Infinite-dimensional porous media equations and optimal transportation

Luigi Ambrosio[*]
Scuola Normale Superiore, Pisa

Edoardo Mainini [†]
Scuola Normale Superiore, Pisa

Abstract

In this paper we study a class of non linear diffusion equations in a Hilbert space X,

$$\partial_t \mu_t - \nabla \cdot (\nabla (L \circ \rho_t)\gamma) = 0 \qquad \text{in } X \times (0, +\infty),$$

with respect to a log-concave reference probability measure γ. We obtain existence, uniqueness and stability properties, in the framework of gradient flows in spaces of probability measures.

1 Introduction

In the last few years, starting from the seminal papers [O1, JKO], many studies have been devoted to the description of classical and non-classical PDE's as evolution problems of gradient flow type in the space of probability measures, endowed with the quadratic optimal transportation distance W_2. Here we just mention [A, CG1, CG2, O2, O3, O4] and we refer to the monographs [AGS, VI] for a detailed (but already not completely up to date) description of the literature. It turns out that this interpretation as a gradient flow, when associated to a convex structure, is extremely useful to derive existence, stability results and trends to equilibrium. A systematic theory of these evolution problems, which covers also infinite-dimensional state spaces, has been developed in [AGS]. In [ASZ], building upon many results in [AGS], the authors obtained general existence and stability results for infinite-dimensional Fokker-Planck equations in Hilbert spaces associated to log-concave probability measures γ; the idea is to view the PDE as the gradient flow of the relative entropy functional

$$\rho\gamma \mapsto \int_X \rho \ln \rho \, d\gamma$$

with respect to W_2, and log-concavity of γ is (see [AGS]) precisely the property needed for convexity. More recently this results have been extended to the Ornstein-Uhlenbeck operator in Wiener spaces (see [FSS, MA]).

[*]l.ambrosio@sns.it
[†]edoardo.mainini@sns.it

In this paper we investigate more in detail the nonlinear counterpart of these results, corresponding to general energies

$$\mu = \rho\gamma \mapsto \mathscr{F}(\mu) := \int_X F(\rho)\, d\gamma, \qquad \rho\gamma \in \mathscr{P}_2(X) \tag{1.1}$$

(set equal to $+\infty$ if μ is not absolutely continuous with respect to γ). Here we shall denote by $\mathscr{P}(X)$ (resp. $\mathscr{P}_2(X)$) the space of probability measures (resp. probability measures with finite quadratic moment) on the separable Hilbert space X. In particular we obtain well-posedness and regularizing properties for nonlinear evolution equations of the form

$$\begin{cases} \partial_t \mu_t - \nabla \cdot (\nabla (L \circ \rho_t)\gamma) = 0 \quad \text{in } X \times (0, +\infty), \\ \lim_{t \downarrow 0} \mu_t = \bar{\mu}, \end{cases} \tag{1.2}$$

where ρ_t represents the density of μ_t with respect to γ and $L = L_F : \mathbb{R} \to \mathbb{R}$ is the Legendre transform of F (so that the linear Fokker-Planck equation of [ASZ] corresponds to $F(z) = z \ln z$). The reader may consult [DaP, DaPZ] for a systematic study of evolution PDE's in infinite dimensions and the monograph [VA] for the finite-dimensional theory of porous media equations.

It should be emphasized that, as soon as a convex structure is identified, the results in [AGS] provide existence and uniqueness of the gradient flow, and several equivalent formulations of the evolution problem; but, the interpretation of this evolution in conventional PDE terms might not be immediate; in the case of Fokker-Planck equations, the connection with the point of view of Dirichlet forms and of Markov processes is completely analyzed in [ASZ], and tools from the theory of optimal transportation are used to show closability of the Dirichlet form $\int \|\nabla u\|^2\, d\gamma$.

In the nonlinear context provided by (1.1), our goal is relate the evolution semigroup in $\mathscr{P}_2(X)$ to the classical viewpoint based on Sobolev spaces and integration by parts. To this aim, we assume that an orthonormal basis (that we shall denote by $\mathbf{e}_j$) of X exists, such that $\partial_{\mathbf{e}_j}\gamma \ll \gamma$ for all $j \geq 1$; notice that this assumption is consistent with the model case of Gaussian measures γ. Notice however that it is not needed for the existence of the evolution semigroup in $\mathscr{P}_2(X)$. On the other hand, in order to have a convex structure we need some structural assumptions on F which cover all nonlinearities $F(z) = z^m$, $m > 1$ (see Assumption 4.1) and the log-concavity of γ. This last hypothesis covers all measures γ of the form $e^{-V}\gamma_G$ with γ_G Gaussian and V convex and lower semicontinuous, but we don't need any absolute continuity assumption w.r.t. a Gaussian.

One of the possible equivalent descriptions of the gradient flow in $\mathscr{P}_2(X)$ takes the form of a continuity equation (in the weak sense of duality with cylindrical functions)

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \tag{1.3}$$

coupled with a constitutive equation relating $v_t \in L^2(\mu_t; X)$ to μ_t, namely $-v_t = \partial^0 \mathscr{F}(\mu_t)$. In this context, $\partial^0 \mathscr{F}(\rho\gamma)$ is the element with minimal $L^2(X, \rho\gamma; X)$ norm of $\partial \mathscr{F}(\rho\gamma)$ and the subdifferential relation $\xi \in \partial \mathscr{F}(\rho\gamma)$ reads

$$\mathscr{F}(\sigma) \geq \mathscr{F}(\rho\gamma) + \int_X \langle \xi(x), \mathbf{t}(x) - x \rangle\, \rho(x)\, d\gamma(x) \qquad \forall \sigma \in \mathscr{P}_2(X).$$

Here we are denoting by $\mathbf{t}$ the optimal transport map between $\rho\gamma$ and σ, and it turns out that the absolute continuity of all measures $\partial_{\mathbf{e}_j}\gamma$ suffices to show in Theorem 3.2 (following with minor variants [AGS, 6.2.10]) existence and uniqueness of optimal transport maps. In comparison with [AGS] our analysis is simplified by the choice of the quadratic exponent and by the existence of optimal maps, so that Kantorovich plans do not play an explicit role.

So, most of this paper will be devoted to the identification of $\partial^0\mathscr{F}(\rho\gamma)$ and, in comparison to the linear Gaussian case considered in [AGS, 10.4.8], new difficulties are due to the nonlinearity and to the generality of γ. If $\rho \in L^\infty(X,\gamma)$, we shall prove that $\partial\mathscr{F}(\rho\gamma)$ is not empty if and only if $L_F \circ \rho \in W^{1,1}(X,\gamma)$ and $\nabla(L_F \circ \rho)/\rho \in L^2(X,\rho\gamma;X)$; if this is the case, then

$$\frac{\nabla(L_F \circ \rho)}{\rho} = \partial^0\mathscr{F}(\rho\gamma). \tag{1.4}$$

In the case of unbounded densities ρ, membership to the Sobolev space can not be defined because we assume only $\partial_{\mathbf{e}_j}\gamma \ll \gamma$ (the assumption $|\partial_{\mathbf{e}_j}\gamma| \le C\gamma$ would be incompatible even with the Gaussian case) and the integration by parts formula does not make sense. To overcome this difficulty, we define (in the same spirit as [BBGG, DMOP]) generalized Sobolev spaces $GW^{1,1}(X,\gamma)$ in the "entropy" sense, by requiring that the truncated functions $T_\alpha(\rho) = -\alpha \vee \rho \wedge \alpha$ belong to $W^{1,1}(X,\gamma)$ for all $\alpha \ge 0$ (see also [C] for a definition of entropy solutions to some degenerate evolution equations). In this class a gradient can still be defined and (1.4) remains true. Replacing (1.4) into (1.3) leads to the following result:

Theorem 1.1 *Assume that $L = L_F$, with F satisfying* Assumption 4.1*, and that γ satisfies* Assumption 2.5*. Then, for all $\bar\mu \in \mathscr{P}_2(X)$ there exists a distributional solution $\mu_t = \rho_t\gamma$ to* (1.2)*, satisfying $L_F \circ \rho_t \in GW^{1,1}(X,\gamma)$ for a.e. $t > 0$ and:*

$$\left\| \frac{\nabla(L_F \circ \rho_t)}{\rho_t} \right\|_{L^2(X,\mu_t;X)} \in L^2_{loc}(0,+\infty). \tag{1.5}$$

In the class of solutions μ_t satisfying (1.5) *this solution is unique. Furthermore, if $\bar\mu \le C\gamma$, then $\rho_t \le C$ γ-a.e. for all $t > 0$ and therefore $L_F \circ \rho_t \in W^{1,1}(X,\gamma)$ for a.e. $t > 0$.*

The solution inherits from the gradient flow representation also additional properties, listed in Theorem 3.8 and Remark 3.9: here we just mention that it is described by a contraction semigroup on $\mathscr{P}_2(X)$. We conclude noticing that our strategy (based on the perturbation argument, as in [AGS, Remark 10.4.7]) identifies only the element with minimal norm and not the whole $\partial\mathscr{F}(\rho\gamma)$, in contrast with the known finite-dimensional result, recalled in Theorem 4.6. Since the differential inclusion $v_t \in -\partial\mathscr{F}(\mu_t)$ is equivalent to the equation $v_t = -\partial^0\mathscr{F}(\mu_t)$, our result is sufficient to identify the PDE (1.2). A direct analysis of the subdifferential relation seems to require change of variables formulas relative to γ, a problem still open under our weak assumptions on γ.

1.1 Plan of the paper

In Section 2 we will introduce the notation and the main technical tools used in the paper. In Section 3 we will recall, mostly from [AGS], the abstract theory of gradient flows, with particular attention to the case of geodesically convex functionals on $(\mathscr{P}_2(X), W_2)$. In Section 4 we will introduce all the hypotheses and properties enjoyed by the internal energy functional $\mathscr{F}$, quoting in particular the known results in the finite dimensional case. Moreover, we will introduce some useful Γ-convergence approximation techniques. Finally in Section 5 we will characterize equation (1.2) as the Wasserstein gradient flow of $\mathscr{F}$ and prove the main result.

2 Notation and tools

2.1 Probability spaces

Let X be a separable Hilbert space with norm $\|\cdot\|$. The set of probability measures $\mathscr{P}(X)$ will be endowed with the usual weak topology, induced by the duality with continuous and bounded functions on X. If (μ_n) weakly converges to μ, we will write $\mu_n \rightharpoonup \mu$. The set of probability measures on X with finite quadratic moment will be denoted with $\mathscr{P}_2(X)$, that is

$$\mathscr{P}_2(X) := \left\{ \mu \in \mathscr{P}(X) \quad s.t. \quad \int_X \|x\|^2 d\mu(x) < +\infty \right\}. \tag{2.1}$$

We will say that μ_n converge to μ in $\mathscr{P}_2(X)$ if $\mu_n \rightharpoonup \mu$ and

$$\lim_{n\to\infty} \int_X \|x\|^2 d\mu_n = \int_X \|x\|^2 d\mu. \tag{2.2}$$

In this case we will write $\mu_n \to \mu$; this convergence is equivalent to the one induced by the distance W_2 defined in (3.1) below, see for instance [AGS, Proposition 7.1.5].

We now introduce the *push forward* notation: given a Borel map $\mathbf{t} : X \to Y$ and $\mu \in \mathscr{P}(X)$, $\mathbf{t}_\# \mu \in \mathscr{P}(Y)$ is defined by $(\mathbf{t}_\# \mu)(A) = \mu(\mathbf{t}^{-1}(A))$, for any Borel set $A \subset X$.

Next, given an orthonormal basis $(\mathbf{e}_i)$ of X (a specific choice of $(\mathbf{e}_i)$, induced by γ, will be specified later on), we consider the canonical projection maps $\pi^d(x) : X \to \mathbb{R}^d$, of the form

$$\pi^d(x) = (\langle x, \mathbf{e}_1\rangle, \ldots, \langle x, \mathbf{e}_d\rangle).$$

Definition 2.1 (Smooth cylindrical functions) *We say that $\varphi : X \to \mathbb{R}$ is a smooth cylindrical function if $\varphi = \psi \circ \pi^d$, where π^d is a projection map and $\psi \in C_c^\infty(\mathbb{R}^d)$. The set of smooth cylindrical functions on X will be denoted by* $\mathrm{Cyl}(X)$.

Definition 2.2 (Cylindrical projections) *If $\nu = u\gamma \ll \gamma$ and $\gamma_d := \pi^d_\# \gamma$, then $\pi^d_\# \nu \ll \gamma_d$ and its density u_d is explicitly given by*

$$u_d(x) = \int_X u(y)\, d\gamma_x(y), \tag{2.3}$$

where γ_x is the family of measures, concentrated on $(\pi_d)^{-1}(x)$, which disintegrate γ with respect to γ_d. We shall call u_d cylindrical projections of u.

