bar\eta_{t,x}(u)\, d\bar\mu(t,x) \qquad (2.21) \\
&= \int_{I\times\Gamma} \langle G(u(t))\dot u(t), \dot u(t)\rangle\, d\bar\eta(t,u) \\
&= \int_{I\times\Gamma} |u'|^2(t)\, d\bar\eta(t,u) = \frac{1}{T}\int_0^T |\mu'|^2(t)\, dt < +\infty.
\end{aligned}$$

Now we prove that (2.14) holds. Taking $\varphi \in C^\infty_c(\mathbb{R}^n)$, the mapping $t \mapsto \int_{\mathbb{R}^n} \varphi\, d\mu_t$ is absolutely continuous. Indeed, for every $s,t \in I$, taking an optimal plan (i.e., a minimizer in (1.34))

$\gamma_{s,t} \in \Gamma(\mu_s, \mu_t)$, we have

$$\begin{aligned}\left|\int_{\mathbb{R}^n} \varphi\, d\mu_t - \int_{\mathbb{R}^n} \varphi\, d\mu_s\right| &\leq \int_{\mathbb{R}^n\times\mathbb{R}^n} |\varphi(y) - \varphi(x)|\, d\gamma_{s,t}(x,y) \\ &\leq C \sup_{x\in\mathbb{R}^n} \|\nabla\varphi(x)\| \int_{\mathbb{R}^n\times\mathbb{R}^n} d(x,y)\, d\gamma_{s,t}(x,y) \\ &\leq C \sup_{x\in\mathbb{R}^n} \|\nabla\varphi(x)\|\, W_G(\mu_s,\mu_t).\end{aligned}$$

Then, for $\mathscr{L}^1$-a.e. $t \in I$,

$$\frac{d}{dt}\int_{\mathbb{R}^n} \varphi\, d\mu_t = \frac{d}{dt}\int_\Gamma \varphi(e_t(u))\, d\eta(u) = \int_\Gamma \langle \nabla\varphi(e_t(u))), \dot{u}(t)\rangle\, d\eta(u)$$

and taking into account the definition of $\tilde{\boldsymbol{v}}$ we obtain

$$\frac{d}{dt}\int_{\mathbb{R}^n} \varphi\, d\mu_t = \int_{\mathbb{R}^n} \langle \nabla\varphi, \tilde{\boldsymbol{v}}_t\rangle\, d\mu_t \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.22}$$

Since this pointwise derivative is also a distributional derivative, we can conclude.

Now we prove (2.15) and (2.16). Using the same argument of the proof of (2.21) we obtain that for every $[a,b] \subset I$

$$\int_a^b \|\tilde{\boldsymbol{v}}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}\, dt \leq \int_a^b |\mu'|^2(t)\, dt \tag{2.23}$$

and it follows that

$$\|\tilde{\boldsymbol{v}}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} \leq |\mu'|(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{2.24}$$

On the other hand for every $\boldsymbol{v} \in L^2(\bar{\mu};\mathbb{R}^n)$ such that $(\mu, \boldsymbol{v})$ satisfies the continuity equation, by Theorem 8.2.1 of [5] (see also Theorem 12 of [3]) there exists $\zeta \in \mathscr{P}(\Gamma)$ such that $(e_t)_\# \zeta = \mu_t$ and ζ is concentrated on the set $\{u \in AC^2(I;\mathbb{R}^n) : u \text{ is an integral solution of } \dot{u}(t) = \boldsymbol{v}_t(u(t))\}$. Taking $s, t \in I$ with $s < t$ and $\gamma_{s,t} := (e_s, e_t)_\# \zeta \in \Gamma(\mu_s, \mu_t)$, we have

$$\begin{aligned}W_G^2(\mu_s,\mu_t) &\leq \int_{\mathbb{R}^n\times\mathbb{R}^n} d^2(x,y)\, d\gamma_{s,t}(x,y) = \int_\Gamma d^2(e_s(u), e_t(u))\, d\zeta(u) \\ &\leq \int_\Gamma (t-s)\int_s^t \langle G(u(r))\dot{u}(r), \dot{u}(r)\rangle\, dr\, d\zeta(u) \\ &= \int_\Gamma (t-s)\int_s^t \langle G(u(r))\boldsymbol{v}_r(u(r)), \boldsymbol{v}_r(u(r))\rangle\, dr\, d\zeta(u) \\ &= (t-s)\int_s^t \int_{\mathbb{R}^n} \langle G(x)\boldsymbol{v}_r(x), \boldsymbol{v}_r(x)\rangle\, d\mu_r(x)\, dr,\end{aligned} \tag{2.25}$$

where the last equality follows by Fubini-Tonelli Theorem. Lebesgue differentiation Theorem implies that

$$|\mu'|(t) \leq \|\boldsymbol{v}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I, \tag{2.26}$$

and (2.16) follows recalling (2.24). Now (2.15) is an obvious consequence of (2.24) and (2.26) applied to $\tilde{\boldsymbol{v}}$.
The uniqueness of $\tilde{\boldsymbol{v}}_t$ is a consequence of the linearity of the continuity equation with respect to the vector field, the uniform convexity of the norm of $L^2_G(\mu_t, \mathbb{R}^n)$ (indeed it is a Hilbert space) and the minimality property (2.16). □

Corollary 2.5 (Benamou-Brenier formula for W_G)**.** *For every μ^0, $\mu^1 \in \mathscr{P}_2(\mathbb{R}^n_G)$ we have*

$$W^2_G(\mu^0, \mu^1) = \inf \left\{ \int_0^1 \|\boldsymbol{v}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} \, dt : (\mu, \boldsymbol{v}) \in EC(\mathbb{R}^n),\ \mu_0 = \mu^0,\ \mu_1 = \mu^1 \right\}. \tag{2.27}$$

Proof. If $(\mu, \boldsymbol{v}) \in EC(\mathbb{R}^n)$ and μ connects μ^0 to μ^1, by (2.25) we obtain

$$W^2_G(\mu^0, \mu^1) \leq \int_0^1 \|\boldsymbol{v}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} \, dt.$$

On the other hand, since $\mathbb{R}^n_G$ is a geodesic space, by Proposition 1 in [24], $\mathscr{P}_2(\mathbb{R}^n_G)$ is a geodesic space too. Taking μ a constant speed geodesic connecting μ^0 to μ^1 and $\tilde{\boldsymbol{v}}$ the associated vector field given by Theorem 2.4 we obtain

$$W^2_G(\mu^0, \mu^1) = \int_0^1 |\mu'|^2(t) \, dt = \int_0^1 \|\tilde{\boldsymbol{v}}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} \, dt.$$

□

3 Proof of the main results

The proof of Theorem 1.1 is based on the general theory of curves of maximal slope in metric spaces as developed in [5]. Here we recall the definition of curve of maximal slope for ϕ together with the notions of slope of ϕ and upper gradient for ϕ.

Definition 3.1 (Local slope)**.** The functional $|\partial\phi|_G : \mathscr{P}_2(\mathbb{R}^n_G) \to [0, +\infty]$ defined by

$$|\partial\phi|_G(\mu) := \limsup_{W_G(\mu,\nu)\to 0} \frac{(\phi(\mu) - \phi(\nu))^+}{W_G(\mu, \nu)} \tag{3.1}$$

is called the *local slope* of the functional ϕ.

