Bollettino
Unione Matematica Italiana

Luigi Ambrosio

Gradient Flows in Metric Spaces and in the Spaces of Probability Measures, and Applications to Fokker-Planck Equations with Respect to Log-Concave Measures

Bollettino dell'Unione Matematica Italiana, Serie 9, Vol. **1** (2008), n.1, p. 223–240.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2008_9_1_1_223_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana,
2008.

Bollettino U. M. I.
(9) **I** (2008), 223-240

Gradient Flows in Metric Spaces and in the Spaces of Probability Measures, and Applications to Fokker-Planck Equations with Respect to Log-Concave Measures (*)

Luigi Ambrosio

Abstract. – *A survey on the main results of the theory of gradient flows in metric spaces and in the Wasserstein space of probability measures obtained in* [3] *and* [4], *is presented.*

1. – Introduction and motivations.

In this paper I will make a short survey on the main results of the theory of gradient flows in metric spaces and in the Wasserstein space of probability measures obtained in [3], see also [4]. In the final part of the paper I will focus mainly on the applications to Fokker-Planck equations and Markov semigroupus in infinite-dimensional spaces, along the lines of [5]. The content of this paper reflects, with some additional comments and outlines of proofs, the talk delivered at the joint UMI-DMV meeting in Perugia, and it is by no means a systematic exposition: we refer to the above-mentioned references and to the bibliographies therein for more accurate informations on this wide subject.

I will start by describing informally some motivations of our work. Let us consider the Fokker-Planck equation (also called forward Kolmogorov equation) in $(0,+\infty)\times\mathbb{R}^n$:

$$\frac{d}{dt}f_t = \varepsilon \Delta f_t + \nabla\cdot(f_t\nabla V), \qquad f(0,\cdot)=\bar\mu.$$

Here the initial condition $\bar\mu$ is attained in a weak sense, and we are particularly interested to the case when $\bar\mu$ is a nonnegative measure.

If V is convex, $C^{1,1}$ and $\int\exp(-V)\,dx=1$, then $f_t\to\gamma:=\exp(-V)\mathscr{L}^n$ weakly in $\mathbb{R}^n$, as $t\to\infty$. The Itó formula (see for instance [37]) provides the link of FP equations with solutions to the stochastic differential equation

$$\text{(SDE)} \qquad dX_t = -\nabla V(X_t)\,dt+\sqrt{2\varepsilon}\,dW_t, \qquad X(0)=x$$

(*) Conferenza tenuta a Perugia il 22 giugno 2007 in occasione del “Joint Meeting U.M.I. - D.M.V.”.

when $\bar{\mu} = \delta_x$: indeed, f_t is the density of the law of $X_t(x)$ (i.e. the expectation of the random variable $\phi(X_t(x))$ is equal to $\int_{\mathbb{R}^n} \phi f_t \, dy$ for all bounded Borel functions ϕ).

The SDE can classically be viewed as a stochastic perturbation of an ODE whose behaviour is very simple to describe, under (uniform) convexity assumptions.

Our main goals are:

• extend this picture to general convex V's (with no growth condition or regularity assumption on V);

• extend these results to infinite-dimensional state spaces H;

• show that these extensions are unique, and jointly continuous with respect to V, finite-dimensional approximations and convergence of the initial conditions.

Problems with constraints and (SDE) with reflection.

For instance, if we consider $V \in C^{1,1}(K)$, with $V = +\infty$ out of K, the (SDE) becomes (from now on I shall assume that $\varepsilon = 1$)

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2}dW_t + \boldsymbol{n}_K(X_t)\,dL_t, \qquad X_0(x) = x,$$

where L_t is reflection process occurring on the boundary of K [11]. If V has an intermediate behaviour (regular up to the boundary in some parts, and going to $+\infty$ on other parts), the precise description of the SDE becomes more complex, while we will see that the problem can be dealt with in a unified way at the PDE level.

Stochastic partial differential equations.

On the other hand, since several evolution PDE's can be viewed as infinite-dimensional ODE's, allowing for an infinite-dimensional state space in the (SDE) leads to stochastic partial differential equations, a subject on which a large literature is available, see for instance [14], [15] and the references therein.

Gibbs measures.

Measures of the form $Z^{-1}\exp(-V)\mu$, typically with V convex and μ product measure, even on an infinite-dimensional state spaces E arise quite often in Statistical Mechanics (for instance $E = \mathbb{R}^{Z_d}$).

Our methods stem from a *variational* formulation of the FP equation first introduced, for the heat equation (corresponding to the "degenerate" case $V = 0$), by Jordan-Kinderlehrer-Otto in [25] (in [30] adopted a similar viewpoint to study the large time asymptotics of the porous medium equation). This point of view has been systematically explored, for some classes of linear and nonlinear equations in the book [3]. This point of view leads, for instance, to an intrinsic and essentially geometric description of the evolution problem, which formally corresponds to a gradient flow in the space of probability measures. Although this approach relies very much on the one hand on convexity (it allows only for controlled perturbations of this condition) and on the other hand on the additive structure of the noise, in this context it produces quite strong and general results.