In addition, using for instance (2.3), one can prove that if $u \in L^p(X,\gamma)$, $p \in [1,+\infty)$, then $u_d \in L^p(X,\gamma_d)$ and

$$u_d \circ \pi^d \to u \quad \text{in } L^p(X,\gamma) \text{ as } d \to \infty. \tag{2.4}$$

Notice that (2.3) makes sense (componentwise) also for maps u taking values in X, and if $u \in L^p(X,\gamma;X)$, then $u^d \circ \pi_d \to u$ in $L^p(X,\gamma;X)$.

In order to deal with sequences of pairs $(\boldsymbol{\rho}_n,\mu_n)$, where $\boldsymbol{\rho}_n \in L^p(X,\mu_n;X)$ and μ_n are measures on X, we will need the following notion of convergence.

Definition 2.3 *Let $(\mu_n) \in \mathscr{P}(X)$ be weakly convergent to μ. Let $\boldsymbol{\rho}_n \in L^1(X,\mu_n;X)$ and $\boldsymbol{\rho} \in L^1(X,\mu;X)$. We say that $\boldsymbol{\rho}_n$ weakly converge to $\boldsymbol{\rho}$ if*

$$\lim_{n\to\infty} \int_X \zeta(x)\rho_n^j d\mu_n(x) = \int_X \zeta(x)\rho^j d\mu(x) \tag{2.5}$$

for any $\zeta \in \mathrm{Cyl}(X)$ and $j \in \mathbb{N}$, where ρ_n^j and ρ^j are respectively the components of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_n$ along the basis $(\mathbf{e}_j)$.
We say that $\boldsymbol{\rho}_n$ strongly converge to $\boldsymbol{\rho}$ in L^p, $p > 1$, if in addition it holds

$$\lim_{n\to\infty} \|\boldsymbol{\rho}_n\|_{L^p(X,\mu_n;X)} = \|\boldsymbol{\rho}\|_{L^p(X,\mu;X)}. \tag{2.6}$$

Analogously, in the scalar case we say that $\rho_n \in L^1(X,\mu_n)$ weakly converge to $\rho \in L^1(X,\mu)$ if

$$\lim_{n\to\infty} \int_X \zeta(x)\rho_n d\mu_n(x) = \int_X \zeta(x)\rho d\mu(x) \qquad \forall \zeta \in \mathrm{Cyl}(X) \tag{2.7}$$

and strong L^p convergence requires $\|\rho_n\|_{L^p(X,\mu_n)} \to \|\rho\|_{L^p(X,\mu)}$ as $n \to \infty$.

Moreover, in the sequel we will take advantage of the following result (see [AGS, Theorem 5.4.4]).

Lemma 2.4 *Let $\mu_n \to \mu$ in $\mathscr{P}_2(X)$. If ρ_n strongly converge to ρ in $L^2(X,\mu)$ then*

$$\lim_{n\to\infty} \int_X f(x,\rho_n(x))d\mu_n(x) = \int_X f(x,\rho(x))d\mu \tag{2.8}$$

for every continuous function f with at most 2-growth, that is

$$|f(x,y)| \le A + B(\|x\|^2 + \|y\|^2) \quad \forall (x,y) \in X \times X \tag{2.9}$$

for some A, $B \in \mathbb{R}$. More generally, (2.8) holds also if strong L^2 convergence is replaced by

$$\lim_{n\to\infty} \int_X g(\rho_n(x))d\mu_n = \int_X g(\rho(x))d\mu$$

for some strictly convex function $g : X \to \mathbb{R}$ with at least 2-growth at infinity.

2.2 Partial derivatives and gradient in Hilbert spaces

Let γ be a probability measure on X and $v \in X$, $v \neq 0$. The Fomin distributional derivative (see for instance [B]) $\partial_v \gamma$ is defined by the canonical duality

$$\langle \partial_v \gamma, \varphi \rangle = -\int_X \partial_v \varphi \, d\gamma, \quad \varphi \in \mathrm{Cyl}(X)$$

where $\partial_v \varphi$ is the partial derivative of φ in the direction v. We say that $\partial_v \gamma$ is an absolutely continuous measure with respect to γ if there exists $g \in L^1(X,\gamma)$ such that

$$\int_X \partial_v \varphi d\gamma = -\int_X g\varphi d\gamma, \quad \forall \varphi \in \mathrm{Cyl}(X). \tag{2.10}$$

Throughout this paper we shall make the following assumption:

Assumption 2.5 *$\partial_{\mathbf{e}_j} \gamma \ll \gamma$ for all $j \geq 1$. The corresponding Radon-Nikodym derivatives will be denoted by g^j.*

Now we can define the distributional partial derivative of a bounded function (see for instance [B]).

Definition 2.6 (Partial derivative, gradient, Sobolev spaces) *Under* Assumption 2.5*, a function $u \in L^\infty(X,\gamma)$ has partial derivative $\eta^j \in L^1(X,\gamma)$ if*

$$\int_X \partial_{\mathbf{e}_j} \zeta(x) u(x) \, d\gamma(x) = -\int_X \eta^j(x)\zeta(x) \, d\gamma(x) + \int_X u(x)\zeta(x) g^j(x) \, d\gamma(x) \quad \forall \zeta \in \mathrm{Cyl}(X). \tag{2.11}$$

In this case, we write $\eta^j := \partial^\gamma_{\mathbf{e}_j} u$, and simply $\partial_{\mathbf{e}_j} u$ when no ambiguity arises. In addition, if this happens for all $j \geq 1$ and $\sqrt{\sum_j (\partial_{\mathbf{e}_j} u)^2} \in L^p(X,\gamma)$, we write $u \in W^{1,p}(X,\gamma)$ and set

$$\nabla u := \sum_{j=1}^\infty (\partial_{\mathbf{e}_j} u)\mathbf{e}_j \in L^p(X,\gamma;X).$$

We shall also use the fact that

$$\partial_{\mathbf{e}_j} u_d = (\partial_{e_j} u)_d \tag{2.12}$$

whenever $\partial_{\mathbf{e}_j} u$ exists and $d \geq j$.

We shall also need a chain rule formula and an existence result γ-a.e. of directional derivatives of Lipschitz functions; we recall briefly their proofs, that can be achieved by standard arguments.

Theorem 2.7 (Chain rule) *Let $u \in L^\infty(X,\gamma)$ with $\partial_{\mathbf{e}_j} u \in L^1(X,\gamma)$, and let $f \in Lip(\mathbb{R})$. Then $\partial_{\mathbf{e}_j}(f \circ u) \in L^1(X,\gamma)$ and*

$$\partial_{\mathbf{e}_j}(f \circ u) = f'(u)\partial_{\mathbf{e}_j} u \quad \gamma\text{-a.e. in } X. \tag{2.13}$$

More precisely, denoting by Σ the set where f is not differentiable, both $\partial_{\mathbf{e}_j} u = 0$ and $\partial_{\mathbf{e}_j}(f \circ u) = 0$ γ-a.e. on $u^{-1}(\Sigma)$, where (2.13) does not make sense.

Proof. We denote by Y the orthogonal subspace to $\mathbf{e}_j$, by $\pi : X \to Y$ the orthogonal projection and for $y \in Y$ we denote by $\gamma_y \in \mathscr{P}(\mathbb{R})$ the conditional probability measures, namely

$$\gamma(B) = \int_Y \gamma_y\big(\{t : \ y + t\mathbf{e}_j \in B\}\big)\, d\pi_\# \gamma$$

for all Borel sets $B \subset X$. We claim that $\gamma_y \ll \mathcal{L}^1$ for $\pi_\#\gamma$-a.e. y. To prove this, we shall prove that γ_y has derivative equal to $f_y\gamma_y$, where $f_y(t) = \beta_j(y + t\mathbf{e}_j)$, and use the well known fact that this property, on the real line, implies absolute continuity.

To prove the claim, fix $\zeta \in \mathrm{Cyl}(Y)$ and $\psi \in C^1_c(\mathbb{R})$ and notice that

$$\begin{aligned}
\int_Y \zeta(y) \int_{\mathbb{R}} \psi'(t)\, d\gamma_y(t)\, d\pi_\#\gamma(y) &= \int_X \zeta(\pi(x))\psi'(x_j)\, d\gamma(x) \\
&= -\int_X \zeta(\pi(x))\psi(x_j) g^j(x)\, d\gamma(x) \\
&= -\int_Y \zeta(y) \int_{\mathbb{R}} \psi(t) f_y(t)\, d\gamma_y(t)\, d\pi_\#\gamma(y).
\end{aligned}$$

Since ζ is arbitrary, $\int_{\mathbb{R}} \psi'(t)\, d\gamma_y(t) = -\int_{\mathbb{R}} \psi(t) f_y(t)\, d\gamma_y(t)$ for $\pi_\#\gamma$-a.e. y. We can find a $\pi_\#\gamma$-negligible set $Y' \subset Y$ such that the equality holds for all $y \in Y \setminus Y'$ and all ψ in a countable dense set in $C^1_c(\mathbb{R})$. By density, the claimed property holds for all $y \in Y \setminus Y'$.

With a very similar argument one can prove a second claim, that $u_y(t) := u(y + tv)$ is differentiable according to (2.11) with $X = \mathbb{R}$, $\gamma = \gamma_y$, for $\pi_\#\gamma$-a.e. y, with $\partial^{\gamma_y} u_y(t) = \partial^\gamma_{\mathbf{e}_j} u(y + t\mathbf{e}_j)$. Having proved the claims, the conclusion of the proof is standard: first the statement is proved for u_y, γ_y, and then, using the conditional probability representation of γ, it is extended to u, γ.

So, it remains to prove the chain rule formula in the case when $X = \mathbb{R}$, $\gamma = h\mathcal{L}^1$, with $h' = hg \in L^1(\mathbb{R})$. In this case we shall use the fact that this property holds for the classical distributional derivative (see for instance [EG, Chapter 4]), or [AFP, Theorem 3.99] for a more general result); we can read the integration by parts formula

$$\int_{\mathbb{R}} uh\zeta'\, dt = \int_{\mathbb{R}} ugh\zeta\, dt - \int_{\mathbb{R}} \partial^\gamma u h\zeta\, dt \tag{2.14}$$

by saying that $v := uh \in W^{1,1}(\mathbb{R})$ and $h\partial^\gamma u = v' - uh'$. Since h is continuous it follows that $u = v/h \in W^{1,1}_{\mathrm{loc}}(\{h > 0\})$ and the classical product rule in Sobolev spaces gives $\partial^\gamma u = u'$ in $\{h > 0\}$. Conversely, if a bounded function w belongs to $W^{1,1}_{\mathrm{loc}}(\{h > 0\})$ and $w' \in L^1(\gamma)$, then $w \in W^{1,1}(\mathbb{R}, \gamma)$ and $\partial^\gamma w = w'$: indeed, under these assumptions (2.14) with $u = w$ holds when ζ has support contained in $\{h > 0\}$, and by approximation it holds for all ζ of the form $\tilde\zeta h/\sqrt{h^2 + \varepsilon^2}$ with $\tilde\zeta \in C^1_c(\mathbb{R})$. Letting $\epsilon \to 0$ easily gives

$$\int_{\mathbb{R}} wh\tilde\zeta'\, dt = \int_{\mathbb{R}} wgh\tilde\zeta\, dt - \int_{\mathbb{R}} \partial^\gamma wh\tilde\zeta\, dt$$

because the extra term

$$\int_{\mathbb{R}} wh\zeta\Big(\frac{h}{\sqrt{h^2 + \varepsilon^2}}\Big)'\, dt$$

coming from the differentiation of $h/\sqrt{h^2+\varepsilon^2}$ can be estimated, up to the multiplicative constant $\sup|w\zeta|$, by

$$\frac{\epsilon^2 h|h'|}{\sqrt{h^2+\epsilon^2}^3} \le \frac{\epsilon^2|h'|}{h^2+\epsilon^2} \le |h'|$$

and tends to 0 pointwise.