Clearly $D(|\partial\phi|_G) \subset D(\phi)$.

In the case $A = I$, under our assumptions on F and V, the local slope of ϕ can be characterized by (see [5] Theorem 10.4.6)

$$|\partial\phi|_I(\mu) = \begin{cases} \left\| \frac{\nabla f(u)}{u} + \nabla V \right\|_{L^2_I(\mu;\mathbb{R}^n)} & \text{if } \mu = u\mathscr{L}^n \in D(|\partial\phi|_I) \\ +\infty & \text{otherwise} \end{cases} \tag{3.2}$$

where the effective domain of the local slope is

$$D(|\partial\phi|_I) = \{\mu = u\mathscr{L}^n \in D(\phi) : f(u) \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n), \frac{\nabla f(u)}{u} + \nabla V \in L^2(\mu;\mathbb{R}^n)\}. \tag{3.3}$$

The derivation of the explicit expression (3.2) of the slope is based on the convexity of ϕ along geodesics on $\mathscr{P}_2(\mathbb{R}^n_I)$ with respect to the distance W_I, (the convexity is ensured by the assumption (1.7) on F and the convexity of V). Since, in general, the functional ϕ is not convex along geodesics on $\mathscr{P}_2(\mathbb{R}^n_G)$ with respect to the distance W_G (see Remark 1.6), the computation of the slope could be more difficult.

A possible generalization of the modulus of the gradient for functionals defined in the metric space $\mathscr{P}_2(\mathbb{R}^n_G)$ is the following notion of upper gradient.

Definition 3.2 (Upper gradient)**.** A Borel function $g : \mathscr{P}_2(\mathbb{R}^n_G) \to [0,+\infty]$ is called a *strong upper gradient* for ϕ if, for every $\mu \in AC^2_{\mathrm{loc}}(I;\mathscr{P}_2(\mathbb{R}^n_G))$ such that $g(\mu)|\mu'| \in L^1_{\mathrm{loc}}(I)$ we have

$$|\phi(\mu_t) - \phi(\mu_s)| \le \int_s^t g(\mu_r)|\mu'|(r)\,dr \qquad \forall s,t \in I, \quad s < t. \tag{3.4}$$

In particular, if $g(\mu)|\mu'| \in L^1_{\mathrm{loc}}(I)$ then $\phi \circ \mu$ is locally absolutely continuous and

$$\left|\frac{d}{dt}\phi(\mu_t)\right| \le g(\mu_t)|\mu'|(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I. \tag{3.5}$$

As showed in [5], $|\partial\phi|_I$ is a strong upper gradient for ϕ in the space $\mathscr{P}_2(\mathbb{R}^n_I)$.

The notions of metric derivative (see (1.44)) and upper gradient allow to define the concept of curve of maximal slope. We refer to the introduction and [5] for the motivation of this definition.

Definition 3.3 (Curve of maximal slope)**.** We say that $\mu : [0,+\infty) \to \mathscr{P}_2(\mathbb{R}^n_G)$ is a *curve of maximal slope* for ϕ, with respect to the strong upper gradient g, if $\mu \in AC^2_{\mathrm{loc}}([0,+\infty);\mathscr{P}_2(\mathbb{R}^n_G))$ and

$$\frac{1}{2}\int_s^t g^2(\mu_r)\,dr + \frac{1}{2}\int_s^t |\mu'|^2(r)\,dr = \phi(\mu_s) - \phi(\mu_t) \qquad \forall s,t \in [0,+\infty), \quad s < t. \tag{3.6}$$

The equality (3.6) is called *energy identity.*

The first fundamental Lemma 3.4 is a sort of chain rule in the space $\mathscr{P}_2(\mathbb{R}^n_G)$. The proof of this Lemma uses the similar result in the "flat" space $\mathbb{R}^n_I$, which holds since the functional ϕ is geodesically convex in the space $\mathscr{P}_2(\mathbb{R}^n_I)$.

Lemma 3.4. *Let $\mu \in AC^2_{\mathrm{loc}}(I;\mathscr{P}_2(\mathbb{R}^n_G))$, $\tilde{\boldsymbol{v}}_t$ be the vector field given by Theorem 2.4 and g be defined in (1.42). If*

$$t \mapsto g(\mu_t)|\mu'|(t) \in L^1_{\mathrm{loc}}(I) \tag{3.7}$$

then $t \mapsto \phi \circ \mu_t$ is locally absolutely continuous and the chain rule holds:

$$\frac{d}{dt}\phi(\mu_t) = \int_{\mathbb{R}^n} \left\langle \frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x), \tilde{\boldsymbol{v}}_t(x) \right\rangle d\mu_t(x) \qquad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in I, \tag{3.8}$$

where $\mu_t = u_t\mathscr{L}^n$.

Proof. By the equivalence of the distances W_G and W_I, (3.2), (1.42) and (3.7) we have that $\int_J |\partial\phi|_I(\mu_t)|\mu'|_I(t)\,dt < +\infty$ for every bounded interval $J \subset I$, where $|\mu'|_I$ denotes the metric derivative of μ with respect to W_I. Since ϕ is convex along geodesics in $\mathscr{P}_2(\mathbb{R}^n_I)$, the Wasserstein chain rule result of Section 10.1.2 of [5] shows that $\phi \circ \mu_t$ is absolutely continuous and

$$\frac{d}{dt}\phi(\mu_t) = \int_{\mathbb{R}^n} \left\langle \frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x), \boldsymbol{v}_t(x) \right\rangle d\mu_t(x) \qquad \text{for a.e. } t \in I, \tag{3.9}$$

where $\boldsymbol{v}_t$ is the vector field associated to the curve μ_t given by Theorem 2.4 in the case of the identity matrix $G = I$.
Moreover, by Corollary 10.3.15 of [5], the vector field $\frac{\nabla f(u_t(x))}{u_t(x)}$ belongs to the closure of $\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^n)\}$ in the topology of $L^2(\mu_t;\mathbb{R}^n)$. Since Theorem 2.4 implies that $\operatorname{div}((\boldsymbol{v}_t - \tilde{\boldsymbol{v}}_t)\mu_t) = 0$ in the sense of distribution, then by density

$$\int_{\mathbb{R}^n} \left\langle \frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x), \boldsymbol{v}_t(x) \right\rangle d\mu_t(x) = \int_{\mathbb{R}^n} \left\langle \frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x), \tilde{\boldsymbol{v}}_t(x) \right\rangle d\mu_t(x),$$

and (3.8) follows from (3.9). □

In order to state the second lemma, we define the relaxed slope $|\partial^-\phi|_G$ by

$$|\partial^-\phi|_G(\mu) := \inf\left\{ \liminf_{n\to+\infty} |\partial\phi|_G(\mu_n) : \mu_n \to \mu \text{ narrowly}, \sup_n W_G(\mu_n,\mu) < +\infty, \sup_n \phi(\mu_n) < +\infty \right\}. \tag{3.10}$$

Lemma 3.5. *The function g defined in (1.42) is a strong upper gradient for ϕ, and the following inequality holds*

$$g(\mu) \le |\partial^-\phi|_G(\mu) \qquad \forall \mu \in D(g). \tag{3.11}$$

In the following proofs the functional ϕ is often written as $\phi(\mu) = \mathscr{F}(\mu) + \mathscr{V}(\mu)$, where the internal energy functional $\mathscr{F} : \mathscr{P}_2(\mathbb{R}^n_G) \to (-\infty,+\infty]$ is defined by

$$\mathscr{F}(\mu) := \begin{cases} \int_{\mathbb{R}^n} F(u(x))\,dx & \text{if } \mu = u\mathscr{L}^n \in \mathscr{P}_2^r(\mathbb{R}^n) \\ +\infty & \text{otherwise,} \end{cases}$$

and the potential energy functional $\mathscr{V} : \mathscr{P}_2(\mathbb{R}^n_G) \to (-\infty,+\infty]$ by

$$\mathscr{V}(\mu) := \int_{\mathbb{R}^n} V(x)\,d\mu(x).$$

We postpone the proof of Lemma 3.5 until after the proof of Theorem 1.1.