2. – Background on optimal transportation.

In this section I will describe some essential features of the theory of optimal transportation. Besides [3], a good introduction to this topic is [39].

The problem, raised by Monge in 1781, can be informally described as follows: given $X, Y \subset \mathbb{R}^n$, we have two distributions of mass $\rho(x)$ in X and $\rho'(y)$ in Y satisfying the *mass balance condition*

$$\int_X \rho(x)\,dx = \int_Y \rho'(y)\,dy$$

and we want to move ρ into ρ' in such a way that the work done is minimal.

The admissible movements are descrived by a *transport map* $T : X \to Y$ such that

$$\int_{T^{-1}(E)} \rho(x)\,dx = \int_E \rho'(y)\,dy \qquad \text{for all } E \subset Y \text{ Borel.}$$

Since *work* = *mass*×*displacement*, we have to minimize

$$\mathcal{E}(T) := \int_X |T(x) - x|\rho(x)\,dx$$

among all admissible transport maps T.

Despite its very classical and "natural" structure, for a long time this variational problem has not been considered so much, in contrast with the variational problems, for instance, raised by Mechanics: for instance the problem of minimizing an action functional

$$\mathcal{A}(x) := \int_0^1 L(t, x(t), \dot{x}(t))\,dt.$$

As a matter of fact, even some basic issues, as the analogue in this context of the Euler-Lagrange equations

$$\frac{d}{dt} L_{\dot{x}}(t, x(t), \dot{x}(t)) = L_x(t, x(t), \dot{x}(t))$$

were not understood, until much more recent times.

Indeed, the problem could be attacked successfully only with the modern tools of Measure Theory and Functional Analysis, with the seminal work of Kantorovich, in 1940. In even more recent times (last 20 years) many more connections are emerging between this theory and many other fields: Shape Optimization, Geometric and Functional inequalities, Nonlinear diffusion, Partial Differential Equations, Riemannian Geometry. Here I will mainly focus on the relations between optimal transportation and PDE's.

2.1 – *A modern formulation of the optimal transport problem.*

We consider:
- a probability measure μ in X;
- a probability measure ν in Y;
- a function $c : X \times Y \to [0, +\infty]$.

Then, we minimize the energy

$$\mathcal{E}(T) := \int_X c(x, T(x))\, d\mu(x)$$

among all maps T satisfying

$$\mu(T^{-1}(E)) = \nu(E) \qquad \forall E \subset Y.$$

In short, from now on we will denote by $T_{\#}$ the push forward operator induced by T, mapping measures in X to measures in Y, and write $T_{\#}\mu = \nu$. An even more general formulation, allowing *transport plans* instead of transport maps, was considered by Kantorovich, and is very popular and studied in Probability: *find a law in $X \times Y$ whose marginals are μ and ν, and such that the expectation of c is minimal.* Equivalently, we say that $\eta \in \mathscr{P}(X \times Y)$ is a *transport plan* between μ and ν if its first and second marginals are μ and ν respectively (i.e. $\mu(A) = \eta(A \times Y)$ and $\nu(B) = \eta(X \times B)$ whenever A and B are Borel). Then, we consider the problem

$$\inf\left\{ \int_{X\times Y} c\,\eta \,:\, \eta \text{ plan between } \mu \text{ and } \nu \right\}. \tag{2.1}$$

Even though the theory of optimal transportation becomes in many respects simpler and more flexible, working at the level of plans, it turns out that some applications of the theory really depend on the existence of optimal transport maps; for this reason we review some basic results in this direction.

2.2 – *Cost=distance*^{2}.

In this case the optimal transport problem has a unique solution, at least when the initial measure is absolutely continuous with respect to Lebesgue measure, and the optimal transport map exhibits a remarkable structure.

THEOREM 2.1 (Brenier, Knott–Smith, '80). – *Assume that* $X = Y = \mathbb{R}^n$, $c(x, y) = \frac{1}{2}|x - y|^2$, *and* $\mu \ll \mathscr{L}^n$. *Then there exists a unique optimal transport map. Furthermore, this map is the gradient of a convex function.*

The classical proofs of this fact use the differentiability of the cost function. Another even more remarkable fact is that the theory extends to manifolds, where the distance fails to be differentiable in the large:

THEOREM 2.2 (McCann, '01). – *Assume that* $X = Y = M$ *compact* C^3 *Riemannian manifold without boundary,* $c(x, y) = \frac{1}{2}d_M^2(x, y)$, *and* $\mu \ll \mathrm{Vol}_M$. *Then there exists a unique optimal transport map, representable by*

$$T(x) = \exp_x(-\nabla\varphi(x)) \qquad \text{for } \mu\text{-a.e. } x \in M$$

for a suitable potential function φ. *Furthermore,* $[0, T] \ni t \mapsto \exp_x(-t\nabla\varphi(x))$ *is a minimizing geodesic in* M *for all* $T < 1$.