Obviously $w = f(u)$ is locally Sobolev on $\{h > 0\}$ and $w' = f'(u)u'$ on $\{h > 0\} \setminus u^{-1}(\Sigma)$, and equal to 0 on $u^{-1}(\Sigma)$. See Proposition 3.92 and Theorem 3.99 in [AFP]. $\square$

Theorem 2.8 (Partial derivatives) *Let $f : X \to \mathbb{R}$ be Lipschitz and assume that $\partial_v\gamma \ll \gamma$. Then*

$$\exists \lim_{t\downarrow 0} \frac{f(x+tv)-f(x)}{t} \qquad \text{for } \gamma\text{-a.e. } x. \tag{2.15}$$

Proof. As in the proof of Theorem 2.7, one can prove that the conditional measures γ_y induced by the map $x \mapsto x - \langle x, v\rangle v$, indexed by $y \in v^\perp$, are absolutely continuous with respect to $\mathcal{L}^1$ for $\pi_\#\gamma$-a.e. $y \in v^\perp$, where π is the orthogonal projection on $v^\perp$. Then, the existence $\mathcal{L}^1$-a.e. of the derivative of $t \mapsto f(y+tv)$ yields existence of the derivative γ_y-a.e. in X. We conclude that the limit (2.15) exists γ-a.e. in X. $\square$

Definition 2.6 makes sense for $L^\infty(X,\gamma)$ functions. In order to treat the unbounded case, we will need a generalized definition of Sobolev spaces, based on truncation. For $u : X \to \mathbb{R}$ and $\alpha \ge 0$, define the α-truncate of u by

$$T_\alpha(u) := -\alpha \vee u \wedge \alpha. \tag{2.16}$$

Suppose that $T_n(u) \in W^{1,1}(X,\gamma)$ for every integer n. Thanks to Theorem 2.7, there holds $\nabla T_n u = 0$ γ-a.e. on $\{|u| > n\}$. Moreover,

$$\nabla T_n u = \nabla T_m u \qquad \gamma\text{-a.e. on } \{|u| < n\} \tag{2.17}$$

for $n < m$, since the two functions are equal on $\{|u| < n\}$. Hence we can define

$$\partial^\gamma_{\mathbf{e}_j} u := \partial^\gamma_{\mathbf{e}_j} T_n(u) \qquad \gamma\text{-a.e. on } \{|u| < n\}, \tag{2.18}$$

$$\nabla u := \nabla T_n(u) \qquad \gamma\text{-a.e. on } \{|u| < n\} \tag{2.19}$$

and this is a good definition, up to γ-negligible sets, because of (2.17) (and because we used only a countable set of truncation levels).

Definition 2.9 (Generalized Sobolev spaces) *We say that a Borel map $u : X \to \mathbb{R}$ belongs to $GW^{1,p}(X,\gamma)$ if $T_\alpha(u) \in W^{1,p}(X,\gamma)$ for all $\alpha \ge 0$. The partial derivatives and the gradient of u are defined as in* (2.18) *and* (2.19)*.*

Notice that we might equivalently require only $T_n(u) \in W^{1,p}(X,\gamma)$ for all integers n: this follows by applying the chain rule with $f = T_\alpha$ to the identity $T_\alpha = T_\alpha \circ T_n$, for $n > \alpha$. Similarly one can prove that any unbounded sequence of truncation levels would provide an equivalent definition.

3 Wasserstein structure and gradient flows in probability spaces

In $\mathscr{P}_2(X)$ we introduce the following distance

$$W_2^2(\mu,\nu) = \inf\left\{ \int_{X\times X} \|y-x\|^2 d\beta(x,y) \quad s.t. \quad \beta \in \Gamma(\mu,\nu) \right\} \tag{3.1}$$

where $\Gamma(\mu,\nu)$ denote the subset of $\mathscr{P}(X\times X)$ of measures with first marginal μ and second marginal ν. It is well known that the infimum is achieved, and we shall denote by $\Gamma_0(\mu,\nu)$ the set of *optimal transport plans*, corresponding to solutions of the Kantorovich optimal transport problem. A transport plan can be seen as a multivalued generalization of a *transport map*, that is, a Borel map $\mathbf{t}: X \to X$ such that $\mathbf{t}_{\#}\mu = \nu$.

Let us consider a functional $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$, and define its effective domain as

$$D(\phi) = \{\mu \in \mathscr{P}_2(X) : \phi(\mu) < +\infty\}.$$

We say that the functional is proper if $D(\phi) \neq \emptyset$.

If $\mu \in D(\phi)$, the *metric slope* of ϕ at μ is defined by

$$|\partial\phi|(\mu) = \limsup_{\nu\to\mu} \frac{(\phi(\mu)-\phi(\nu))^+}{W_2(\mu,\nu)}.$$

From now on, the following hypotheses on the functional will be assumed:

Assumption 3.1 *$\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ is a proper lower semicontinuous functional, it is bounded from below and such that for all $\mu \in D(\phi)$, $\nu \in \mathscr{P}_2(X)$ there exists an optimal transport map $\mathbf{t}$ from μ to ν.*

As a matter of fact, existence of optimal maps simplifies considerably some proofs and constructions, although almost all arguments can be reproduced working with transport plans. The assumption will be satisfied if $\phi(\mu)$ finite implies $\mu \ll \gamma$ and Assumption 2.5 holds:

Theorem 3.2 (Existence of optimal maps) *Assume that $\partial_{\mathbf{e}_j}\gamma \ll \gamma$ for all $j \geq 1$, $\mu, \nu \in \mathscr{P}_2(X)$ and $\mu \ll \gamma$. Then there exists a unique optimal transport plan from μ to ν, and this plan is induced by a map.*

Proof. The proof follows the traditional routine (see for instance [AGS, 6.2.10] for closely related results, involving measures μ vanishing on Gauss null sets): one reduces to the case when ν has a bounded support and finds an optimal plan β and a maximizing pair (φ,ψ) of Kantorovich potentials, so that $\varphi(x)+\psi(y) \leq |x-y|^2$ and equality holds on suppβ; since

$$\varphi(x) = \inf_{y\in \operatorname{supp}\nu} |x-y|^2 - \psi(y)$$

we have that φ is Lipschitz on bounded sets. Then, by applying a local version of Theorem 2.8, we find a γ-negligible set $N \subset X$ such that $\partial_{\mathbf{e}_j}\varphi$ exists at all points of $X \setminus N$ for all $j \geq 1$. Since

$|x' - y|^2 - \varphi(x')$ attains its minimum at $x' = x$ (equal to $-\psi(y)$) for points $(x, y) \in \mathrm{supp}\beta$, if $x \notin N$ partial differentiation gives

$$2\langle x - y, \mathbf{e}_j\rangle = \partial_{\mathbf{e}_j}\varphi(x), \qquad \forall j \geq 1.$$

Since $\beta(N \times X) = \mu(N) = 0$, this proves that y is uniquely determined by x β-a.e., hence β is concentrated on a graph. This provides the optimal transport map. Since any optimal plan β is concentrated on the graph of a map, the optimal map is unique (otherwise a combination of two optimal maps would produce an optimal plan not concentrated on a graph) and, as a consequence, β is unique as well. □

Remark 3.3 (Stability of optimal maps) Let $\mu \in \mathscr{P}_2(X)$ with $\mu \ll \gamma$. Arguing as in [AGS, Lemma 8.5.3], uniqueness at the level of transport plans provides a strong continuity property of optimal transport maps, namely $\nu_n \to \nu$ in $\mathscr{P}_2(X)$ implies convergence in $L^2(\mu; X)$ of the optimal transport maps $\mathbf{t}_n$ from μ to ν_n to the optimal transport map $\mathbf{t}$ from μ to ν.

The next definition corresponds to the standard Fréchet subdifferential in Hilbert spaces.

Definition 3.4 (Wasserstein subdifferential) *Let $\mu \in D(\phi)$. The Wasserstein subdifferential $\partial\phi(\mu)$ of ϕ at μ is the set of vectors $\xi \in L^2(X, \mu; X)$ such that*

$$\phi(\nu) - \phi(\mu) \geq \int_X \langle \xi, \mathbf{t}(x) - x\rangle \, d\mu(x) + o(W_2(\mu, \nu)),$$

where $\mathbf{t}$ is the optimal transport map between μ and ν.

Definition 3.5 (Convexity along geodesics) *We say that ϕ is convex along geodesics if, for all μ, $\nu \in D(\phi)$, we have*

$$\phi(\mu_t) \leq (1 - t)\phi(\mu) + t\phi(\nu) \qquad \forall t \in [0, 1], \tag{3.2}$$

where $\mu_t = ((1 - t)Id + t\mathbf{t})_{\#}\mu$ is the constant speed Wasserstein geodesic connecting μ to ν.

We point out that for convex functionals along geodesics the subdifferential relation can be equivalently written as

$$\phi(\nu) - \phi(\mu) \geq \int_X \langle \xi(x), \mathbf{t}(x) - x\rangle \, d\mu(x). \tag{3.3}$$

See Section 10.1.1 in [AGS].

The next proposition shows the connection between Wasserstein subdifferential and metric slope (and it is analogous to the case of Fréchet subdifferential in Hilbert spaces). See for example [AGS, Lemma 10.1.5].

Proposition 3.6 (Slope and minimal selection in $\partial\phi$) *Let ϕ be a convex functional along geodesics, and let $\mu \in D(\phi)$. Then*

$$|\partial\phi|(\mu) = \min\left\{\|\xi\|_{L^2(X,\mu:X)} : \xi \in \partial\phi(\mu)\right\}$$

with the convention $\min\emptyset = +\infty$. If $|\partial\phi|(\mu)$ is finite, the minimum point is unique and we denote it with $\partial^0\phi(\mu)$.

If ϕ is convex along geodesics, the application $\mu \mapsto |\partial\phi|(\mu)$ is also lower semicontinuous in $(\mathscr{P}_2(X), W_2)$, as this fact holds true in general metric spaces (see [AGS, Corollary 2.4.10]).

We will also need the following stronger notion of convexity:

Definition 3.7 *We say that $\phi : \mathscr{P}_2(X) \to [0, +\infty]$ is strongly convex if it is convex along geodesics and for any $\mu, \nu, \sigma \in D(\phi)$ there exists a continuous curve $\mu_t : [0,1] \to \mathscr{P}_2(X)$, with $\mu_0 = \mu$ and $\mu_1 = \nu$, such that*

$$\begin{cases} W_2^2(\mu_t, \sigma) \le (1-t) W_2^2(\mu, \sigma) + t W_2^2(\nu, \sigma) - t(1-t) W_2^2(\mu, \nu) \\ \phi(\mu_t) \le (1-t)\phi(\mu) + t\phi(\nu) \end{cases} \qquad \forall t \in [0,1]. \tag{3.4}$$

3.1 The gradient flow equation

We say that a curve $\mu_t : (0, +\infty) \to \mathscr{P}_2(X)$ is absolutely continuous, and we write $\mu_t \in AC_{\rm loc}((0,+\infty), \mathscr{P}_2(X))$, if for some $m \in L^1_{\rm loc}((0,+\infty))$ there holds

$$W_2(\mu_s, \mu_t) \le \int_s^t m(y)\, dy, \qquad \forall s \le t.$$

We recall the following abstract result for absolutely continuous curves in probability spaces (see [AGS, Theorem 8.3.1]). If μ_t is absolutely continuous there exists a Borel vector field $\mathbf{v}_t \in L^2(X, \mu_t; X)$ with $\|v_t\|_{L^2} \in L^1_{\rm loc}(0,+\infty)$ such that the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \tag{3.5}$$

holds in the sense of distributions (i.e. in the duality with $\mathrm{Cyl}(X)$). The vector field $\mathbf{v}_t$ is not unique, but there exists the one of minimal L^2 norm, which can be thought as the velocity vector associated to the curve μ_t. In the sequel, $\mathbf{v}_t$ will always be understood as this optimal tangent vector.

The curve μ_t is then said to be the gradient flow of functional ϕ if

$$-\mathbf{v}_t \in \partial\phi(\mu_t), \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0. \tag{3.6}$$

If ϕ is convex along geodesics, recalling (3.3) the gradient flow equation can be equivalently rewritten as

$$-\int_X \langle \mathbf{v}_t(x), \mathbf{t}(x) - x \rangle\, d\mu_t \le \phi(\nu) - \phi(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0, \tag{3.7}$$

for all $\nu \in D(\phi)$. We can therefore think to the gradient flow equation itself as a system containing the subdifferential inclusion and the general continuity equation (3.5). We shall say that a gradient flow μ_t starts from $\bar\mu$ if $\mu_t \to \bar\mu$ as $t \downarrow 0$.

We also recall the following formula for the derivative of the Wasserstein distance along absolutely continuous curves (see [AGS, Theorem 8.4.7]). If $\mu_t \in AC_{\rm loc}((0,+\infty), \mathscr{P}_2(X))$ and $\nu \in \mathscr{P}_2(X)$ there holds

$$\frac{1}{2}\frac{d}{dt} W_2^2(\mu_t, \nu) = \int_{X \times X} \langle x - y, \mathbf{v}_t(x) \rangle\, d\gamma_t(x,y), \qquad \forall \gamma_t \in \Gamma_0(\mu_t, \nu).$$

Since the left hand side in (3.7) is (in our simplified setting where optimal transport maps do exist) the derivative of the squared Wasserstein distance, (3.7) is also equivalent to the following inequality:

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\nu) \le \phi(\nu) - \phi(\mu_t), \qquad \text{for } \mathcal{L}^1\text{-a.e. } t>0, \tag{3.8}$$

for all $\nu \in D(\phi)$.