Proof of Theorem 1.1. By (1.5) and (1.6), the functional $\mathscr{F}$ is lower semi continuous with respect to the narrow convergence in W_G-bounded sets, i.e. it satisfies

$$\sup_n W_G(\mu_n,\mu) < +\infty, \quad \mu_n \to \mu \text{ narrowly} \quad \Rightarrow \quad \liminf_{n\to+\infty} \mathscr{F}(\mu_n) \ge \mathscr{F}(\mu), \tag{3.12}$$

(see e.g. Chapter 9 of [5] and also general lower semi-continuity results in [4]); by (1.4), the functional $\mathscr{V}$ is lower semi continuous with respect to the narrow convergence, i.e.

$$\mu_n \to \mu \text{ narrowly} \qquad \Rightarrow \qquad \liminf_{n\to+\infty} \mathscr{V}(\mu_n) \geq \mathscr{V}(\mu) \tag{3.13}$$

(see Chapter 5 in [5]); the Wasserstein distance W_G is lower semi continuous with respect to the narrow convergence, i.e.

$$\mu_n \to \mu, \nu_n \to \nu \text{ narrowly} \qquad \Rightarrow \qquad \liminf_{n\to+\infty} W_G(\mu_n, \nu_n) \geq W_G(\mu, \nu) \tag{3.14}$$

(it is a standard basic property that can be found e.g. in [32]). The application of Dunford Pettis theorem yields that W_G-bounded sublevel sets of ϕ are narrowly sequentially compact, i.e.

$$\begin{gathered} \text{if } \{\mu_n\} \text{ is a sequence in } \mathscr{P}_2(\mathbb{R}^n_G) \text{ with} \quad \sup_{m,n} W_G(\mu_n, \mu_m) < +\infty, \quad \sup_n \phi(\mu_n) < +\infty \\ \text{then } \{\mu_n\} \text{ admits a narrowly convergent subsequence.} \end{gathered} \tag{3.15}$$

The functional

$$\Phi(\mu) := \frac{1}{2\tau} W_G^2(\mu, M) + \phi(\mu) \tag{3.16}$$

defined in $\mathscr{P}_2(\mathbb{R}^n_G)$ for fixed $M \in D(\phi)$ and $\tau > 0$, is bounded from below. Indeed by the condition on behaviour near zero in (1.6) we have that there exist c_1, $c_2 \geq 0$ such that the negative part of $F \circ u$ satisfies

$$F^-(u(x)) \leq c_1 u(x)^\alpha + c_2 u(x). \tag{3.17}$$

Then, for $\mu = u\mathscr{L}^n \in D(\mathscr{F})$, we have

$$\begin{aligned} \mathscr{F}(\mu) &= \int_{\mathbb{R}^n} F^+(u(x))\, dx - \int_{\mathbb{R}^n} F^-(u(x))\, dx \geq -c_1 \int_{\mathbb{R}^n} u(x)^\alpha\, dx - c_2 \int_{\mathbb{R}^n} u(x)\, dx \\ &= -c_1 \int_{\mathbb{R}^n} u(x)^\alpha\, dx - c_2. \end{aligned} \tag{3.18}$$

Since it is not restrictive to assume $\alpha < 1$, an application of Hölder's inequality yields

$$\int_{\mathbb{R}^n} u(x)^\alpha\, dx \leq \left(\int_{\mathbb{R}^n} (1 + |x|^2) u(x)\, dx \right)^\alpha \left(\int_{\mathbb{R}^n} (1 + |x|^2)^{-\alpha/(1-\alpha)}\, dx \right)^{1-\alpha} < +\infty \tag{3.19}$$

because of the condition $\alpha > \frac{n}{n+2}$ and $\mu \in \mathscr{P}_2(\mathbb{R}^n)$. The inequality $d^2(x, y) \geq \lambda|x - y|^2 \geq \lambda(\frac{1}{2}|y|^2 - |x|^2)$ yields

$$W_G^2(\mu, M) \geq \frac{\lambda}{2} \int_{\mathbb{R}^n} |x|^2\, d\mu(x) - \lambda \int_{\mathbb{R}^n} |x|^2\, dM(x). \tag{3.20}$$

From (3.18), (3.19), (3.20) and (1.4) we obtain

$$\begin{aligned}\phi(\mu) + \frac{1}{2\tau} W_G^2(\mu, M) \geq -c_2 + \inf V - c_3 \left(1 + \int_{\mathbb{R}^n} |x|^2 \, d\mu(x)\right)^\alpha \\ + \frac{\lambda}{4\tau} \int_{\mathbb{R}^n} |x|^2 \, d\mu(x) - \frac{\lambda}{2\tau} \int_{\mathbb{R}^n} |x|^2 \, dM(x),\end{aligned} \tag{3.21}$$

which implies that

$$\inf_{\mu \in \mathscr{P}_2(\mathbb{R}^n_G)} \Phi(\mu) > -\infty. \tag{3.22}$$

Since the properties (3.12), (3.13), (3.14), (3.15) show that the functional Φ defined in (3.16) for fixed $M \in D(\phi)$ and $\tau > 0$, is narrow lower semi continuous in sublevel sets of Φ, and the sublevel sets of Φ are narrowly sequentially relatively compact, for (3.22) a minimum of Φ exists.

The properties (3.12), (3.13), (3.14), (3.15), (3.22) are exactly the assumptions needed in Chapter 2 of [5] to work with the theory of gradient flows in metric spaces (in our case the metric space is $\mathscr{P}_2(\mathbb{R}^n_G)$ and the weak topology σ is the narrow topology). In particular the narrow compactness of the family of piecewise constant solutions $\overline{M}_\tau$ and consequently the existence of the curve $\mu \in AC^2_{\text{loc}}([0, +\infty); \mathscr{P}_2(\mathbb{R}^n_G))$ satisfying the property (1.46) is a consequence of Proposition 2.2.3 of [5].

We recall now Theorem 2.3.3 of [5]. It states that under the assumptions (3.12), (3.13), (3.14), (3.15), (3.22), if $|\partial^- \phi|_G$ is a strong upper gradient for ϕ and $\mu \in AC^2_{\text{loc}}([0, +\infty); \mathscr{P}_2(\mathbb{R}^n_G))$ is a limit point, in the sense of (1.46), of the approximate piecewise constant $\overline{M}_\tau$, then μ is a curve of maximal slope for ϕ with respect to the upper gradient $|\partial^- \phi|_G$ and the convergence properties (1.50), (1.51), (1.52) hold.