This result has been the starting point of the applications of optimal transportation to Riemannian Geometry, which led in particular (in a series of papers by Lott-Villani and Sturm-Von Renesse) to a weak definition of nonnegative Ricci curvature based on optimal transportation.

2.3 – *The Wasserstein distance.*

The infimum in Monge's problem can be used to define a distance, called Wasserstein distance, between probability measures in X. In the case cost=distance^2, and for measures μ with no atom [] (for general measures Kantorovich's formulation (2.1) should be used), we can set

$$W_2(\mu, \nu) := \inf\left\{ \sqrt{\int_X d^2(x, T(x))\, d\mu(x)} : \ T_\#\mu = \nu \right\}.$$

The "manifold" $\mathscr{P}(X)$ of probability measures on X becomes in this way a metric space, which inherits many properties of X (e.g., compact if X is compact, complete if X is complete, positively curved in the Alexandrov sense if X is positively curved,....). Moreover, if X is a length space, then $\mathscr{P}(X)$ is a length space as well, and if T is an optimal transport map between μ and ν, then

$$\mu_t := (T_t)_\#\mu \qquad \text{with } d(x, T_t(x)) = td(x, T(x)),\ t \in [0, 1]$$

is a constant speed geodesic between μ and ν. These "nonlinear" interpolations, quite different from the classical ones

$$t \mapsto (1 - t)\mu + t\nu$$

were first introduced by McCann in [29]. As a matter of fact, the conventional interpolation above has even an *infinite* length with respect to W_2.

3. – Gradient flows and their variational formulation.

On a Riemannian manifold $\mathcal{M}$, the gradient flow of a function $F : \mathcal{M} \to \boldsymbol{R}$ starting from $\bar{x}$ is defined by

$$\begin{cases} x'(t) = -\nabla F(x(t)) \\ x(0) = \bar{x}. \end{cases}$$

A key property fulfilled by gradient flows is the *energy dissipation identity*:

$$\frac{d}{dt} F(x(t)) = -|\nabla F(x(t))|^2.$$

The Riemannian metric of $\mathcal{M}$ is necessary, in the definition of gradient flows, because of the identification of the differential dF (a covector) with a vector ∇F. It is natural to look, in a nonsmooth setting, to equivalent formulations where the metric (or, more precisely, the distance) plays a more explicit role. Essentially all weak definitions fit in one of the these three groups:

- (EDI) A definition based on the energy dissipation rate. This appears in the paper [16] by De Giorgi and collaborators, and in an unpublished work by Perelman-Petrunin [31], see also [26].
- (EVI) A definition based on a family of Evolution Variational Inequalies; this is inspired by [6].
- A discrete in time solution, provided by the classical implicit Euler scheme.

In [3] we systematically explore these concepts and use, in particular, the third one as a tool to prove existence, passing to the limit as the time step goes to zero either in the (EDI) or in the (EVI) sense.

3.1 – *Energy dissipation rate.*

De Giorgi's formulation of the energy dissipation identity is the following:

$$\text{(EDI)} \qquad \frac{d}{dt} F(x(t)) \le -\frac{1}{2}|x'(t)|^2 - \frac{1}{2}|\nabla F(x(t))|^2.$$

While the plain energy dissipation identity (a single equation) does not encode the ODE $x' = -\nabla F(x)$ (a system of equations), (EDI) surprisingly does! Indeed

$$\begin{aligned} \frac{d}{dt} F(x(t)) &= -\langle x'(t), \nabla F(x(t))\rangle \\ &\ge -|x'(t)||\nabla F(x(t))| \ \ge\ -\frac{1}{2}|x'(t)|^2 - \frac{1}{2}|\nabla F(x(t))|^2 \end{aligned}$$

and the first inequality holds if and only if $-x'(t)$ and $\nabla F(x(t))$ are parallel (i.e. they differ by multiplication of some nonnegative scalar), while the second inequality holds if and only if and only if the two vectors have the same modulus.

Furthermore, all terms in (EDI) make sense even for evolution curves $x(t)$ in a metric space (E, d), provided we understand $|x'(t)|$ as the *metric derivative*

$$|x'(t)| := \lim_{h\to 0} \frac{d(x(t+h), x(t))}{|h|}$$

and $|\nabla F(x(t))|$ as the *descending slope*

$$|\partial F(u)| := \limsup_{v\to u} \frac{[F(u) - F(v)]^+}{d(u,v)}.$$

This formulation is very useful to handle passage to limits, and we will see that the Euler scheme contains a discrete version of (EDI). Furthermore, if F is convex along constant speed geodesics, equality holds in (EDI), the map $t \mapsto F(x(t))$ is locally absolutely continuous in $(0, +\infty)$ and the classical energy dissipation identity can be recovered (see [3, Theorem 2.3.3 and § 2.4]).