We now recall the main result about gradient flows of convex functionals along geodesics (see [AGS, Theorem 11.2.1], also for a detailed discussion on the properties of flows).

Theorem 3.8 *Let $\phi : \mathscr{P}_2(X) \to (-\infty,+\infty]$ be a strongly convex functional. Then, for all $\bar\mu \in \overline{D(\phi)}$, there exists a unique gradient flow μ_t starting from $\bar\mu$, generating a contraction semigroup $S(t)$ on $\overline{D(\phi)}$. In addition μ_t satisfies* (3.8)*, belongs to $D(|\partial\phi|)$ for any $t>0$ and the map $t \mapsto |\partial\phi|(\mu_t)$ is nonincreasing.*

Remark 3.9 The solution provided by Theorem 3.8 satisfies the following additional properties.

i) For every $t>0$ there holds

$$\phi(\mu_t) \le \frac{1}{2t}W_2^2(\mu_0,\nu) + \phi(\nu) \quad \forall \nu \in D(\phi), \tag{3.9}$$

$$|\partial\phi|^2(\mu_t) \le |\partial\phi|^2(\nu) + \frac{1}{t^2}W_2^2(\mu_0,\nu) \quad \forall \nu \in D(|\partial\phi|). \tag{3.10}$$

ii) The following energy identity holds:

$$\int_a^b \int_X |\mathbf{v}_t|^2\, d\mu_t\, dt = \phi(\mu_a) - \phi(\mu_b) \qquad \forall\, 0 \le a < b < +\infty. \tag{3.11}$$

iii) If μ_0 is a minimum point for ϕ and $t>0$, then

$$|\partial\phi|(\mu_t) \le \frac{W_2(\bar\mu,\mu_0)}{t}, \qquad \phi(\mu_t) - \phi(\mu_0) \le \frac{W_2^2(\bar\mu,\mu_0)}{2t}$$

and the map $t \mapsto W_2(\mu_t,\mu_0)$ is nonincreasing.

3.2 The discrete scheme

The gradient flow μ_t of functional ϕ is the limit of the Euler discrete scheme: given $\bar\mu \in D(\phi)$, one constructs a sequence $(\mu_\tau^k) \subset \mathscr{P}_2(X)$, with $\mu_\tau^0 = \bar\mu$, whose k-th element is found minimizing the functional

$$\Phi_\tau(\nu,\mu_\tau^{k-1}) := \phi(\nu) + \frac{1}{2\tau}W_2^2(\nu,\mu_\tau^{k-1}). \tag{3.12}$$

For $t>0$ and $k>0$, we can define a discrete gradient flow as

$$\widetilde{\mu}_\tau(t) := \mu_\tau^k \quad \text{if } t \in ((k-1)\tau, k\tau].$$

In fact, under the same assumptions as Theorem 3.8, another consequence of [AGS, Theorem 11.2.1] is the following

Proposition 3.10 *If ϕ is strongly convex and $\mu_\tau^0 = \bar{\mu}$, we have $\widetilde{\mu}_\tau(t) \to \mu_t$ for all $t \geq 0$, where μ_t is the gradient flow provided by* Theorem 3.8.

Let us focus the attention on the discrete problem. We will give some more precise results.

Proposition 3.11 *Let ϕ be a strongly convex functional, let $\mu \in \overline{D(\phi)}$ and $\tau > 0$. Then $\Phi_\tau(\cdot, \mu)$ admits a unique minimizer $\mu_\tau \in \mathscr{P}_2(X)$.*

Proof. Define

$$\psi_\tau(\mu) := \inf_{\nu \in \mathscr{P}_2(X)} \left\{ \frac{1}{2\tau} W_2^2(\mu, \nu) + \phi(\nu) \right\}. \tag{3.13}$$

ψ_τ depends continuously on μ (with respect to the W_2 convergence), as shown in [AGS, Lemma 3.1.2] in a more general framework. Suppose now that (ν_n) is a minimizing sequence for $\Phi_\tau(\cdot, \mu)$. Then, since $\mu \in \overline{D(\phi)}$, there exists a sequence $(\mu_n) \subset D(\phi)$ converging to μ in $\mathscr{P}_2(X)$ such that

$$\limsup_{n\to\infty} \Phi_\tau(\nu_n, \mu_n) = \limsup_{n\to\infty} \Phi_\tau(\nu_n, \mu) \leq \psi_\tau(\mu). \tag{3.14}$$

Let us now take advantage of (3.4), choosing a continuous curve μ_t, $t \in [0, 1]$, connecting ν_n to ν_m, with μ_n as a base point. We obtain

$$\begin{aligned} \Phi_\tau(\mu_{1/2}, \mu_n) &= \phi(\mu_{1/2}) + \frac{1}{2\tau} W_2^2(\mu_{1/2}, \mu_n) \\ &\leq \frac{1}{2} \left(\phi(\nu_n) + \frac{1}{2\tau} W_2^2(\nu_n, \mu_n) \right) + \frac{1}{2} \left(\phi(\nu_m) + \frac{1}{2\tau} W_2^2(\nu_m, \mu_n) \right) - \frac{1}{8\tau} W_2^2(\nu_n, \nu_m). \end{aligned} \tag{3.15}$$

Now notice that the left hand side can be bounded from below with $\psi_\tau(\mu_n)$, while the first two terms in the right hand side are asymptotically smaller than $\frac{1}{2}\psi_\tau(\mu)$, by (3.14). We conclude that $W_2^2(\nu_n, \nu_m) \to 0$ as n, $m \to \infty$, whence $\nu_n \to \nu$ in $\mathscr{P}_2(X)$. Then, it follows easily that ν is a minimizer. Since the minimizing sequence was chosen arbitrarily, we also conclude that ν is the unique minimizer. □

Remark 3.12 Let $\mu_\tau \in D(\phi)$ be a minimizer over $\mathscr{P}_2(X)$ of $\Phi_\tau(\cdot, \mu)$, where Φ is the functional defined in (3.12), and assume that $\mathbf{t}_\tau$ is an optimal transport map between μ_τ and μ. Then (see Lemma 10.1.2 in [AGS]) one can construct a vector $\boldsymbol{\omega}_\tau \in \partial\phi(\mu_\tau)$ by

$$\boldsymbol{\omega}_\tau = \frac{\mathbf{t}_\tau - \mathbf{I}}{\tau}. \tag{3.16}$$

The following approximation result of the minimal selection in terms of vectors as in (3.16) will be useful in the sequel.

Lemma 3.13 *Let $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be a convex functional along geodesics and let $\mu \in D(|\partial\phi|)$. If μ_τ is a minimizer of $\Phi_\tau(\cdot, \mu)$ and $\boldsymbol{\omega}_\tau$ is constructed as in* (3.16)*, then there exist $\tau_n \to 0$ such that, as $n \to \infty$, $\mu_{\tau_n} \to \mu$, $\phi(\mu_{\tau_n}) \to \phi(\mu)$ and, more precisely,*

$$|\partial\phi|^2(\mu) = \lim_{n\to\infty} \frac{\phi(\mu) - \phi(\mu_{\tau_n})}{\tau_n} = \lim_{n\to\infty} \|\boldsymbol{\omega}_{\tau_n}\|^2_{L^2(X,\, \mu_{\tau_n};\, X)}\,. \tag{3.17}$$

Moreover, $\boldsymbol{\omega}_{\tau_n} \in L^2(X, \mu_{\tau_n}; X)$ *converge, in the sense of* Definition 2.3*, to the unique vector* $\partial^0\phi(\mu)$ *with minimal norm in* $\partial\phi(\mu)$*.*

Proof. The proof follows from Lemma 10.3.10 and 10.3.11 in [AGS], simply reducing the notation therein in terms of standard vector subdifferentials in $\mathscr{P}_2(X)$. $\square$

4 Internal energy functional

Given a Borel probability measure γ on $\mathbb{R}^d$, we define the finite-dimensional internal energy functional relative to γ as follows:

$$\mathscr{F}_d(\mu|\gamma) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} F\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise .} \end{cases}$$

The definition can be extended easily to the case of a Borel probability measure γ in an infinite-dimensional Hilbert space X:

$$\mathscr{F}(\mu|\gamma) = \begin{cases} \displaystyle\int_{X} F\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise .} \end{cases}$$

Assumption 4.1 *We consider the following assumptions on the integrand* $F : [0, +\infty) \to (-\infty, +\infty]$*:*

i) *F is strictly convex;*

ii) *the map* $s \mapsto e^s F(e^{-s})$ *is convex and nonincreasing in* $\mathbb{R}$*;*

iii) $F(0) = 0$*;*

iv) *F has a superlinear growth at infinity.*

Condition *ii)* is needed for the geodesic convexity of $\mathscr{F}$, and in fact it has been introduced in [AGS] as a dimension-free extension of the one introduced by McCann (see [Mc]) for the d-dimensional case, namely

$$x \mapsto x^d F(x^{-d}) \text{ is convex and nonincreasing in } (0, +\infty). \tag{4.1}$$

Indeed, it can be shown that *ii)* implies (4.1). It is convenient to introduce the continuous function

$$L_F(x) := xF'_+(x) - F(x), \tag{4.2}$$

where F'_+ denotes the right derivative. In fact, we will write the velocity vector field of the gradient flow of $\mathscr{F}_d$ and $\mathscr{F}$ in terms of L_F, which will indeed be the same function L as equation

(1.2). Notice also that the monotonicity condition in *(ii)* is equivalent to $xL_F'(x) - L_F(x) \geq 0$, while the convexity condition yields

$$e^s F(e^{-s}) - F'(e^{-s}) + e^{-s}F''(e^{-s}) \geq 0,$$

which implies convexity of F.

Let us introduce (see [AGS, Lemma 9.4.4]) the following dual representation of $\mathscr{F}$:

$$\mathscr{F}(\mu|\gamma) = \sup\left\{\int_X g(x)d\mu(x) - \int_X F^*(g(x))d\gamma(x) : g \in C_b^0(X)\right\}, \tag{4.3}$$

where F^* denotes the Fenchel conjugate of F. We notice that from (4.3) $\mathscr{F}$ is sequentially l.s.c. with respect to the weak convergence. For μ, $\nu \in \mathscr{P}_2(X)$, we also introduce the notation

$$\Phi_\tau^{\mathscr{F}}(\nu, \mu) := \mathscr{F}(\nu|\gamma) + \frac{1}{2\tau}W_2^2(\nu, \mu). \tag{4.4}$$

The typical example of function F one can consider is the n-th power:

$$F(x) = \frac{x^n}{n-1}, \quad n > 1, \tag{4.5}$$

with $L_F(x) = x^n$. Another important example is $F(x) = x \log x$, corresponding to the relative entropy functional (see Remark 4.7 below), whose gradient flow is a linear Fokker-Planck equation (see [JKO] and the infinite-dimensional theory in [ASZ]).

4.1 Geodesic convexity of $\mathscr{F}$

In this subsection we recall some results on the convexity properties of $\mathscr{F}$.

Definition 4.2 (Log-concavity) *A probability measure on X is said to be log-concave if, for any couple of open sets A, B in X, there holds*

$$\log \gamma((1-t)A + tB) \geq (1-t)\log\gamma(A) + t\log\gamma(B). \tag{4.6}$$

If $X = \mathbb{R}^d$ and γ is non-degenerate (i.e. it is not supported in a proper subspace of X), then Borell proved (see also [AGS, Theorem 9.4.10]) that γ is log-concave if and only if $\gamma = e^{-V}\mathscr{L}^d$ for some lower semicontinuous and convex function $V : \mathbb{R}^d \to (-\infty, +\infty]$ whose domain has nonempty interior.

For the internal energy functional relative to γ, convexity along geodesics is strictly related to the log-concavity of γ, as shown by the following result (see [AGS, Theorem 9.4.12]).

Theorem 4.3 *Let F be satisfying* Assumption 4.1 *ii)-iii)-iv), and suppose that γ is log-concave. Then $\mathscr{F}(\cdot|\gamma)$ is strongly convex in $\mathscr{P}_2(X)$.*

Remark 4.4 Let $\gamma \in \mathscr{P}(X)$ be log-concave, let $\mu \in \mathscr{P}_2(X)$ and consider the constrained minimization problem

$$\min_{\nu \leq M\gamma} \Phi_\tau^{\mathscr{F}}(\nu, \mu).$$

Then this problem admits a unique minimizer, as the unconstrained one. In fact, the functional

$$\mathscr{F}^M(\mu|\gamma) := \begin{cases} \displaystyle\int_X F^M\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.7}$$

where

$$F^M(z) := \begin{cases} +\infty & \text{if } z > M, \\ F(z) & \text{otherwise,} \end{cases}$$

trivially satisfies the hypotheses of Theorem 4.3, so it is strongly convex and we can apply Proposition 3.11 with $\phi = \mathscr{F}^M$.