Since Lemma 3.5 yields that $|\partial^- \phi|_G$ is a strong upper gradient for ϕ, Theorem 2.3.3 of [5] can be applied. Consequently the following energy identity holds

$$\frac{1}{2} \int_s^t |\partial^- \phi|_G^2(\mu_r) \, dr + \frac{1}{2} \int_s^t |\mu'|^2(r) \, dr = \phi(\mu_s) - \phi(\mu_t) \qquad \forall s, t \in [0, +\infty), \quad s < t, \tag{3.23}$$

and (1.50), (1.51), (1.52) hold. By (3.11) we obtain that

$$\frac{1}{2} \int_s^t g^2(\mu_r) \, dr + \frac{1}{2} \int_s^t |\mu'|^2(r) \, dr \leq \phi(\mu_s) - \phi(\mu_t) \qquad \forall s, t \in [0, +\infty), \quad s < t \tag{3.24}$$

and $g(\mu) \in L^2_{\text{loc}}([0, +\infty))$. Since by Lemma 3.5 g is a strong upper gradient for ϕ, we have that

$$\phi(\mu_s) - \phi(\mu_t) \leq \int_s^t |\mu'|(r) g(\mu_r) \, dr \leq \frac{1}{2} \int_s^t g^2(\mu_r) \, dr + \frac{1}{2} \int_s^t |\mu'|^2(r) \, dr, \tag{3.25}$$

then (3.6) follows by (3.24) and (3.25).

Now by (3.23), (3.24) and (3.25) we have that

$$\begin{aligned}\phi(\mu_s) - \phi(\mu_t) &= \frac{1}{2}\int_s^t |\partial^- \phi|_G^2(\mu_r)\,dr + \frac{1}{2}\int_s^t |\mu'|^2(r)\,dr \\ &= \frac{1}{2}\int_s^t g^2(\mu_r)\,dr + \frac{1}{2}\int_s^t |\mu'|^2(r)\,dr \\ &= \int_s^t |\mu'|(r) g(\mu_r)\,dr \qquad \forall s,t \in [0,+\infty), \quad s<t.\end{aligned} \tag{3.26}$$

Since $|\mu'|g(\mu) \in L^1_{\text{loc}}(0,+\infty)$, then $\phi(\mu)$ is locally absolutely continuous and (1.48) follow easily from (3.26). It is a direct consequence of (3.26) that

$$|\partial^-\phi|_G^2(\mu_t) = g^2(\mu_t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,+\infty). \tag{3.27}$$

Now we show that $\mu_t = u_t \mathscr{L}^n$ is a weak solution of the equation (1.1).
Let $\tilde{\boldsymbol{v}}$ be the vector field associated to μ given by Theorem 2.4. From (3.26) we have

$$-\frac{d}{dt}\phi(\mu_t) = \frac{1}{2}|\mu'|^2(t) + \frac{1}{2}g^2(\mu_t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,+\infty)$$

which can be rewritten, by (2.15) and the definition of g, as

$$-\frac{d}{dt}\phi(\mu_t) = \frac{1}{2}\,\|\tilde{\boldsymbol{v}}_t\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} + \frac{1}{2}\left\| A\frac{\nabla f(u_t)}{u_t} + A\nabla V\right\|^2_{L^2_G(\mu_t;\mathbb{R}^n)} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,+\infty).$$

On the other hand (3.8) implies that

$$\frac{d}{dt}\phi(\mu_t) = \int_{\mathbb{R}^n} \langle G(x)A(x)\left(\frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x)\right), \tilde{\boldsymbol{v}}_t(x)\rangle\, d\mu_t(x) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,+\infty).$$

Then equality holds in Cauchy-Schwartz inequality in the Hilbert space $L^2_G(\mu_t;\mathbb{R}^n)$, and this implies that

$$\tilde{\boldsymbol{v}}_t(x) = -A(x)\frac{\nabla f(u_t(x))}{u_t(x)} - A(x)\nabla V(x) \qquad \text{for } \mu_t\text{-a.e. } x \in \mathbb{R}^n. \tag{3.28}$$

Now (1.49) follows from (3.28) since $(\mu, \tilde{\boldsymbol{v}})$ satisfies the continuity equation in the sense of (2.14). □

Proof of Lemma 3.5. First of all we prove that g is an upper gradient.
Let $\mu \in AC^2_{\text{loc}}(I;\mathscr{P}_2(\mathbb{R}^n_G))$ such that $t \mapsto g(\mu_t)|\mu'|(t) \in L^1_{\text{loc}}(I)$. By Lemma 3.4, Cauchy-Schwartz inequality and Theorem 2.4, we have

$$\begin{aligned}\left|\frac{d}{dt}\phi(\mu_t)\right| &= \left|\int_{\mathbb{R}^n}\left\langle \frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x), \tilde{\boldsymbol{v}}_t(x)\right\rangle d\mu_t(x)\right| \\ &= \left|\int_{\mathbb{R}^n}\langle G(x)A(x)\left(\frac{\nabla f(u_t(x))}{u_t(x)} + \nabla V(x)\right), \tilde{\boldsymbol{v}}_t(x)\rangle\, d\mu_t(x)\right| \\ &\le \left\| A\frac{\nabla f(u_t)}{u_t} + A\nabla V\right\|_{L^2_G(\mu_t;\mathbb{R}^n)} \|\tilde{\boldsymbol{v}}_t\|_{L^2_G(\mu_t;\mathbb{R}^n)} = g(\mu_t)|\mu'|(t)\end{aligned}$$

which states that g is an upper gradient.

Now we prove that

$$g(\mu) \le |\partial\phi|_G(\mu) \qquad \forall \mu \in D(g). \tag{3.29}$$

Let $\boldsymbol{\xi} \in C_c^\infty(\Omega;\mathbb{R}^n)$ be a regular vector field, where $\Omega := \mathrm{Int}(D(V))$, and let $X(t,x)$ be the flow associated to $\boldsymbol{\xi}$: i.e. the unique solution of the Cauchy problem

$$\dot X(t,x) = \boldsymbol{\xi}(X(t,x)), \quad X(0,x) = x, \quad t \in \mathbb{R}. \tag{3.30}$$

Fixing $\mu \in D(g)$ we consider the curve $\mu_t := X(t,\cdot)_\#\mu$ and we observe that for t sufficiently small $\mu_t \in D(\phi)$. Denoting, as usual, by $\mu = u\mathscr{L}^n$ and $\mu_t = u_t\mathscr{L}^n$, we observe that $u_t \rightharpoonup u$ weakly in $L^1(\Omega)$ as $t \to 0$ (indeed μ_t narrowly converges to μ and the family $\{u_t\}_{t\in(-\varepsilon,\varepsilon)}$ is equiintegrable). By the L^1 weak convergence of u_t to u, the regularity of $\boldsymbol{\xi}$ and the ellipticity condition (1.3), the mapping $t \mapsto \|\boldsymbol{\xi}\|^2_{L^2_G(\mu_t;\mathbb{R}^n)}$ is continuous at 0. Consequently we have

$$\lim_{t\downarrow 0} \frac{1}{t}\int_0^t \|\boldsymbol{\xi}\|^2_{L^2_G(\mu_s;\mathbb{R}^n)}\, ds = \|\boldsymbol{\xi}\|^2_{L^2_G(\mu;\mathbb{R}^n)}\,. \tag{3.31}$$