3.2 – *Evolution Variational Inequalities.*

Assume that F is convex in $\mathcal{M} = \mathbb{R}^n$, and assume for simplicity that F is everywhere differentiable: then we have the *energy inequality*

$$F(v) \geq F(u(t)) + \langle \nabla F(u(t)), v - u(t)\rangle \qquad \forall v,$$

and the *derivative of distance* $\frac{d}{dt}|u(t) - v|^2 = 2\langle u'(t), u(t) - v\rangle$, so that we can write the gradient flow equation as

$$F(v) \geq F(u(t)) - \langle u'(t), v - u(t)\rangle = F(u(t)) + \frac{d}{dt}\frac{1}{2}|u(t) - v|^2.$$

As in the case of (EDI), and with an even minor effort, we see that this formulation of the gradient flow can be immediately translated in a general metric space (E, d): we say that $u(t)$ is an (EVI) of F if, in the dense of distributions in $(0, +\infty)$, we have

$$\text{(EVI)} \qquad \frac{1}{2}\frac{d}{dt}\, d^2(u(t), v) + F(u(t)) \leq F(v) \qquad \forall v \in E.$$

This property is very strong, and it leads quite easily to uniqueness (contractivity), stability both w.r.t. $\bar{u}$ and F, regularizing effects, that we will mention more precisely later on. It should not be surprising that existence is much harder in this formulation, compared with the one based on energy dissipation.

As an illustration of the strength of this definition, we sketch the proof of uniqueness of (EVI): given u^1, u^2 solutions of (EVI), both starting from $\bar{u}$, insert $v = u^2(t)$ into

$$\frac{1}{2}\frac{d}{dt}\, d^2(u^1(t), v) \le F(v) - F(u^1(t))$$

and $v = u^1(t)$ into

$$\frac{1}{2}\frac{d}{dt}\, d^2(v, u^2(t)) \le F(v) - F(u^2(t))$$

to obtain

$$\frac{d}{dt}\, d^2(u^1(t), u^2(t)) \le 0.$$

This argument is not completely rigorous, because each of the differential inequalities in (EVI) has an exceptional set of times, possibly depending on v; however, if we understand the derivatives not in the pointwise sense, but in the distributional one, the gap can be filled (see [3, Lemma 4.3.4] for the precise argument, inspired by Kruzkhov method of doubling of variables).

Besides the contraction property, solutions to (EVI) satisfy (see [3, Theorem 4.0.4] and [4]):

REGULARIZATION. – For $t > 0$ we have

$$F(u(t)) \le \inf_{v\in E}\left\{F(v) + \frac{1}{2t} d^2(\bar{u}, v)\right\}$$

and

$$|\nabla F|(u(t))^2 \le \inf_{v\in E}\left\{|\nabla F(v)|^2 + \frac{1}{t^2} d^2(\bar{u}, v)\right\}.$$

These properties are important in some applications: for instance in the case that we will examine later on, when F is the relative entropy functional (with respect to some reference measure γ), and the initial condition is a Dirac mass (thus with an infinite relative entropy), the conditions above tell us in a quantitative way how quickly the Dirac mass spreads, and provide upper bounds on the logarithmic gradient of the density with respect to γ, for positive times.

MONOTONICITY. – (EVI) solutions are also solutions in the (EDI) sense, and have satisfy the typical properties of gradient flows: $t \mapsto F(u(t))$ is nonincreasing in $(0, +\infty)$ and its right derivative equals, for all times, $-|\partial F(u(t))|^2$. Furthermore, $t \mapsto |\nabla F(u(t))|^2$ is nonincreasing as well. Finally, under *uniform* convexity assumptions, these properties can be improved, and lead to exponential convergence to the unique minimum point $x_{\min}$.

3.3 – *Implicit Euler scheme.*

It is a classical fact that a discrete solution (in time) to the gradient flow can be built by the implicit Euler scheme: given a time step $\tau > 0$, we define a piecewise constant (in time) solution $x_\tau(t)$ by

$$x_\tau(t) := x_k, \qquad t \in (k\tau, (k+1)\tau],$$

where $x_0 = \bar{x}$ and (x_k) is built by minimizing recursively

$$y \mapsto F(y) + \frac{1}{2\tau}|y - x_{k-1}|^2.$$

In Hilbert spaces, it is clear why this is an approximation of the gradient flow, by the *first discrete Euler equation*:

$$\frac{x_k - x_{k-1}}{\tau} = -\nabla F(x_k)$$

So, x_k is implicitly defined by this equation, and explicitly by

$$x_k = (\mathrm{Id} + \tau\nabla F)^{-1}(x_{k-1}),$$

whenever $(\mathrm{Id} + \tau\nabla F)$ is invertible. Under convexity assumptions on F, $(\mathrm{Id} + \tau\nabla F)^{-1}$ is not only invertible, but also non-expansive, and this easily leads to error estimates, stability and convergence results as $\tau \downarrow 0$.