4.2 Bounded densities

The following result extends the one of [A, Section 2.1] to the infinite dimensional case, basically with the same proof.

Lemma 4.5 *Let F satisfy* Assumption 4.1 *and suppose that γ is log-concave. Let $\mu = \rho\gamma \in \mathscr{P}_2(X)$, with $\rho \leq M$ γ-a.e. in X. Then there exists a unique minimizer μ_τ of $\Phi_\tau^{\mathscr{F}}(\cdot, \mu)$, and $\mu_\tau \leq M\gamma$.*

Proof. We assume without loss of generality that M is a point of differentiability for F. As a first step, we consider the problem of minimizing $\Phi_\tau^{\mathscr{F}}(\cdot, \mu)$ under the constraint $\nu \leq M'\gamma$, where $M' \geq M$. In view of Remark 4.4, we know that in this case there exists a unique minimizer $\overline{\mu}_\tau = \overline{\rho}_\tau\gamma \leq M'\gamma$.

Let β denote the optimal transport plan between μ and $\overline{\mu}_\tau$. Suppose by contradiction that $\overline{\rho}_\tau > M$ on some Borel set $\Omega \subset X$ with $\gamma(\Omega) > 0$ and let Ω^c be the complement of Ω in X.

Now let $\beta_\Omega = \chi_{\Omega^c \times \Omega}\beta$. It is clear that $\pi^1_\# \beta_\Omega \leq \mu$ and $\pi^2_\# \beta_\Omega \leq \overline{\mu}_\tau$. Then, letting $\tilde{\rho}$ and $\tilde{\rho}_\tau$ be the densities with respect to γ of the first and second marginal of β_Ω, we have

$$\tilde{\rho} \leq \rho \quad \text{and} \quad \tilde{\rho}_\tau \leq \overline{\rho}_\tau. \tag{4.8}$$

Moreover, the following properties are easily seen to hold γ-a.e.:

$$\tilde{\rho} \leq M, \qquad \tilde{\rho} = 0 \ \text{ on } \Omega, \qquad \tilde{\rho}_\tau = 0 \ \text{ on } \Omega^c. \tag{4.9}$$

Let us introduce the competitor of μ_τ as

$$\rho_\tau^\varepsilon\gamma := (\overline{\rho}_\tau + \varepsilon(\tilde{\rho} - \tilde{\rho}_\tau))\,\gamma. \tag{4.10}$$

By the definition of $\tilde\rho$ and $\tilde\rho_\tau$ it is immediate to check that $\int_X \tilde\rho\, d\gamma = \int_X \tilde\rho_\tau\, d\gamma = \beta(\Omega^c \times \Omega)$. As a consequence $\rho^\varepsilon_\tau \gamma \in \mathscr{P}_2(X)$. Moreover, since $\overline{\rho}_\tau > M$ γ-a.e. in Ω, making use of (4.8) and (4.9) we obtain, for small enough ε,

$$\rho^\varepsilon_\tau = \overline{\rho}_\tau - \varepsilon\tilde\rho_\tau > 0 \qquad \gamma\text{-a.e. on } \Omega. \tag{4.11}$$

Then, denoting by F'_- and F'_+ respectively the left and right derivative of F, thanks to the convexity of F we have, for small enough ε,

$$\begin{aligned}
\int_X \left(F(\rho^\varepsilon_\tau) - F(\overline{\rho}_\tau)\right)\, d\gamma &\le \int_{\Omega^c} \left(F(\overline{\rho}_\tau + \varepsilon\tilde\rho) - F(\overline{\rho}_\tau)\right)\, d\gamma + \int_\Omega \left(F(\overline{\rho}_\tau - \varepsilon\tilde\rho_\tau) - F(\overline{\rho}_\tau)\right)\, d\gamma \\
&\le \varepsilon \int_{\Omega^c} F'_+(\overline{\rho}_\tau + \varepsilon\tilde\rho)\tilde\rho\, d\gamma - \varepsilon \int_\Omega F'_-(\overline{\rho}_\tau - \varepsilon\tilde\rho_\tau)\tilde\rho_\tau\, d\gamma \\
&\le \varepsilon \int_{\Omega^c} F'_+(M + \varepsilon\tilde\rho)\tilde\rho\, d\gamma - \varepsilon \int_\Omega F'_-(M - \varepsilon\tilde\rho_\tau)\tilde\rho_\tau\, d\gamma \\
&= \varepsilon \int_{X\times X} \left[F'_+(M + \varepsilon\tilde\rho(x)) - F'_-(M - \varepsilon\tilde\rho_\tau(y))\right]\, d\beta_\Omega(x,y) \\
&= \varepsilon \int_{X\times X} o(1)\, d\beta_\Omega(x,y).
\end{aligned}$$

Since $\tilde\rho$ and $\tilde\rho_\tau$ are bounded above γ-a.e. by M', we conclude that

$$\int_X \left(F(\rho^\varepsilon_\tau) - F(\overline{\rho}_\tau)\right)\, d\gamma \le o(\varepsilon). \tag{4.12}$$

On the other hand, let $\mathbf{t} : X \times X \to X \times X$ be defined by $\mathbf{t}(x,y) = (x,x)$, and let

$$\beta_\varepsilon = \beta - \varepsilon\beta_\Omega + \varepsilon\mathbf{t}_\#\beta_\Omega.$$

By the composition rule of the push forward we have $\pi^2_\#\mathbf{t}_\#\beta_\Omega = (\pi^2 \circ \mathbf{t})_\#\beta_\Omega = \pi^1_\#\beta_\Omega$, so that the second marginal of $\mathbf{t}_\#\beta_\Omega$ is equal to the first marginal of β_Ω, namely $\tilde\rho$; analogously the first marginal of $\mathbf{t}_\#\beta_\Omega$ coincides with the first marginal of β_Ω. Hence it is clear that $\beta_\varepsilon \in \Gamma(\mu, \rho^\varepsilon_\tau\gamma)$. So we can estimate

$$W_2^2(\rho^\varepsilon_\tau\gamma, \mu) - W_2^2(\rho_\tau\gamma, \mu) \le \int_{X\times X} |x-y|^2\, d(\beta_\varepsilon - \beta)(x,y) = -\varepsilon \int_{\Omega^c\times\Omega} |x-y|^2\, d\beta(x,y). \tag{4.13}$$

Together with (4.12), this gives

$$\Phi^{\mathscr{F}}_\tau(\rho^\varepsilon_\tau\gamma, \mu) - \Phi^{\mathscr{F}}_\tau(\overline{\rho}_\tau\gamma, \mu) \le -\frac{\varepsilon}{2\tau} \int_{\Omega^c\times\Omega} |x-y|^2\, d\beta(x,y) + o(\varepsilon). \tag{4.14}$$

But consider that

$$\beta(\Omega\times\Omega) \le \beta(\Omega\times X) = \int_X \chi_\Omega(x)\, d(\pi^1_\#\beta)(x) = \int_\Omega \rho(x)\, d\gamma(x) \le M\gamma(\Omega). \tag{4.15}$$

This forces $\beta(\Omega^c \times \Omega)$ to be strictly positive, otherwise

$$\beta(\Omega \times \Omega) = \beta(X \times \Omega) = \overline{\mu}_\tau(\Omega) = \int_\Omega \overline{\rho}_\tau(x)\, d\gamma(x) > M\gamma(\Omega)$$

against (4.15). Back to (4.14), if ε is chosen small enough, we contradict the minimality of $\overline{\mu}_\tau = \overline{\rho}_\tau \gamma$. We have proved that $\overline{\rho}_\tau \leq M$, independently of the initial choice of M'.

Since these properties hold for all $M' > M$, it turns out that the minimizer is independent of M', hence $\overline{\mu}_\tau$ is a minimizer under the constraint $\nu = \rho\gamma$ with $\rho \in L^\infty(\gamma)$. Then, a simple truncation argument provides the minimality of $\overline{\mu}_\tau$ in the unconstrained problem. □

4.3 The finite-dimensional case

A key ingredient of our analysis will be the finite-dimensional framework, which has been studied in detail in [AGS, Chapter 10]. We now recall the main result therein (see [AGS, Theorem 10.4.9]).

Theorem 4.6 *Let $\gamma = e^{-V}\mathcal{L}^d$ be a non-degenerate log-concave probability measure on $\mathbb{R}^d$, let Ω be the nonempty interior of $D(V)$ and consider the functional $\mathscr{F}_d(\cdot|\gamma)$ and $\mu = \rho\gamma \in D(\mathscr{F}_d)$. Then $\rho \in D(|\partial\mathscr{F}_d|)$ if and only if*

$$L_F \circ \rho \in W^{1,1}(\Omega) \qquad \text{and} \qquad \frac{\nabla(L_F \circ \rho)}{\rho} \in L^2(\mathbb{R}^d, \mu). \tag{4.16}$$

If these conditions hold, $\frac{\nabla(L_F\circ\rho)}{\rho}$ realizes the minimal selection in $|\partial\mathscr{F}_d(\cdot|\gamma)|$ at the point μ, so that

$$\frac{\nabla(L_F \circ \rho)}{\rho} = \partial^0 \mathscr{F}_d(\mu|\gamma) \qquad \text{and} \qquad \left\| \frac{\nabla(L_F \circ \rho)}{\rho} \right\|_{L^2(\mathbb{R}^d,\mu)} = |\partial\mathscr{F}_d(\mu|\gamma)|. \tag{4.17}$$

From this characterization of the Wasserstein subdifferential of $\mathscr{F}_d$, we learn that the gradient flow of $\mathscr{F}_d$ satisfies equation (1.2) in its finite dimensional version.

Remark 4.7 In the case $X = \mathbb{R}^d$, let $\gamma = e^{-V}\mathcal{L}^d$, where V is a convex l.s.c. potential, and $\mu_t = u_t\mathcal{L}^d$. As a consequence we have $\rho_t = u_t e^V$ and (1.2) becomes

$$\partial_t u_t - \nabla \cdot (\nabla(L_F \circ u_t) + u_t \nabla V) = 0. \tag{4.18}$$

In (4.18) we recognize different PDEs. In particular, if $V = 0$ and $L_F(x) = x^m$, $m > 1$ (which corresponds to $F = (m-1)^{-1}x^m$) we obtain the porous media equations. If $F(x) = x\log x$, then $\mathscr{F}_d$ is the well known entropy functional

$$\mathscr{H}_d(\mu|\gamma) = \begin{cases} \displaystyle\int_X \left(\frac{d\mu}{d\gamma}\right) \log\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise}. \end{cases} \tag{4.19}$$

In this case $L_F(x) = x$ and (4.18) becomes the linear Fokker Planck equation with potential V:

$$\partial_t u_t - \Delta u_t - \nabla \cdot (u_t \nabla V) = 0. \tag{4.20}$$

See [ASZ] for a detailed comparison between the different approaches to (4.20) in infinite dimensions.

4.4 Γ-convergence results

For the characterization of the subdifferential of $\mathscr{F}$, we will perform finite dimensional approximations, and we need a Γ-convergence result. First of all, we introduce the following:

Definition 4.8 (Γ-convergence) *We say that $\phi_n : \mathscr{P}_2(X) \to [-\infty, +\infty]$ $\Gamma(\mathscr{P}_2(X))$-converge to ϕ if*

i) for any sequence $(\mu_n) \subset \mathscr{P}_2(X)$ weakly convergent to μ, there holds

$$\phi(\mu) \le \liminf_{n\to\infty} \phi_n(\mu_n); \tag{4.21}$$

ii) for any $\mu \in \mathscr{P}_2(X)$ there exists $(\mu_n) \subset \mathscr{P}_2(X)$ converging to μ in $\mathscr{P}_2(X)$ such that

$$\lim_{n\to\infty} \phi_n(\mu_n) = \phi(\mu). \tag{4.22}$$

Γ-convergence guarantees the convergence of minimizers to minimizers, as in the next lemma.