By the definition of W_G, using the admissible plan $(\boldsymbol{i}(\cdot), X(t,\cdot))_\#\mu$, where $\boldsymbol{i}$ denotes the identity map on $\mathbb{R}^n$, the definition of Riemannian distance d, and (3.31) we have

$$\begin{aligned} W_G^2(\mu,\mu_t) &\le \int_{\mathbb{R}^n} d^2(x, X(t,x))\, d\mu(x) \le \int_{\mathbb{R}^n} t\int_0^t \langle G(X(s,x))\dot X(s,x), \dot X(s,x)\rangle\, ds\, d\mu(x) \\ &= t\int_{\mathbb{R}^n}\int_0^t \langle G(X(s,x))\boldsymbol{\xi}(X(s,x)), \boldsymbol{\xi}(X(s,x))\rangle\, ds\, d\mu(x) \\ &= t\int_0^t \|\boldsymbol{\xi}\|^2_{L^2_G(\mu_s;\mathbb{R}^n)}\, ds = t^2(\|\boldsymbol{\xi}\|^2_{L^2_G(\mu;\mathbb{R}^n)} + o(t)). \end{aligned} \tag{3.32}$$

The definition of $|\partial\phi|_G$ and (3.32) yield

$$|\partial\phi|_G(\mu) \ge \lim_{t\downarrow 0} \frac{\phi(\mu_t)-\phi(\mu)}{W_G(\mu_t,\mu)} \ge \frac{1}{\|\boldsymbol{\xi}\|_{L^2_G(\mu;\mathbb{R}^n)}} \lim_{t\downarrow 0}\frac{\phi(\mu_t)-\phi(\mu)}{t}. \tag{3.33}$$

In order to compute $\lim_{t\downarrow 0}\frac{\phi(\mu_t)-\phi(\mu)}{t}$ we consider the decomposition $\phi(\mu) = \mathscr{F}(\mu) + \mathscr{V}(\mu)$. By the regularity of the flow $X(t,\cdot)$, changing variable in the integral, for t sufficiently small we obtain

$$\begin{aligned} \mathscr{F}(\mu_t) - \mathscr{F}(\mu) &= \int_{\mathbb{R}^n} F\left(\frac{u(x)}{\det(\nabla(X(t,x)))}\right)\det(\nabla(X(t,x))) - F(u(x))\, dx \\ &= \int_{\mathbb{R}^n} \Psi(u(x), \det(\nabla(X(t,x))) - \Psi(u(x),1)\, dx, \end{aligned}$$

where $\Psi(z,s) := sF(\frac{z}{s})$ is defined for $z \in [0,+\infty)$ and $s \in (0,+\infty)$. An elementary computation shows that $\frac{\partial}{\partial s}\Psi(z,s) = -f(\frac{z}{s})$. Using the monotonicity of f, the doubling condition (1.8) and the inequality

$$F(w) \le wF'(w) \le F(2w) - F(w),$$

for $s > 1/2$ we have

$$0 \le f\Big(\frac{z}{s}\Big) \le f(2z) = F'(2z)2z - F(2z) \le F(4z) - 2F(2z) \le F(4z) - 4F(z) \\ \le C(1 + 4F(z)) + 4F^-(z). \tag{3.34}$$

Since $\det(\nabla(X(0,x))) = 1$ and $\frac{d}{dt}\det(\nabla(X(t,x))) = \operatorname{div}\boldsymbol{\xi}(X(t,x))\det(\nabla(X(t,x)))$, and $\boldsymbol{\xi}$ has compact support, by (3.34) we can pass to the limit

$$\begin{aligned}
\lim_{t\downarrow 0}\frac{\mathscr{F}(\mu_t) - \mathscr{F}(\mu)}{t} &= \int_{\mathbb{R}^n} \frac{d}{dt}\Psi(u(x), \det(\nabla(X(t,x)))_{|t=0}\,dx \\
&= -\int_{\mathbb{R}^n} f(u(x))\frac{d}{dt}\det(\nabla(X(t,x)))_{|t=0}\,dx \\
&= -\int_{\mathbb{R}^n} f(u(x))\operatorname{div}\boldsymbol{\xi}(x)\,dx \\
&= \int_{\mathbb{R}^n} \langle \nabla f(u(x)), \boldsymbol{\xi}(x)\rangle\,dx.
\end{aligned} \tag{3.35}$$

Moreover, since V is locally Lipschitz in Ω and $\boldsymbol{\xi}$ has compact support in Ω, we have

$$\begin{aligned}
\lim_{t\downarrow 0}\frac{\mathscr{V}(\mu_t) - \mathscr{V}(\mu)}{t} &= \lim_{t\downarrow 0}\frac{1}{t}\int_{\mathbb{R}^n}(V(X(t,x)) - V(x))u(x)\,dx \\
&= \int_{\mathbb{R}^n}\langle \nabla V(x), \boldsymbol{\xi}(x)\rangle u(x)\,dx.
\end{aligned} \tag{3.36}$$

Finally (3.35), (3.36) and (3.33) yield

$$\begin{aligned}
|\partial\phi|_G(\mu) &\ge \frac{1}{\|\boldsymbol{\xi}\|_{L^2_G(\mu;\mathbb{R}^n)}}\int_{\mathbb{R}^n}\langle \nabla f(u(x)) + \nabla V(x)u(x), \boldsymbol{\xi}(x)\rangle\,dx \\
&= \frac{1}{\|\boldsymbol{\xi}\|_{L^2_G(\mu;\mathbb{R}^n)}}\int_{\mathbb{R}^n}\langle G(x)(A(x)\frac{\nabla f(u(x))}{u(x)} + A(x)\nabla V(x)), \boldsymbol{\xi}(x)\rangle\,d\mu(x).
\end{aligned}$$

By the density of $C^\infty_c(\Omega;\mathbb{R}^n)$ in $L^2_G(\mu;\mathbb{R}^n)$ (recall that the support of μ is contained in $\overline{\Omega}$) and duality formula for the norm in $L^2_G(\mu;\mathbb{R}^n)$ we obtain (3.29).