3.4 – *How (EDI) and (EVI) can be read in the Euler scheme.*

Let us start from (EDI). The first step is to write down a *second discrete Euler* equation, that reads in this context as a slope estimate: whenever y minimizes

$$z \mapsto F(z) + \frac{1}{2\tau}\, d^2(x, z) \tag{3.1}$$

we have

$$|\partial F(y)| \le \frac{d(x, y)}{\tau}. \tag{3.2}$$

Indeed, the minimality of y gives

$$F(y) - F(y') \le \frac{1}{2\tau}(d^2(x, y') - d^2(x, y)) \le \frac{d(y, y')}{2\tau}(d(x, y) + d(x, y')$$

for all $y' \in E$. Dividing by $d(y, y')$ and letting $y' \to y$ gives (3.2). Now, De Giorgi suggested to look to minimizers y_σ, $\sigma \in (0, \tau]$, of

$$z \mapsto F(z) + \frac{1}{2\sigma} d^2(x, z)$$

as a "variational" interpolation between x and y. By looking at the derivative of the energy $F(y_\sigma)$, one obtains (see [3, Theorem 3.1.4]) the discrete energy identity

$$F(x) - F(y) = \frac{d^2(x,y)}{2\tau} + \frac{1}{2}\int_0^\tau \frac{d^2(y_\sigma, x)}{\sigma^2}\, d\sigma$$

which, taking (3.2) and the construction of (x_k) and x_τ into account, leads to a discrete integral version of (EDI):

$$F(x_\tau(k\tau)) - F(x_\tau((k+1)\tau)) \geq \tau \frac{d^2(x_\tau(k\tau), x_\tau((k+1)\tau))}{2\tau^2} + \frac{1}{2}\int_{k\tau}^{(k+1)\tau} |\partial F|^2(\tilde{x}_\tau(s))\, ds,$$

where the interpolant $\tilde{x}_\tau$ is now made using the De Giorgi variational interpolation. After that this has been estabilished, routine arguments provide (EDI) curves in the limit, provided that the slope is lower semicontinuous (see [3, Chapter 3], where also more general assumptions are considered).

Now, let us look at (EVI). In this case we shall be able to derive an Euler equation corresponding to a discrete version of (EVI) under the following assumption:

WEAK CONVEXITY. – For all z, x_0, x_1 there exists a continuous curve x_t between x_0 and x_1 satisfying:

$$\begin{cases} d^2(x_t, z) \leq \ (1-t)\, d^2(x_0, z) + t d^2(x_1, z) \ - t(1-t) d^2(x_0, x_1) \\ F(x_t) \leq \ (1-t) F(x_0) + t F(x_1) \end{cases} \tag{3.3}$$

for all $t \in [0,1]$. Of course this assumption reduces to the usual convexity along geodesics if we choose as x_t the constant speed geodesic between x_0 and x_1; but, in order to have the first inequality involving the squared distance, this forces the space (E, d) to be NPC (*non positively curved*) in the sense of Alexandrov. This would be too restrictive for the applications we have in mind to $\mathscr{P}_2(H)$, a space that is actually PC (*positively curved*) (see the heuristic computation in [30] and [3, Theorem 7.3.2]). For these reasons we keep more freedom in the choice of the interpolating curve, and use it to derive the *third discrete Euler* equation: whenever y minimizes (3.1), we have

$$\frac{d^2(y,v) - d^2(x,v)}{2\tau} + F(y) \leq F(v) \qquad \forall v \in E, \tag{3.4}$$

which is obviously a discrete version of (EVI). To prove (3.4), let us apply (3.3) with $x_0 = y$, $x_1 = v$ and $z = x$ to obtain for $t \geq 0$

$$F(y) + \frac{1}{2\tau} d^2(y,x)$$
$$\leq F(x_t) + \frac{1}{2\tau} d^2(x_t, x) \leq (1-t) F(y) + t F(v) + \frac{1-t}{2\tau} d^2(y,x) + \frac{t}{2\tau} d^2(x,v)$$

Rearranging terms and dividing by $t > 0$ we pass to the limit as $t \downarrow 0$ to obtain (3.4).

The discrete Euler equation (3.4) can be used to show that the family x_τ converges as $\tau \downarrow 0$ (no compactness is needed here, besides the one needed to have existence of the minimizers in the discrete Euler scheme). Denoting by $\mathcal{S}\bar{x}(t)$ the discrete semigroup and by $\mathcal{S}_\tau\bar{x}(t)$ the discrete one, we have the universal estimate [5]

$$d(\mathcal{S}\bar{\mu}(t), \mathcal{S}_\tau\bar{\mu}(t)) \leq 8\sqrt{\tau F(\bar{\mu})}.$$

More general error estimates (even when the time mesh is not uniform) are discussed in [3, Chapter 4], whose proof is inspired by previous work by Baiocchi, Nochetto, Savaré, Verdi in Banach spaces.