Lemma 4.9 *Let $\phi_h : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be geodesically convex functionals satisfying* Assumption 3.1 *and $\Gamma(\mathscr{P}_2(X))$-convergent to ϕ, still satisfying* Assumption 3.1*. Assume also that for all $M > 0$ the set*

$$\bigcup_{h=1}^{\infty} \{\mu \in \mathscr{P}_2(X) : \ \phi_h(\mu) \le M\} \tag{4.23}$$

is relatively compact in the weak topology of $\mathscr{P}(X)$.
Then, for τ fixed, (μ^h_τ) has limit points in $\mathscr{P}_2(X)$ and $\boldsymbol{\omega}^h_\tau \in \partial\phi_h(\mu^h_\tau)$, constructed as in Remark 3.12*, have strong limit points in the sense of* Definition 2.3*. If (h_n) is any subsequence along which we have convergence, and $\mu_\tau, \boldsymbol{\omega}_\tau$ are the limits, then μ_τ is a minimizer of $\Phi_\tau(\cdot, \mu)$ and $\boldsymbol{\omega}_\tau$ belongs to $\partial\phi(\mu_\tau)$. Moreover*

$$\phi_{h_n}(\mu^{h_n}_\tau) \to \phi(\mu_\tau).$$

Proof. It follows from [AGS, Lemma 10.3.17], which is stated for the general $\mathscr{P}_p(X)$ case. The case $p = 2$ and the fact that we consider maps instead of plans leads to considerable simplifications of that proof, see also [ASZ]. In the proof of Lemma 10.3.17 the equi-tightness assumption (ensuring compactness of minimizing sequences) is not present, and replaced by a stronger lower semicontinuity property than (4.21), involving duality with cylindrical functions. □

It is clear from Lemma 3.13 that there exist μ_{τ_n}, minimizers of $\Phi_{\tau_n}(\cdot, \mu)$, such that the respective subdifferentials converge to $\partial^0\phi(\mu)$. With the next result we want to show that the approximating $\boldsymbol{\omega}_n$ can also be chosen to be subdifferentials of functionals ϕ_n, if ϕ_n is Γ-convergent to ϕ. We include the proof, following with minor variants [AGS, Lemma 10.3.16].

Theorem 4.10 *Let $\phi_n, \phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ be as in* Lemma 4.9. *Then, for every $\mu \in D(|\partial\phi|)$ there exist a subsequence $n(m)$, $\mu_{n(m)}$ converging to μ in $\mathscr{P}_2(X)$ and subdifferentials $\boldsymbol{\omega}_{n(m)} \in \partial\phi_{n(m)}(\mu_{n(m)})$ such that*

$$\boldsymbol{\omega}_{n(m)} \to \partial^0\phi(\mu) \in L^2(X, \mu; X) \quad \textit{strongly in } L^2 \textit{ as in } \text{Definition 2.3} \tag{4.24}$$

and

$$\lim_{m\to\infty} \phi_{n(m)}(\mu_{n(m)}) = \phi(\mu). \tag{4.25}$$

In particular, since $|\partial\phi|(\mu)$ is the $L^2(\mu)$ norm of the minimal selection in $\partial\phi(\mu)$, this means that

$$\limsup_{m\to\infty} \int_X \|\boldsymbol{\omega}_{n(m)}\|^2 d\mu_{n(m)} \leq |\partial\phi|^2(\mu). \tag{4.26}$$

Proof. We construct the approximating sequence in the following way. Let $\mu^h \to \mu$ in $\mathscr{P}_2(X)$ with $\phi_h(\mu^h) \to \phi(\mu)$ (such a sequence exists by Γ-convergence). Let μ_τ^h be a minimizer of

$$\Phi_\tau^h(\cdot, \mu^h) := \phi_h(\cdot) + \frac{1}{2\tau} W_2^2(\cdot, \mu^h).$$

Let moreover $\boldsymbol{\omega}_\tau^h$ be constructed as in Remark 3.12. We will show that there is a subsequence of the family $\{\boldsymbol{\omega}_\tau^h : h \in \mathbb{N}, \tau > 0\}$ such that (4.24) holds. First, for fixed τ, we know from Lemma 4.9 that there is a subsequence $\mu_\tau^{h_n}$ converging in $\mathscr{P}_2(X)$ to μ_τ, where μ_τ minimizes $\Phi_\tau(\cdot, \mu)$. Moreover, the corresponding sequence $\boldsymbol{\omega}_\tau^{h_n}$ converge to $\boldsymbol{\omega}_\tau \in \partial\phi(\mu_\tau)$ in the sense of Definition 2.3. Hence, given $\varepsilon > 0$, for n large enough we have

$$\left| \int_X |\boldsymbol{\omega}_\tau^{h_n}|^2 \, d\mu_\tau^{h_n} - \int_X |\boldsymbol{\omega}_\tau|^2 \, d\mu_\tau \right| < \frac{\varepsilon}{2}. \tag{4.27}$$

and (taking Lemma 4.9 into account)

$$\left| \phi_{h_n}(\mu_\tau^{h_n}) - \phi(\mu_\tau) \right| < \frac{\varepsilon}{2}. \tag{4.28}$$

On the other hand, we know from Lemma 3.13 that there exists an infinitesimal sequence (τ_m) such that

$$\lim_{m\to\infty} \left| \int_X |\boldsymbol{\omega}_{\tau_m}|^2 \, d\mu_{\tau_m} - \int_X |\boldsymbol{\omega}|^2 \, d\mu \right| = 0 \tag{4.29}$$

and

$$\lim_{m\to\infty} |\phi(\mu_{\tau_m}) - \phi(\mu)| = 0. \tag{4.30}$$

Now, with $\tau = \tau_m$ and $\varepsilon = 1/m$ we can suitably choose $h_n = h_n(m)$ in (4.27) and (4.28) to conclude with a diagonal argument. □

Now we state the particular Γ-convergence result for our functionals.

Theorem 4.11 *If γ_n converge weakly to γ, then $\mathscr{F}(\cdot|\gamma_n)$ $\Gamma(\mathscr{P}_2(X))$-converge to $\mathscr{F}(\cdot|\gamma)$ and satisfy the equi-tightness condition* (4.23)*. Moreover, if $\mu \in \mathscr{P}_2(X)$ and $\gamma_n = \pi^n_\# \gamma$, a sequence satisfying condition* (4.22) *is $\pi^n_\# \mu$, so that*

$$\lim_{n\to\infty} \mathscr{F}(\pi^n_\# \mu|\gamma_n) = \mathscr{F}(\mu|\gamma).$$

Proof. We first prove the equi-tightness condition (4.23). Fix $\varepsilon > 0$ and two constants M', M'' large enough such that $M/M' < \varepsilon/2$ and $F(x) > M'x$ for $x > M''$ (this is possible in view of the superlinear growth of F at infinity). Let moreover K_ε be a compact subset of X such that $\gamma_n(K_\varepsilon) > 1 - \frac{\varepsilon}{2M''}$ for every n (the sequence (γ_n) is tight, since it is weakly convergent). If $\mu \in \mathscr{P}_2(X)$ satisfies $\mathscr{F}(\mu|\gamma_n) \le M$ for some n, we have

$$\begin{aligned}\mu(X \setminus K_\varepsilon) = \frac{1}{M'} \int_{X\setminus K_\varepsilon} M' \, d\mu &< \frac{1}{M'} \int_{(X\setminus K_\varepsilon)\cap\{\rho > M''\}} \frac{F(\rho)}{\rho} \, d\mu + \int_{(X\setminus K_\varepsilon)\cap\{\rho\le M''\}} M'' \, d\gamma_n \\ &\le \frac{M}{M'} + M''\gamma_n(X \setminus K_\varepsilon) < \varepsilon.\end{aligned}$$

This shows that the set introduced in (4.23) is tight, hence relatively compact.

In order to prove $\Gamma-$convergence, let $\mu_n \rightharpoonup \mu$. For any $g \in C^0_b(X)$ there holds

$$\int_X g(x)d\mu(x) - \int_X F^*(g(x))d\gamma(x) = \lim_{n\to\infty} \left(\int_X g(x)\, d\mu_n - \int_X F^*(g(x))\, d\gamma_n \right) \le \liminf_{n\to\infty} \mathscr{F}(\mu_n|\gamma_n), \tag{4.31}$$

so that

$$\sup_{\mu\in C^0_b(X)} \left(\int_X g(x)d\mu(x) - \int_X F^*(g(x))d\gamma(x) \right) \le \liminf_{n\to\infty} \mathscr{F}(\mu_n|\gamma_n).$$

Taking into account the duality formula (4.3), the liminf inequality *i)* of the definition of Γ-convergence follows. The limsup inequality *ii)* and the last statement are proven exactly as in [ASZ, Lemma 6.2]. $\square$

Now consider finite dimensional approximations of the measure γ: letting $\gamma_n = \pi^n_\# \gamma$, from Theorem 4.11 we know that $\mathscr{F}(\cdot|\gamma_n)$ Γ-converge to $\mathscr{F}(\cdot|\gamma)$. From the next result it will follow that, if the role of Γ-converging functionals of Theorem 4.10 is played by $\mathscr{F}(\cdot|\gamma_n)$ and we choose a limit point $\mu \in L^\infty(X,\gamma)$, then the approximating μ_n can be chosen so that their densities have uniformly bounded $L^\infty(X,\gamma_n)$ norms.

Corollary 4.12 *For all μ with $\mu \le M\gamma$ and $|\partial\mathscr{F}(\mu|\gamma)|$ finite, there exist μ_n with $\mu_n \le M\gamma_n$, $\mu_n \to \mu$ in $\mathscr{P}_2(X)$, $\mathscr{F}(\mu_n|\gamma_n) \to \mathscr{F}(\mu|\gamma)$. In addition, there exist $\boldsymbol{\omega}_n \in \partial\mathscr{F}(\cdot|\gamma_n)(\mu_n)$ such that*

$$\boldsymbol{\omega}_n \to \partial^0\mathscr{F}(\cdot|\gamma)(\mu) \in L^2(X,\mu;X) \quad \textit{strongly in the sense of } \text{Definition 2.3.} \tag{4.32}$$

Proof. It suffices to revisit in this particular case the proof of Theorem 4.10: first, the measures μ_n satisfy $\mu_n \le M\gamma_n$ by Theorem 4.11; second, the minimizers of the problems

$$\nu \mapsto \mathscr{F}(\nu|\gamma_n) + \frac{1}{2\tau} W_2^2(\nu,\mu_n)$$

satisfy $\mu^n_\tau \le M\gamma_n$ by Lemma 4.5. $\square$

5 Wasserstein subdifferential of $\mathscr{F}$

We will now characterize the subdifferential of $\mathscr{F}$. In this section we make Assumption 2.5 on γ, besides the log-concavity.

In the sequel we are using the stability of generalized Sobolev spaces under composition with L_F, namely $\rho \in GW^{1,1}(X,\gamma)$ implies $L_F \circ \rho \in GW^{1,1}(X,g)$. Indeed, since $L_F(z) \to +\infty$ as $z \to +\infty$ and L_F is strictly increasing, we have

$$T_\alpha(L_F \circ \rho) = L_F \circ T_{L_F^{-1}(\alpha)}(\rho), \tag{5.1}$$

(here T_α is the truncation operator) and since $T_\beta(\rho) \in W^{1,1}(X,\gamma)$ for any $\beta > 0$ we conclude that $L_F \circ \rho \in GW^{1,1}(X,\gamma)$ thanks to the chain rule.

We begin by giving the following:

Definition 5.1 (Generalized Fisher information) *Let $\rho \in L^\infty(X,\gamma)$ and $\rho \in W^{1,1}(X,\gamma)$. Assume that*

$$\sum_{j=1}^{\infty} \int_X \left| \frac{\partial_{\mathbf{e}_j}(L_F \circ \rho)}{\rho} \right|^2 d\mu(x) < +\infty. \tag{5.2}$$

We define the generalized Fisher information functional as follows:

$$\mathscr{G}(\rho\gamma|\gamma) := \left\| \frac{\nabla(L_F \circ \rho)}{\rho} \right\|^2_{L^2(X,\mu;X)}.$$

In the general case $\rho \in L^1(X,\gamma)$, $\rho \in GW^{1,1}(X,\gamma)$, the generalized Fisher information is defined by the same formula, using the fact that $L_F \circ \rho \in GW^{1,1}(X,\gamma)$, so its gradient is still well defined.