Now we prove that g is lower semi continuous with respect to the narrow convergence in bounded sublevel sets of ϕ. Precisely we prove that:

$$\mu_n \to \mu \quad \text{narrowly}, \qquad \sup_{m,n} W_G(\mu_m,\mu_n) < +\infty, \quad \sup_n \phi(\mu_n) < +\infty \quad \Longrightarrow \\ \liminf_{n\to\infty} g(\mu_n) \ge g(\mu). \tag{3.37}$$

Setting

$$l := \liminf_n g(\mu_n),$$

it is not restrictive to assume (if necessary extracting a subsequence) that

$$\sup_n g(\mu_n) < +\infty, \qquad \lim_n g(\mu_n) = l < +\infty, \tag{3.38}$$

since the case $l = +\infty$ is obvious.
Denoting by $\boldsymbol{w}_n u_n := A^{\frac{1}{2}}(\nabla f(u_n) + \nabla V u_n)$ we can write $g(\mu_n) = \int_{\mathbb{R}^n} |\boldsymbol{w}_n|^2 \, d\mu_n$. By (3.38) and the narrow convergence of μ_n to μ, Theorem 5.4.4 of [5] states that there exists a vector field $\boldsymbol{w} \in L^2(\mu; \mathbb{R}^n)$ such that, up to extracting a subsequence,

$$\lim_n \int_{\mathbb{R}^n} \langle \varphi, \boldsymbol{w}_n \rangle \, d\mu_n = \int_{\mathbb{R}^n} \langle \varphi, \boldsymbol{w} \rangle \, d\mu \qquad \forall \varphi \in C_c^0(\mathbb{R}^n; \mathbb{R}^n), \tag{3.39}$$

and

$$\liminf_n \int_{\mathbb{R}^n} |\boldsymbol{w}_n|^2 \, d\mu_n \geq \int_{\mathbb{R}^n} |\boldsymbol{w}|^2 \, d\mu. \tag{3.40}$$

We must only prove that

$$A^{\frac{1}{2}}(x)(\nabla f(u(x)) + \nabla V(x) u(x)) = \boldsymbol{w}(x) u(x) \qquad \text{for } \mathscr{L}^n\text{-a.e. } x \in \mathbb{R}^n. \tag{3.41}$$

Since $\sup_n \mathscr{F}(\mu_n) < +\infty$, $\sup_n W_G(\mu_n, \mu) < +\infty$ and F is superlinear, by weak compactness in $L^1(\Omega)$ we have that $u_n \rightharpoonup u$ weakly in $L^1(\Omega)$. Since ∇V is locally bounded in Ω we have

$$\lim_n \int_{\mathbb{R}^n} \langle \varphi, A^{\frac{1}{2}} \nabla V(x) \rangle u_n(x) \, dx = \int_{\mathbb{R}^n} \langle \varphi, A^{\frac{1}{2}} \nabla V(x) \rangle u(x) \, dx \qquad \forall \varphi \in C_c^0(\Omega; \mathbb{R}^n). \tag{3.42}$$

Now we show that the sequence $f(u_n)$ is bounded in $BV_{\text{loc}}(\mathbb{R}^n)$, i.e., for every $\tilde{\Omega} \subset\subset \mathbb{R}^n$ the sequence $f(u_n)_{|\tilde{\Omega}}$ is bounded in $BV(\tilde{\Omega})$.
The convexity of F yields $f(u) \leq F(2u) - 2F(u)$, and using the doubling condition (1.8), we obtain that $f(u_n)$ is bounded in $L^1_{\text{loc}}(\mathbb{R}^n)$. Moreover

$$\int_{\mathbb{R}^n} |\nabla f(u_n(x))| \, dx = \int_{\mathbb{R}^n} \left| \frac{\nabla f(u_n(x))}{u_n(x)} \right| u_n(x) \, dx \leq \left(\int_{\mathbb{R}^n} \left| \frac{\nabla f(u_n(x))}{u_n(x)} \right|^2 u_n(x) \, dx \right)^{\frac{1}{2}} \tag{3.43}$$

which is bounded by (3.38) and (1.3).
By compactness in $BV_{\text{loc}}(\mathbb{R}^n)$ (see Theorem 3.23 of [4]) there exists a function $L \in BV_{\text{loc}}(\mathbb{R}^n)$ such that, up to considering a subsequence, $f(u_n) \to L$ in $L^1_{\text{loc}}(\mathbb{R}^n)$. Again, up to considering a subsequence, we can suppose that $f(u_n(x)) \to L(x)$ for $\mathscr{L}^n$-a.e. $x \in \mathbb{R}^n$.
The monotonicity of the mapping $z \mapsto f(z)$ and a truncation argument yield that $L(x) = f(u(x))$ for $\mathscr{L}^n$-a.e. $x \in \mathbb{R}^n$. Then for every $\varphi \in C_c^0(\mathbb{R}^n; \mathbb{R}^n)$ we have (see for instance Proposition 3.13 of [4])

$$\lim_n \int_{\mathbb{R}^n} \langle \varphi, \nabla f(u_n) \rangle \, dx = \int_{\mathbb{R}^n} \langle \varphi, \nabla f(u) \rangle \, dx. \tag{3.44}$$

Since for every measurable subset $B \subset \mathbb{R}^n$ we have

$$\begin{aligned} \int_B |\nabla f(u_n(x))| \, dx &= \int_B \left| \frac{\nabla f(u_n(x))}{\sqrt{u_n(x)}} \right| \sqrt{u_n(x)} \, dx \\ &\leq \left(\int_B \frac{|\nabla f(u_n(x))|^2}{u_n(x)} \, dx \right)^{\frac{1}{2}} \left(\int_B u_n(x) \, dx \right)^{\frac{1}{2}} \end{aligned}$$

the equintegrability of $\{u_n\}$ (recall that u_n is L^1 weakly convergent) and the bound (3.38) imply the equintegrability of $\{\nabla f(u_n)\}$. Then the convergence of $\nabla f(u_n)$ is also weak in L^1, and the symmetry of A implies that

$$\lim_n \int_{\mathbb{R}^n} \langle \varphi, A^{\frac{1}{2}} \nabla f(u_n) \rangle \, dx = \int_{\mathbb{R}^n} \langle \varphi, A^{\frac{1}{2}} \nabla f(u) \rangle \, dx \qquad \forall \varphi \in C_c^0(\mathbb{R}^n; \mathbb{R}^n). \tag{3.45}$$

Taking into account (3.39), (3.42) and (3.45) we deduce (3.41).

The inequality (3.11) is a consequence of the definition (3.10) of $|\partial^- \phi|_G$, the property (3.37) and the inequality (3.29). □

Proof of Theorem 1.3. The proof follows by the analogous result for the case $A = I$ Theorem 2.1 of [12] (or Lemma 2.4.13 of [5]) and the ellipticity condition (1.3) on A. The existence and uniqueness of the stationary state μ_∞ is exactly the same for the case $A = I$ and follows by the geodesically strict convexity of ϕ on $\mathscr{P}_2(\mathbb{R}^n_I)$ (see e.g. [25] and [12]).
By the ellipticity of A, (3.2) and the definition of g, we have

$$g(\mu)^2 = \int_{\mathbb{R}^n} \langle A \left(\frac{\nabla f(u)}{u} + \nabla V \right), \frac{\nabla f(u)}{u} + \nabla V \rangle \, d\mu \geq \lambda \int_{\mathbb{R}^n} \left| \frac{\nabla f(u)}{u} + \nabla V \right|^2 d\mu = \lambda |\partial \phi|_I^2(\mu). \tag{3.46}$$

Since by Theorem 2.1 of [12] we have

$$|\partial \phi|_I^2(\mu) \geq 2\alpha(\phi(\mu) - \phi(\mu_\infty)) \qquad \forall \mu \in D(\phi),$$

(1.58) follows from (3.46).
Let $t \mapsto \mu_t$ be the curve given from Theorem 1.1. It is a consequence of (3.26) that $t \mapsto \phi(\mu_t)$ is locally absolutely continuous and

$$\frac{d}{dt} \phi(\mu_t) = -g^2(\mu_t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, +\infty). \tag{3.47}$$