4. – Gradient flows in Wasserstein spaces.

The general framework that we outlined works well for functionals F in the space $\mathscr{P}_2(H)$ of probability measures with finite second order moments (here H is an Hilbert space, but also more general state spaces H could be considered, see in particular [33]), provided we understand properly the concept of convexity. Let us also recall that convergence in $\mathscr{P}_2(H)$ is equivalent to weak convergence, in the duality with $C_b(H)$, plus convergence of the second order moments.

Convexity along Wasserstein geodesics, the so-called McCann displacement convexity

$$t \mapsto F(((1-t)I + t\,T))_{\#}\mu) \text{ convex in } [0,1], \ T \text{ optimal}$$

is sufficient to obtain existence in the (EDI) sense. Furthermore, these solutions can be proved to coincide with (EVI) solutions, and are therefore unique (see [Theorem 11.1. and Theorem 11.1.4]ags)

On the other hand, as we saw in Section 3.4, a stronger (or, rather, different) convexity condition is needed to obtain also convergence, *with error estimates*, of the implicit Euler scheme. In this case, the condition is:

STRONG DISPLACEMENT CONVEXITY. – For all $\bar{\mu}$, ν_0, ν_1 there exists a continuous curve ν_t between ν_0 and ν_1 satisfying:

$$\begin{cases} W_2^2(\nu_t, \bar{\mu}) \leq (1-t)\,W_2^2(\nu_0, \bar{\mu}) + tW_2^2(\nu_1, \bar{\mu}) - t(1-t)W_2^2(\nu_0, \nu_1) \\ F(\nu_t) \leq (1-t)F(\nu_0) + tF(\nu_1) \end{cases} \tag{4.1}$$

for all $t \in [0,1]$.

Notice that if ν_t is the constant speed geodesic between ν_0 and ν_1 the *opposite* inequality always holds [3, Theorem 7.3.2]

$$W_2^2(\nu_t, \bar{\mu}) \geq (1-t)W_2^2(\nu_0, \bar{\mu}) + tW_2^2(\nu_1, \bar{\mu}) - t(1-t)W_2^2(\nu_0, \nu_1)$$

and in general equality fails (this, as we said, is a manifestation of the fact that $\mathscr{P}_2(H)$ is a positively curved metric space). However, luckily enough, there is a different family of curves on which the first inequality in (4.1) holds: using for simplicity just transport maps, they are given by

$$t \mapsto \big((1-t)T_0 + tT_1\big)_{\#}\bar{\mu} \tag{4.2}$$

where T_i, $i = 0, 1$, are the optimal transport maps from $\bar{\mu}$ to ν_i.

The second lucky fact is that the conditions that, typically, guarantee the convexity of F along geodesics, guarantee also the convexity of F along this more general class of curves, whence (4.1) follows. In the next section we shall illustrate this phenomenon for a particular class of functionals, the relative entropy functionals.

We conclude this section by illustrating how, besides the (EDI) and the (EVI) viewpoints, also the "differential" viewpoint to gradient flows can be recovered in Wasserstein spaces, making in some sense rigorous the formal Riemannian calculus developed in [30]. The basic fact proves in Chapter 8 of [3] is that any Lipschitz (ore, more generally) absolutely continuous curve $\mu_t : [0, T] \to \mathscr{P}_2(H)$ can be canonically endowed with a "tangent" vector field $\boldsymbol{v}_t \in L^2(\mu_t; H)$; it is uniquely characterized by the *continuity* equation

$$\frac{d}{dt} + \operatorname{div}(\boldsymbol{v}_t \mu_t) = 0 \tag{4.3}$$

(in the infinite-dimensional case, in the duality with smooth cylindrical functions) and by the relation with the metric derivative

$$\|\boldsymbol{v}_t\|_{L^2(\mu_t;H)} = |\mu_t'| \qquad \text{for } \mathscr{L}^1\text{-a.e. } \mathrm{t} \in (0, \mathrm{T}). \tag{4.4}$$

Alternatively, (4.3) and

$$v_t \in \overline{\left\{\nabla\phi \,:\, \phi \ \text{ smooth cylindrical}\right\}}^{L^2(\mu_t;H)} \qquad \text{for } \mathscr{L}^1\text{-a.e. } \mathrm{t} \in (0, \mathrm{T}). \tag{4.5}$$

can be used to characterize the tangent vector field $\boldsymbol{v}_t$. Therefore (4.5) suggests to define the tangent space to $\mathscr{P}_2(H)$ at μ by

$$\operatorname{Tan}_\mu \mathscr{P}_2(H) := \overline{\left\{\nabla\phi \,:\, \phi \ \text{ smooth cylindrical}\right\}}^{L^2(\mu;H)}$$

with the scalar product induced by the inclusion in $L^2(\mu; H)$. In this sense, $\mathscr{P}_2(H)$ can be considered as an infinite-dimensional Riemannian manifold, and the "Riemannian" distance induced by this choice of the tangent bundle and the metric on it is precisely the Wasserstein distance (in a slightly different language, this was proved first by Benamou and Brenier in Euclidean spaces).