Lemma 5.2 (Lower semicontinuity of $\mathscr{G}$) *Let $(\rho_n) \subset W^{1,1}(X,\gamma)$, with $\rho_n \to \rho$ γ-a.e. and with $\mathscr{G}(\rho_n\gamma|\gamma)$ uniformly bounded. Then $\rho \in GW^{1,1}(X,\gamma)$ and*

$$\mathscr{G}(\rho\gamma|\gamma) \le \liminf_{n\to\infty} \mathscr{G}(\rho_n\gamma|\gamma).$$

Proof. We set $\rho_{n,k} := T_k(\rho_n)$. By dominated convergence, it is clear that $\rho_{n,k} \to T_k(\rho)$ in $L^2(X,\gamma)$ and that

$$L_F \circ \rho_{n,k} \to L_F \circ T_k(\rho) \quad \text{in } L^2(X,\gamma). \tag{5.3}$$

By the chain rule proven in Theorem 2.7, $\nabla(L_F \circ \rho_{n,k})$ is equal to $L'_F(\rho_{n,k})\nabla\rho_{n,k}$, so it vanishes where $\rho_n > k$ and coincides with $\nabla(L_F \circ \rho_n)$ where $\rho_n \le k$. As a consequence, there holds

$$\int \frac{\|\nabla(L_F \circ \rho_{n,k})\|^2}{\rho_{n,k}}\, d\gamma \le \int \frac{\|\nabla(L_F \circ \rho_n)\|^2}{\rho_n}\, d\gamma, \tag{5.4}$$

where the second term is uniformly bounded by hypothesis. In particular, $L_F \circ \rho_{n,k}$ is bounded in $W^{1,2}(X,\gamma)$ and therefore $L_F \circ T_k(\rho) \in W^{1,2}(X,\gamma)$. Since k is arbitrary, we can use $L_F(k)$ as truncation levels to prove that $L_F \circ \rho \in GW^{1,2}(X,\gamma)$; in addition, $\nabla(L_F \circ \rho_{n,k})$ weakly converge in $L^2(X,\gamma;X)$ to $\nabla(L_F \circ T_k(\rho))$.

We can take advantage of Ioffe's lower semicontinuity Theorem under strong-weak convergence (see for instance [AFP, Theorem 5.8]) to obtain

$$\int \frac{\|\nabla(L_F \circ T_k(\rho))\|^2}{T_k(\rho)}\, d\gamma \le \liminf_{n\to\infty} \int \frac{\|\nabla(L_F \circ \rho_{n,k})\|^2}{\rho_{n,k}}\, d\gamma. \tag{5.5}$$

This, in combination with (5.4), gives

$$\int \frac{\|\nabla(L_F \circ T_k(\rho))\|^2}{T_k(\rho)}\, d\gamma \le \liminf_{n\to\infty} \int \frac{\|\nabla(L_F \circ \rho_n)\|^2}{\rho_n}\, d\gamma.$$

To conclude, it suffices to show that the left hand side converges to $\mathscr{G}(\rho\gamma|\gamma)$ as $k \to \infty$. To this aim, it suffices to remind that $\nabla(L_F \circ T_k(\rho))$ vanishes where $\rho > k$ and coincides with $\nabla(L_F \circ \rho)$ where $\rho \le k$. □

We are ready for the result which identifies the Wasserstein subdifferential of $\mathscr{F}$.

Theorem 5.3 *Let $\mu = \rho\gamma \in \mathscr{P}_2(X)$, and assume that F satisfies* Assumption 4.1. *Then the metric slope of $\mathscr{F}(\cdot|\gamma)$ at μ is finite if and only if*

$$L_F \circ \rho \in GW^{1,1}(X,\gamma) \quad \textit{and} \quad \frac{\|\nabla(L_F \circ \rho)\|^2}{\rho} \in L^1(X,\gamma). \tag{5.6}$$

Moreover, in this case

$$\frac{\nabla(L_F \circ \rho)}{\rho} = \partial^0 \mathscr{F}(\mu|\gamma) \quad \textit{and} \quad \mathscr{G}(\mu|\gamma) = |\partial\mathscr{F}(\mu|\gamma)|^2.$$

Proof. **Step 1.** We prove that finiteness of slope at $\mu = \rho\gamma$ implies the regularity properties (5.6). First, assume $\rho \le M$, set $\phi(\nu) = \mathscr{F}(\nu|\gamma)$ and $\phi_d(\nu) = \mathscr{F}(\nu|\gamma_d)$ and recall that $\gamma_d \rightharpoonup \gamma$. Thanks to Theorem 4.11, ϕ_d $\Gamma(\mathscr{P}_2(X))$-converge to ϕ as $d \to \infty$. By Theorem 4.10 we can find sequences

$$\begin{aligned} &\mu_d \to \mu \quad \text{in } \mathscr{P}_2(X), \quad \phi_d(\mu_d) \to \phi(\mu) \\ &\boldsymbol{\omega}_d \in \partial\phi_d(\mu_d) \text{ such that } \boldsymbol{\omega}_d \to \boldsymbol{\omega} = \partial^0\phi(\mu) \text{ strongly in } L^2 \text{ as in Definition 2.3,} \end{aligned} \tag{5.7}$$

and thanks to (4.26) we have also that $|\partial\phi_d|(\mu_d)$ is finite and uniformly bounded in d. We can also choose μ_d so that the additional property $\mu_d \le M\gamma_d$ holds, by Corollary 4.12.

Since $\gamma_d \rightharpoonup \gamma$ and $\mu_d \to \mu$ in $\mathscr{P}_2(X)$, we have that $\rho_d \to \rho$ in the sense of Definition 2.3, in its scalar version. Together with (5.7), which guarantees convergence of the energies, this also implies, thanks to Lemma 2.4 and the strict convexity of F, that

$$\int_X \varphi(x) L_F \circ \rho_d(x)\, d\gamma_d(x) \to \int_X \varphi(x) L_F \circ \rho(x)\, d\gamma(x) \qquad \forall \varphi \in L^1(X,\gamma). \tag{5.8}$$

Indeed, (5.8) holds independently of the growth of L_F for all $\varphi \in C_b^0(X)$, as ρ and ρ_d are essentially bounded, uniformly with respect to d, and the same uniform bound allows to extend the validity of the formula to all $\varphi \in L^1(X,\gamma)$.

The theorem holds if X is finite-dimensional, and since γ_d is supported in $\pi^d(X)$ we can use the implication in finite dimension (Theorem 4.6) to obtain, for $\zeta \in \mathrm{Cyl}(X)$, $j \le d$ and d large enough (depending on ζ only),

$$\begin{aligned}\int_X \partial_{\mathbf{e}_j}\zeta(x) L_F\circ\rho_d(x)\,d\gamma_d(x) = &-\int_X \partial_{\mathbf{e}_j}(L_F\circ\rho_d)(x)\zeta(x)\,d\gamma_d(x)\\ &+\int_X L_F\circ\rho_d(x)\zeta(x)g_d^j(x)\,d\gamma_d(x),\end{aligned} \tag{5.9}$$

where we used also the fact that $\partial_{\mathbf{e}_j}\gamma = g^j\gamma$ implies $\partial_{\mathbf{e}_j}\gamma_d = g_d^j\gamma_d$, g_d^j being the cylindrical projection of g^j (see Definition 2.2). The finite dimensional result also tells us that

$$\omega_d^j := \frac{\partial_{\mathbf{e}_j}(L_F\circ\rho_d)}{\rho_d} \in L^2(X,\mu_d), \qquad j = 1,\dots,d,$$

so we can rewrite (5.9) as

$$\begin{aligned}\int_X \partial_{\mathbf{e}_j}\zeta(x) L_F\circ\rho_d(x)\,d\gamma_d(x) = &-\int_X \omega_d^j(x)\zeta(x)\,d\mu_d(x)\\ &+\int_X L_F\circ\rho_d(x)\zeta(x)g_d^j(x)\,d\gamma_d(x).\end{aligned} \tag{5.10}$$

Now we pass to the limit in (5.10) as $d\to\infty$. The first term converges to the analogous term involving γ and ρ by (5.8), the second one converges too, thanks to (5.7). Adding and subtracting g^j in the last term and using (5.8) with $\varphi = g^j$ we have also convergence of that term. Hence, we find

$$\begin{aligned}\int_X \partial_{\mathbf{e}_j}\zeta(x) L_F\circ\rho(x)\,d\gamma(x) = &-\int_X \omega^j(x)\zeta(x)\,d\mu(x)\\ &+\int_X L_F\circ\rho(x)\zeta(x)g^j(x)\,d\gamma(x) \quad \forall j\in\mathbb{N},\end{aligned} \tag{5.11}$$

that is, $\partial_{\mathbf{e}_j}(L_F\circ\rho) = \rho\omega^j \in L^1(X,\gamma)$. Finally, since $\boldsymbol{\omega} \in L^2(X,\mu;X)$, we obtain $L_F\circ\rho \in W^{1,1}(X,\gamma)$ and

$$\boldsymbol{\omega} = \frac{\nabla(L_F\circ\rho)}{\rho}, \tag{5.12}$$

and since $\boldsymbol{\omega}$ is the minimal selection we have also

$$\mathscr{G}(\mu|\gamma) = |\partial\mathscr{F}(\mu|\gamma)|^2.$$

We have proven the implication for the bounded case. Now we shall pass to the general one. Let $n\in\mathbb{N}$ and consider functionals $\mathscr{F}^n(\cdot|\gamma)$, defined in (4.7). These functionals are strongly convex, as noticed in Remark 4.4, and $\Gamma(\mathscr{P}_2(X))$-converge to $\mathscr{F}(\cdot|\gamma)$ as $n \to \infty$ (indeed, condition (4.21) is trivial, whereas (4.22) can be achieved by a truncation argument). Moreover, since $\mathscr{F}^n \geq \mathscr{F}$, it is easy to show tightness for the sets corresponding to the ones in (4.23). Then, by means of Theorem 4.10 again, we find subsequences (that we don't relabel) $\mu_n \to \mu$ in $\mathscr{P}_2(X)$ and $\boldsymbol{\omega}_n \in \partial\mathscr{F}^n(\mu_n|\gamma)$ such that $\mathscr{F}^n(\mu_n|\gamma) \to \mathscr{F}(\mu|\gamma)$ and

$$\boldsymbol{\omega}_n \to \boldsymbol{\omega} = \partial^0\mathscr{F}(\mu|\gamma) \text{ strongly in } L^2 \text{ as in Definition 2.3.} \tag{5.13}$$

We have $\rho_n \leq n$, since $\mathscr{F}^n(\mu_n|\gamma)$ is finite. So, the already obtained result for the bounded case entails $L_F \circ \rho_n \in W^{1,1}(X,\gamma)$ and ensures that the square of the metric slope at μ_n is characterized as

$$\mathscr{G}(\rho_n\gamma|\gamma) = \int_X \frac{\|\nabla L_F \circ \rho_n\|^2}{\rho_n}\, d\gamma.$$

Notice that the weak convergence of $\rho_n\gamma$ to $\rho\gamma$ and the convergence of $\mathscr{F}(\rho_n\gamma|\gamma) = \mathscr{F}^n(\rho_n\gamma|\gamma)$ to $\mathscr{F}(\rho\gamma|\gamma)$ imply, thanks to the strict convexity of F, that $\rho_n \to \rho$ in γ-measure (see [VIS, Theorem 3] or [BR]); in particular a subsequence of (ρ_n) converges to ρ γ-a.e. Hence, we can apply Lemma 5.2 to that subsequence to conclude that $L_F \circ \rho \in GW^{1,1}(X,\gamma)$ and that

$$\int_X \frac{\|\nabla(L_F \circ \rho)\|^2}{\rho}\, d\gamma \leq |\partial\mathscr{F}(\mu|\gamma)|^2. \tag{5.14}$$

Step 2. Now we prove that Sobolev regularity of $L_F \circ \rho$ and integrability of $\|\nabla(L_F \circ \rho)\|^2/\rho$ imply the opposite inequality in (5.14), hence finiteness of slope. First, assume that ρ is bounded and distant from zero. Since ρ^{-1} is bounded we have $\|\nabla(L_F \circ \rho)\| \in L^2(X,\gamma)$, and since L_F has a locally Lipschitz inverse by strict convexity of F, Theorem 2.7 yields $\rho \in W^{1,2}(X,\gamma)$. Let ρ_d be the d-dimensional cylindrical projection of ρ. By (2.12), $\rho_d \in W^{1,2}(X,\gamma)$ and again Theorem 2.7 gives

$$L_F \circ \rho_d \in W^{1,2}(X,\gamma). \tag{5.15}$$

Moreover, by the chain rule (2.13) we have

$$\nabla(L_F \circ \rho) = L_F'(\rho)\nabla\rho \quad \text{and} \quad \nabla(L_F \circ \rho_d) = L_F'(\rho_d)\nabla\rho_d, \tag{5.16}$$

and these gradients are respectively 0 γ-a.e. on the set of all x such that L_F is not differentiable at $\rho(x)$, $\rho_d(x)$. Since ρ_d and ρ are distant from zero, by (2.4) there holds

$$\frac{(L_F'(\rho_d))^2\|\nabla\rho_d\|^2}{\rho_d} \to \frac{(L_F'(\rho))^2\|\nabla\rho\|^2}{\rho} \quad \text{in} \quad L^1(X,\gamma).$$

In fact

$$\|\nabla\rho_d - \nabla\rho\|^2 \leq \|(\nabla\rho)_d - \nabla\rho\|^2 + \sum_{j=d+1}^{\infty} |\partial_{\mathbf{e}_j}\rho|^2$$

converges to 0 in $L^1(X,\gamma)$ (we use (2.12) and the fact that the convergence (2.4) of cylindrical projections holds for maps with values in X).