Now (1.59) is a consequence of (1.58), (3.47) and Gronwall's Lemma.
Again for Theorem 2.1 of [12], we have that

$$W_I(\mu, \mu_\infty) \leq \sqrt{\frac{2}{\alpha}(\phi(\mu) - \phi(\mu_\infty))} \qquad \forall \mu \in D(\phi)$$

and (1.60) follows by (1.59) and the inequality

$$W_G(\mu, \nu) \leq \sqrt{\lambda^{-1}} W_I(\mu, \nu).$$

□

Proof of Theorem 1.5. Applying Lemma 4.3.4 of [5] we obtain that the map $t \mapsto W_I(\mu_t^1, \mu_t^2)$ is absolutely continuous and

$$\frac{d}{dt} W_I^2(\mu_t^1, \mu_t^2) \leq \frac{\partial}{\partial s} W_I^2(\mu_s^1, \mu_t^2)_{|s=t} + \frac{\partial}{\partial s} W_I^2(\mu_t^1, \mu_s^2)_{|s=t} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, +\infty). \tag{3.48}$$

Moreover we recall that (see Proposition 8.4.7 and Remark 8.4.8 of [5]) for any absolutely continuous curve μ_t in $\mathscr{P}_2(\mathbb{R}^n)$ and any measure $\sigma \in \mathscr{P}_2(\mathbb{R}^n)$, and for a Borel vector field $\boldsymbol{v}_t$ such that $\int_0^T \int_{\mathbb{R}^n} |\boldsymbol{v}_t(x)|^2 \, d\mu_t(x) \, dt < +\infty$ for every $T > 0$ and $(\mu_t, \boldsymbol{v}_t)$ satisfies the continuity equation, we have

$$\frac{d}{dt} W_I^2(\mu_t, \sigma) = 2 \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x - y, \boldsymbol{v}_t(x) \rangle \, d\gamma(x,y) \qquad \forall \gamma \in \Gamma_o(\mu_t, \sigma) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, +\infty), \tag{3.49}$$

where $\Gamma_o(\mu_t, \sigma)$ denotes the subset of $\Gamma(\mu_t, \sigma)$ consisting of minimizers in (1.34) in the case $A = I$.

We denote by $r : \mathbb{R}^n \to \mathbb{R}^n$ the optimal transport map for the euclidean quadratic cost $|x-y|^2$ from μ_t^1 to μ_t^2 and we recall that $r_\# \mu_t^1 = \mu_t^2$, r is μ_t^1-a.e. differentiable and $\det(\nabla r(x)) > 0$ for μ_t^1-a.e. $x \in \mathbb{R}^n$ and there exists a μ_t^1 negligible set N such that r is strictly monotone on $\mathbb{R}^n \setminus N$. We also recall that r is μ_t^1-essentially injective and the optimal transport map from μ_t^2 to μ_t^1 is the inverse function r^{-1} (for all these properties see Section 6.2 of [5]).

Since u^1 and u^2 are weak solutions given by Theorem 1.1 of the equation (1.1) with $f(u) = u$, by (1.47) $u_t^i \in W^{1,1}(\mathbb{R}^n)$, $i = 1, 2$ and

$$\int_0^T \int_{\mathbb{R}^n} \left(\left| \frac{\nabla u_t^i(x)}{u_t^i(x)} \right|^2 + |\nabla V(x)|^2 \right) u_t^i(x) \, dx \, dt < +\infty$$

for every $T > 0$, $i = 1, 2$. Then we can apply (3.49) with $\gamma = (\boldsymbol{i}, r)_\# \mu_t^1$ and $\boldsymbol{v}_t(x) = -\frac{a(x)\nabla u_t^1(x)}{u_t^1(x)} - a(x)\nabla V(x)$ obtaining that

$$\frac{\partial}{\partial s} W_I^2(\mu_s^1, \mu_t^2)_{|s=t} = -2 \int_{\mathbb{R}^n} \langle x - r(x), a(x) \frac{\nabla u_t^1(x)}{u_t^1(x)} + a(x)\nabla V(x) \rangle u_t^1(x) \, dx, \tag{3.50}$$

and, similarly, with $\gamma = (\boldsymbol{i}, r^{-1})_\# \mu_t^2$ and $\boldsymbol{v}_t(x) = -\frac{a(x)\nabla u_t^2(x)}{u_t^2(x)} - a(x)\nabla V(x)$ obtaining that

$$\frac{\partial}{\partial s} W_I^2(\mu_t^1, \mu_s^2)_{|s=t} = -2 \int_{\mathbb{R}^n} \langle y - r^{-1}(y), a(y) \frac{\nabla u_t^2(y)}{u_t^2(y)} + a(y)\nabla V(y) \rangle u_t^2(y) \, dy. \tag{3.51}$$

Then by (3.48), (3.50) and (3.51)

$$\begin{aligned} \frac{d}{dt} W_I^2(\mu_t^1, \mu_t^2) \le & -2 \int_{\mathbb{R}^n} \langle x - r(x), a(x)\nabla u_t^1(x) \rangle \, dx - 2 \int_{\mathbb{R}^n} \langle y - r^{-1}(y), a(y)\nabla u_t^2(y) \rangle \, dy \\ & - 2 \int_{\mathbb{R}^n} \langle x - r(x), a(x)\nabla V(x) \rangle u_t^1(x) \, dx - 2 \int_{\mathbb{R}^n} \langle y - r^{-1}(y), a(y)\nabla V(y) \rangle u_t^2(y) \, dy. \end{aligned} \tag{3.52}$$

Since $r_\# \mu_t^1 = \mu_t^2$, the second line of (3.52) can be written as

$$\begin{aligned} & - 2 \int_{\mathbb{R}^n} \langle x - r(x), a(x)\nabla V(x) \rangle u_t^1(x) \, dx - 2 \int_{\mathbb{R}^n} \langle y - r^{-1}(y), a(y)\nabla V(y) \rangle u_t^2(y) \, dy = \\ & = -2 \int_{\mathbb{R}^n} \langle x - r(x), a(x)\nabla V(x) - a(r(x))\nabla V(r(x)) \rangle u_t^1(x) \, dx. \end{aligned} \tag{3.53}$$

The estimate of the first line of (3.52) requires a bit more work. Since u_t^1 and u_t^2 belong to $W^{1,1}(\mathbb{R}^n)$ and r and r^{-1} are monotone and $a \in C^1(\mathbb{R}^n)$, using a similar argument of the proof of Lemma 10.4.5 of [5], and a truncation argument, the following weak formula of integration by parts holds ($\operatorname{tr}\nabla$ is the absolutely continuous part of the distributional divergence)

$$\int_{\mathbb{R}^n} \langle a(x)(r(x)-x), \nabla u_t^1(x)\rangle \, dx \le -\int_{\mathbb{R}^n} \operatorname{tr}\nabla(a(x)(r(x)-x))u_t^1(x)\, dx, \tag{3.54}$$

and similarly

$$\int_{\mathbb{R}^n} \langle a(y)(r^{-1}(y)-y), \nabla u_t^2(y)\rangle \, dy \le -\int_{\mathbb{R}^n} \operatorname{tr}\nabla(a(y)(r^{-1}(y)-y))u_t^2(y)\, dy. \tag{3.55}$$