Given this setup, gradient flows can be defined as in the classical situations by requiring the velocity field to be the gradient of F, or more generally to belong to the subdifferential of F. For absolutely continuous measures with respect to $\mathscr{L}^n$ (but everything can be extended to infinite dimensions or, working with plans, to singular measures) the natural definition of subdifferential in this context is:

$$\partial F(\mu) := \left\{ \xi \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n) : F(\nu) \geq F(\mu) + \int_{\mathbb{R}^n} \langle \xi, T_\mu^\nu - \mathrm{Id} \rangle \, d\mu \quad \forall \nu \in \mathscr{P}_2(H) \right\}.$$

It turns out that, for convex functionals along geodesics, the viewpoints (EDI), (EVI) and differential coincide.

5. – Log-concavity and relative entropy.

We consider a probability measure γ in H satisfying the *log-concavity* condition:

$$\ln \gamma\big((1-t)A + tB\big) \geq t \ln \gamma(A) + (1-t) \ln \gamma(B) \qquad \forall t \in (0,1)$$

whenever A, $B \subset H$ are open (this avoids the nontrivial problem of the γ-measurability of $(1-t)A + tB$). We assume that γ is non-degenerate, i.e. that γ is not supported in a proper subspace of H.

The class of log-concave measures includes all Gaussian measures γ_G and all multiplicative perturbations $\gamma = \exp(-V)\gamma_G$, with $V : H \to \mathbb{R} \cup \{+\infty\}$ convex and lower semicontinuous. Moreover, by a classical result due to Borell [8] (see also [3, Theorem 9.4.10]), in Euclidean spaces $\mathbb{R}^n$ non-degenerate log-concave measure are precisely those absolutely continuous with respect to $\mathscr{L}^n$, with a density of the form $\exp(-V)$, with V convex and lower semicontinuous.

Now we can consider the *relative entropy* functional with respect to γ:

$$\mathcal{H}(\mu|\gamma) := \begin{cases} \displaystyle\int_H \rho \log \rho \, d\gamma & \text{if } \mu = \rho\gamma; \\ +\infty & \text{otherwise.} \end{cases}$$

It turns out (see [3, Theorem 9.4.11]) that the three conditions:

- γ is log-concave;
- $\mathcal{H}(\cdot|\gamma)$ is displacement convex in $\mathscr{P}_2(H)$;
- $\mathcal{H}(\cdot|\gamma)$ is convex along the curves (4.2);

are all equivalent. Therefore, for log-concave measures γ, the general theory that we outlined is fully applicable. Then, the byproducts of the general theory are:

• existence, uniqueness, regularizing effects (in particular Dirac masses as initial data are allowed), with the estimate

$$\mathcal{H}(\mathcal{S}\delta_x(t)|\gamma) \le \inf_{\rho\in L^1(H,\gamma)} \left\{ \mathcal{H}(\rho\gamma|\gamma) + \frac{1}{2t} W_2^2(\rho\gamma, \delta_x) \right\} < +\infty;$$

• stability with respect to $\bar{\mu}$ and γ (precisely, convergence of $\bar{\mu}_n$ to $\bar{\mu}$ in $\mathscr{P}_2(H)$ and weak convergence of γ_n to γ, in the duality with $C_b(H)$, implies local uniform convergence in $[0, +\infty)$ of the semigroups);

Let us now translate these result in the more conventional language of Dirichlet forms [19], [27]. Although the Wasserstein viewpoint is genuinely nonlinear, a comparison here is possible because of the special choice $F(\mu) = \mathcal{H}(\mu|\gamma)$. It is important to mention, however, that different choices of the nonlinearity $\rho \ln \rho$ are indeed possible within the Wasserstein theory, as in Otto's work [30] on the porous medium equation.

5.1 – *Wasserstein semigroups and Dirichlet forms.*

The following result has been proved in [5].