On the other hand, $(L'_F(\rho_d))^2/\rho_d$ converge to $(L'_F(\rho))^2/\rho$ in $L^1(X,\gamma)$ and are essentially bounded uniformly in d. Then

$$\lim_{d\to\infty}\int_X \frac{\|\nabla(L_F\circ\rho_d)\|^2}{\rho_d}\,d\gamma_d=\int_X \frac{\|\nabla(L_F\circ\rho)\|^2}{\rho}\,d\gamma. \tag{5.17}$$

In view of (5.15), we can apply Theorem 4.6 and obtain the finiteness of $|\partial\mathscr{F}(\mu_d|\gamma_d)|$, where $\mu_d=\rho_d\gamma_d$, and also $|\partial\mathscr{F}(\mu_d|\gamma_d)|^2=\mathscr{G}(\mu_d|\gamma_d)$. Now we make use of the lower semicontinuity of the metric slope and of (5.17) to infer the finiteness of the slope:

$$|\partial\mathscr{F}(\cdot|\gamma)|^2(\mu)\le\liminf_{d\to\infty}|\partial\mathscr{F}(\cdot|\gamma_d)|^2(\mu_d)\le\int_X \frac{\|\nabla(L_F\circ\rho)\|^2}{\rho}\,d\gamma.$$

Now consider the case in which ρ is bounded but not necessarily distant from 0. Let $\rho_n=\max\{\rho,\frac{1}{n}\}$, so that ρ_n is distant from zero, and $\mu_n=\rho_n\gamma$.

Notice that ρ_n are not probability measures, but the results we apply are obviously still valid if, instead of working in $\mathscr{P}_2(X)$, one works in the space $z\mathscr{P}_2(X)$ with $z>0$ (this can also be seen considering the map $F_z(s)=F(zs)$, to come back to probability measures, as we do in Step 3). Since L_F is nondecreasing, $L_F\circ\rho_n=\max\{L_F\circ\rho,L_F(\frac{1}{n})\}$, and by Theorem 2.7 we can infer that $L_F\circ\rho_n\in W^{1,1}(X,\gamma)$. The chain rule also gives

$$\int_X \frac{\|\nabla(L_F\circ\rho_n)\|^2}{\rho_n}\,d\gamma\le\int_X \frac{\|\nabla(L_F\circ\rho)\|^2}{\rho}\,d\gamma, \tag{5.18}$$

since $\rho_n\ge\rho$ and $\nabla(L_F\circ\rho_n)=0$ γ-a.e. on $\{\rho<1/n\}$. Since we have proven the theorem for the case of a density distant from zero, we have by (5.18) that

$$|\partial\mathscr{F}(\mu_n|\gamma)|^2\le\int_X \frac{\|\nabla(L_F\circ\rho)\|^2}{\rho}\,d\gamma.$$

Using the lower semicontinuity of the slope we conclude.

Finally, in the general unbounded case, we take advantage of the just achieved characterization of the slope at $T_n(\rho)\gamma$. The slope is lower semicontinuous, and reasoning as we did to obtain (5.4), we get

$$\begin{aligned}|\partial\mathscr{F}(\mu|\gamma)|^2\le\liminf_{n\to\infty}|\partial\mathscr{F}(T_n(\rho)\gamma|\gamma)|^2&=\liminf_{n\to\infty}\int_X \frac{\|\nabla(L_F\circ T_n(\rho))\|^2}{T_n(\rho)}\,d\gamma\\&\le\int_X \frac{\|\nabla(L_F\circ\rho)\|^2}{\rho}\,d\gamma.\end{aligned} \tag{5.19}$$

Step 3. Suppose now that either the metric slope at μ is finite or that (5.6) hold. Joining together (5.14) and (5.19) we get the desired equality $|\partial\mathscr{F}(\mu|\gamma)|^2=\mathscr{G}(\mu|\gamma)$. Then, in order to characterize the minimal selection $\partial^0\mathscr{F}(\mu|\gamma)$, we have to show that $\nabla(L_F\circ\rho)/\rho$ belongs to $\partial\mathscr{F}(\mu|\gamma)$. We know from (5.12) that this is true if ρ is bounded. In the general case we check the subdifferential relation (3.3) with $\phi=\mathscr{F}$ and $\xi=\nabla(L_F\circ\rho)/\rho$ by approximation;

thanks to Remark 3.3, it suffices to check the property for all $\nu = f\gamma$ with f bounded. Now we approximate ρ by $\rho_n := z_n^{-1}(\rho \wedge n)$, where $z_n \uparrow 1$ is a normalizing constant, and we write the subdifferential relation for ρ_n, $F_n(s) = F(z_n s)$, to obtain:

$$\int_X F(z_n f(x))\, d\gamma(x) \geq \int_X F(\rho(x) \wedge n)\, d\gamma(x) + \int_X \left\langle \frac{\nabla(L_{F_n} \circ \rho_n)(x)}{\rho_n(x)}, \mathbf{t}_n(x) - x \right\rangle \rho_n(x)\, d\gamma(x),$$

where $\mathbf{t}_n$ are the optimal maps from $\rho_n\gamma$ to ν. Since $L_{F_n}(s) = L_F(z_n s)$, $L_{F_n} \circ \rho_n = L_F \circ (\rho \wedge n)$, and using the chain rule this immediately gives

$$\lim_{n\to\infty} \int_X \left\| \frac{\nabla(L_{F_n} \circ \rho_n)}{\rho_n} - \frac{\nabla(L_F \circ \rho)}{\rho} \right\|^2 \rho_n\, d\gamma = 0.$$

Hence, we need only to check that

$$\lim_{n\to\infty} \int_X \left\langle \frac{\nabla(L_F \circ \rho)(x)}{\rho(x)}, \mathbf{t}_n(x) - x \right\rangle \rho_n(x)\, d\gamma(x) = \int_X \left\langle \frac{\nabla(L_F \circ \rho)(x)}{\rho(x)}, \mathbf{t}(x) - x \right\rangle \rho(x)\, d\gamma(x).$$

By a density argument, it suffices to check that

$$\lim_{n\to\infty} \int_X \langle g(x), \mathbf{t}_n(x) - x \rangle\, \rho_n(x)\, d\gamma(x) = \int_X \langle g(x), \mathbf{t}(x) - x \rangle \rho(x)\, d\gamma(x)$$

for all $g \in C_b(X; X)$. Writing the integrals above in terms of optimal plans, the formula reduces to

$$\lim_{n\to\infty} \int_X \langle g(x), y - x \rangle\, d\beta_n(x, y) = \int_X \langle g(x), y - x \rangle\, d\beta(x, y).$$

The latter is a direct consequence of the tightness of (β_n) (because the marginals are tight), of the fact that any limit point is an optimal plan from $\rho\gamma$ to γ (see for instance [AGS, Proposition 7.1.3]) and of the uniqueness of β proved in Theorem 3.2. $\square$

After Theorem 5.3, we can give a straightforward proof of the main result.

Proof of Theorem 1.1 Notice that the domain $D(\mathscr{F}(\cdot|\gamma))$ is dense in $\mathscr{P}_2(X)$ and, under Assumption 4.1, $\mathscr{F}(\cdot|\gamma)$ is strongly convex. Hence we can apply Theorem 3.8 to obtain, for any $\bar{\mu} \in \mathscr{P}_2(X)$, existence and uniqueness of the gradient flow μ_t of $\mathscr{F}(\cdot|\gamma)$ starting from $\bar{\mu}$. Notice that, by the regularizing effect of the semigroup, $\mu_t \ll \gamma$ for any $t > 0$ even if $\bar{\mu}$ does not have a density with respect to γ. The measures μ_t satisfy (3.5) and (3.6), and with Theorem 5.3 we have characterized, under Assumption 2.5, the Wasserstein subdifferential of $\mathscr{F}(\cdot|\gamma)$ at $\mu_t = \rho_t\gamma$ as $\frac{\nabla(L_F \circ \rho_t)}{\rho_t}$. We deduce that $\mu_t = \rho_t\gamma$ is a solution to (1.2). This solution is unique and satisfies all the additional properties of Remark 3.9.

Finally, if $\bar{\mu} \leq M\gamma$, we know by Lemma 4.5 that such a bound is preserved by the discrete minimizer of functional $\Phi_\tau^{\mathscr{F}}(\cdot, \bar{\mu})$ defined in (4.4) (independently of the value of τ). Since μ_t, in view of Proposition 3.10, is the limit of discrete minimizers, we conclude that $\rho_t \leq M$ γ-a.e. for all $t \geq 0$.

References

[A] M. Agueh Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory, *Adv. Differential Equations*, **10**, (2005), no. 3, 309–360.

[AGS] L. Ambrosio, N. Gigli and G. Savaré: Gradient flows in metric spaces and in the spaces of probability measures, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, (2005).

[AFP] L. Ambrosio, N. Fusco and D. Pallara: Functions of bounded variation and free discontinuity problems, Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, (2000).

[ASZ] L. Ambrosio, G. Savaré, L. Zambotti: Existence and Stability for Fokker-Planck equations with log-concave reference measure, to appear in Probab. Theory Related Fields.

[B] V.I. Bogachev Gaussian measures, *Mathematical Surveys and Monographs*, 62. American Mathematical Society, Providence, RI, 1998.

[BBGG] P. Bénilan, L. Boccardo, T. Gallouët, R. Gariepy, M. Pierre, J.L. Vázquez: An L^1-theory of existence and uniqueness of solutions of nonlinear elliptic equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **22** (1995), no. 2, 241–273.

[BR] H.Brezis: Convergence in $\mathcal{D}'$ and in L^1 under strict convexity. In: Boundary value problems for partial differential equations and applications, RMA Res. Notes Appl. Math., **29** (1993), 43–52.

[C] J. Carrillo Menéndez: Entropy solutions for nonlinear degenerate problems. *Arch. Ration. Mech. Anal.* **147** (1999), no. 4, 269–361.

[CG1] E. A. Carlen, W. Gangbo: Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math.* (2) **157** (2003), 807–846.

[CG2] E. A. Carlen, W. Gangbo: Solution of a model Boltzmann equation via steepest descent in the 2-Wasserstin metric. *Arch. Ration. Mech. Anal.* (2) **172** (2004), 21–64.

[DMOP] G. Dal Maso, F. Murat, L. Orsina, A. Prignet: Renormalized solutions of elliptic equations with general measure data. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **28** (1999), no. 4, 741–808

[DaP] G. Da Prato: An introduction to infinite-dimensional analysis, Springer, 2006.

[DaPZ] G. Da Prato, J. Zabczyk: Second order partial differential equations in Hilbert spaces, London Mathematical Society Lecture Note Series, **293**, Cambridge U. P., 2002.

[EG] L.C. Evans, R.F. Gariepy: Measure theory and fine properties of functions. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[FSS] S. Fang, J. Shao, K.-T. Sturm: Wasserstein space over the Wiener space. *Probab. Theory Relat. Fields*, in press.

[JKO] R. Jordan, D. Kinderlehrer and F. Otto: The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998), 1–17.

[MA] J.Maas: Analysis of infinite-dimensional diffusions, PhD thesis, 2009.

[Mc] R. J. McCann: A convexity principle for interacting gases. *Adv. Math.* **128** (1997), no. 1 153–179.

[Mik] T. Mikami: Dynamical systems in the variational formulation of the Fokker-Planck equation by the Wasserstein metric. *Appl. Math. Optim.* **42** (2000), 203–227.

[O1] F. Otto: Doubly degenerate diffusion equations as steepest descents *Manuscript*, (1996)

[O2] F. Otto: Evolution of microstructure in unstable porous media flow: a relaxational approach. *Comm. Pure Appl. Math.* **52** (1999), 873–915.

[O3] F. Otto: The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, **26** (2001), 101–174.

[O4] F. Otto: Dynamics of Labyrinthine Pattern Formation in Magnetic Fluids: A Mean-Field Theory. *Arch.Rational Mech. Anal.* **141** (1998), 63–103.

[VA] J.L.Vázquez: The porous medium equation. Mathematical Theory. Oxford Mathematical Monographs, Oxford, 2007.

[VI] C. Villani: Optimal transport, old and new. Springer-Verlag (2008)

[VIS] A. Visintin: Strong convergence results related to strict convexity. *Comm. PDE.* **9** (1984), 439–466.