Since

$$\begin{aligned} \operatorname{tr}\nabla(a(x)(r(x)-x)) &= \langle \nabla a(x), r(x)-x\rangle + a(x)\operatorname{tr}\nabla(r(x)-x) \\ &= \langle \nabla a(x), r(x)-x\rangle + a(x)(\operatorname{tr}\nabla r(x) - n), \end{aligned}$$

by (3.54), (3.55) and $r_{\#}\mu_t^1 = \mu_t^2$ we have

$$\begin{aligned} &-2\int_{\mathbb{R}^n} \langle x - r(x), a(x)\nabla u_t^1(x)\rangle \, dx - 2\int_{\mathbb{R}^n} \langle y - r^{-1}(y), a(y)\nabla u_t^2(y)\, dy \\ &\le 2\int_{\mathbb{R}^n} (\langle x-r(x), \nabla a(x)\rangle + (n - \operatorname{tr}\nabla r(x))a(x))u_t^1(x)\, dx \\ &\quad + 2\int_{\mathbb{R}^n} (\langle y - r^{-1}(y), \nabla a(y)\rangle + (n - \operatorname{tr}(\nabla r(r^{-1}(y)))^{-1})a(y))u_t^2(y)\, dy \\ &= 2\int_{\mathbb{R}^n} (\langle x-r(x), \nabla a(x)\rangle + (n - \operatorname{tr}\nabla r(x))a(x))u_t^1(x)\, dx \\ &\quad + 2\int_{\mathbb{R}^n} (\langle r(x) - x, \nabla a(r(x))\rangle + (n - \operatorname{tr}(\nabla r(x))^{-1})a(r(x)))u_t^1(x)\, dx \\ &= 2\int_{\mathbb{R}^n} \langle x - r(x), \nabla a(x) - \nabla a(r(x))\rangle u_t^1(x)\, dx \\ &\quad + 2\int_{\mathbb{R}^n} (na(x) + na(r(x)) - a(x)\operatorname{tr}\nabla r(x) - a(r(x))\operatorname{tr}(\nabla r(x))^{-1})u_t^1(x)\, dx. \end{aligned} \tag{3.56}$$

Since all the eigenvalues $\lambda_i(x)$ of $\nabla r(x)$ are strictly positive, we easily obtain

$$\begin{aligned} -a(x)\operatorname{tr}\nabla r(x) - a(r(x))\operatorname{tr}(\nabla r(x)^{-1}) &= -\sum_{i=1}^n (a(x)\lambda_i(x) + a(r(x))(\lambda_i(x))^{-1}) \\ &\le -2n\sqrt{a(x)}\sqrt{a(r(x))} \end{aligned}$$

and then by (3.56)

$$\begin{aligned} &-2\int_{\mathbb{R}^n} \langle x - r(x), a(x)\nabla u_t^1(x)\rangle \, dx - 2\int_{\mathbb{R}^n} \langle y - r^{-1}(y), a(y)\nabla u_t^2(y)\rangle \, dy \\ &\le 2\int_{\mathbb{R}^n} \big(\langle x - r(x), \nabla a(x) - \nabla a(r(x))\rangle + n\big(\sqrt{a(x)} - \sqrt{a(r(x))}\big)^2\big)u_t^1(x)\, dx. \end{aligned} \tag{3.57}$$

Combining (3.52) with (3.53) and (3.57) and using the assumption (1.65) we obtain

$$\frac{d}{dt} W_I^2(\mu_t^1, \mu_t^2) \leq -2\alpha \int_{\mathbb{R}^n} |x - r(x)|^2 u_t^1(x)\, dx = -2\alpha W_I^2(\mu_t^1, \mu_t^2)$$

and therefore

$$W_I(\mu_t^1, \mu_t^2) \leq e^{-\alpha t} W_I(\mu_0^1, \mu_0^2) \qquad \forall t \in (0, +\infty).$$

□

Acknowledgements

The author would like to thank Giuseppe Savaré for valuable suggestions about this work, and the referees for some useful observations.

References

[1] Martial Agueh. Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. *Adv. Differential Equations*, 10(3):309–360, 2005.

[2] Luigi Ambrosio. Minimizing movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)*, 19:191–246, 1995.

[3] Luigi Ambrosio. Transport equation and cauchy problem for non-smooth vector fields. Lecture Notes of the CIME Summer school in Cetrary, June 27-July 2, 2005, Available on line at http://cvgmt.sns.it/people/ambrosio/, 2005.

[4] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savarè. *Gradient flows in metric spaces and in the Wasserstein spaces of probability measures.* Birkhäuser, 2005.

[6] Anton Arnold, Peter Markowich, Giuseppe Toscani, and Andreas Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2):43–100, 2001.

[7] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[8] E. A. Carlen and W. Gangbo. Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math. (2)*, 157(3):807–846, 2003.

[9] E. A. Carlen and W. Gangbo. Solution of a model Boltzmann equation via steepest descent in the 2-Wasserstein metric. *Arch. Ration. Mech. Anal.*, 172(1):21–64, 2004.

[10] 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.

[11] J.A. Carrillo, R.J. McCann, and C. Villani. Contractions in the 2-wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179:217–263, 2006.

[12] José A. Carrillo, Robert J. McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[13] J. Crank. *The mathematics of diffusion*. Clarendon Press, Oxford, second edition, 1975.

[14] Giuseppe De Cecco and Giuliana Palmieri. Intrinsic distance on a Lipschitz Riemannian manifold. *Rend. Sem. Mat. Univ. Politec. Torino*, 46(2):157–170 (1990), 1988.

[15] Giuseppe De Cecco and Giuliana Palmieri. Intrinsic distance on a LIP Finslerian manifold. *Rend. Accad. Naz. Sci. XL Mem. Mat. (5)*, 17:129–151, 1993.

[16] Giuseppe De Cecco and Giuliana Palmieri. LIP manifolds: from metric to Finslerian structure. *Math. Z.*, 218(2):223–237, 1995.

[17] Ennio De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 81–98. Masson, Paris, 1993.

[18] Ennio De Giorgi, Antonio Marino, and Mario Tosques. Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 68(3):180–187, 1980.

[19] Marco Degiovanni, Antonio Marino, and Mario Tosques. Evolution equations with lack of convexity. *Nonlinear Anal.*, 9(12):1401–1443, 1985.

[20] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[21] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[22] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17 (electronic), 1998.

[23] David Kinderlehrer and Adrian Tudorascu. Transport via mass transportation. *Discrete Contin. Dyn. Syst. Ser. B*, 6(2):311–338 (electronic), 2006.

[24] Stefano Lisini. Characterization of absolutely continuous curves in wasserstein spaces. *Calc. Var. Partial Differential Equations*, 28(1):85–120, 2007.

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

[26] Felix Otto. Doubly degenerate diffusion equations as steepest descent. *Manuscript*, 1996.

[27] Felix Otto. Evolution of microstructure in unstable porous media flow: a relaxational approach. *Comm. Pure Appl. Math.*, 52(7):873–915, 1999.

[28] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[29] Luca Petrelli and Adrian Tudorascu. Variational principle for general diffusion problems. *Appl. Math. Optim.*, 50(3):229–257, 2004.

[30] Karl-Theodor Sturm. Convex functionals of probability measures and nonlinear diffusions on manifolds. *J. Math. Pures Appl. (9)*, 84(2):149–168, 2005.

[31] Juan Luis Vázquez. *The porous medium equation.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, Oxford, 2007. Mathematical theory.

[32] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[33] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.