THEOREM 5.1. – *Let γ be log-concave and non-degenerate, and consider the bilinear form*

$$\mathcal{E}_\gamma(u, v) := \int_H \langle \nabla u, \nabla v \rangle_H \, d\gamma \qquad u,\, v \in C_b^1(H).$$

Then $\mathcal{E}_\gamma$ is closable, its closure is a Dirichlet form and the associated L^2 semigroup P_t^γ is linked to the Wasserstein semigroup by

$$P_t^\gamma f(x) = \int_H f \, d\mathcal{S}^\gamma \delta_x(t) \qquad \forall f \in L^\infty(H, \gamma).$$

Moreover P_t^γ satisfies the strong Feller property:

$$P_t^\gamma : L^\infty(H, \gamma) \hookrightarrow \mathrm{Lip}_b(H) \qquad \forall t > 0.$$

A key property in the proof of the previous theorem is the closability of the Dirichlet form. The classical method to prove this property, based on an integration by parts formula along sufficiently many directions [1], does not seem to be applicable at this level of generality. So, also the proof of this fact in [5] invokes some ideas from optimal transport theory: in particular lower semicontinuity of $\mathcal{E}_\gamma(u, u)$ (which is known to be equivalent to closability) is proved passing through the identity

$$2\sqrt{\mathcal{E}_\gamma(u, u)} = |\partial F(u^2\gamma)| \qquad \text{with} \qquad F(\mu) := \mathcal{H}(\mu|\gamma),$$

using the lower semicontinuity of the slope of geodesically convex functionals.

In the next section we will examine instead the link with the theory of Markov processes.

5.2 – *Wasserstein semigroups and Markov processes.*

THEOREM 5.2 [5]. – *Under the previous assumptions on γ, there exists a unique Markov family $\mathbb{P}^\gamma_x$, $x \in \operatorname{supp}\gamma$, of probability measures in $C\big(]0,+\infty[;H\big)$ such that*

(i) $\mathbb{E}^\gamma_x[f(X_t(x))] = P^\gamma_t f(x)$ *for all* $t > 0$, $f \in L^\infty(H,\gamma)$;

(ii) $\mathbb{E}^\gamma_x[|X_t - x|^2] \to 0$ *as* $t \downarrow 0$.

Moreover, $(x,\gamma) \mapsto \mathbb{P}^\gamma_x$ *enjoys the following stability property: if*

(a) $x_n \in \operatorname{supp}\gamma_n$, $|x_n - x| \to 0$, $x \in \operatorname{supp}\gamma$;

(b) $\gamma_n \to \gamma$ *weakly in H;*

then

$$\langle X(t), h\rangle \text{ under } \mathbb{P}^{\gamma_n}_{x_n} \ \to\ \langle X(t), h\rangle \text{ under } \mathbb{P}^\gamma_x \ \text{ weakly in } C([\delta, T])$$

for all $0 < \delta \le T < +\infty$, $h \in H$.

The proof uses the Ma-Roeckner theory [27], which allows to "lift" the L^2 semigroup of a sufficiently regular Dirichlet form to a Markov process and, for the tightness, the Lyons-Zheng decomposition.

We conclude this survey with a sketch of the proof of stability. Stability of Markov processes depends in essence on the stability of the transition probabilities, i.e. $\mathcal{S}^\gamma\delta_x(t)$. So, essentially we need stability of the Wasserstein semigroup. The uniform error estimates for the Euler scheme allow a reduction to the stability of the discrete (in time) Wasserstein semigroup $\mathcal{S}^\gamma_\tau\delta_x(t)$.

By induction, we are led to the following statement: if $\mu_n \to \mu$ in Wasserstein distance, and ν_n minimize

$$\sigma \mapsto \mathcal{H}(\sigma|\gamma_n) + \frac{1}{2\tau} W_2^2(\sigma, \mu_n),$$

then $\nu_n \to \nu$ in Wasserstein distance, with ν minimizer of

$$\sigma \mapsto \mathcal{H}(\sigma|\gamma) + \frac{1}{2\tau} W_2^2(\sigma, \mu).$$

De Giorgi designed in the '70 (see [13]) a theory to allow these passages to limits, the so-called Γ-convergence. It turns out that $\mathcal{H}(\,\cdot\,|\gamma_n)$ converge to $\mathcal{H}(\,\cdot\,|\gamma)$ *precisely* in the sense of Γ-convergence, see [3, Lemma 10.3.16 and Lemma 10.3.17] or [5] for a precise statement.

Finally, I will brief mention some work in progress by Fang and A.-Savaré-Zambotti on the Ornstein-Uhlenbck process in Wiener spaces. At the level of the

SDE, for $\gamma = \mathcal{N}(0, Q)$, it corresponds to replace

$$dX_t = -Q^{-1}X_t dt + dW_t \qquad \text{with} \qquad dY_t = -Y_t dt + \sqrt{Q} dW_t.$$

At the Wasserstein level, it corresponds to replace the standard quadratic cost function by the Feyel-Ustünel [18] one:

$$c(x, y) := \begin{cases} \|x - y\|^2 & \text{if } x - y \in \mathcal{H}; \\ +\infty & \text{otherwise,} \end{cases}$$

where $(\mathcal{H}, \|\cdot\|)$ is the Cameron-Martin space of γ.

REFERENCES

[1] S. ALBEVERIO - S. KUSUOKA, *Maximality of infinite-dimensional Dirichlet forms and Hegh-Krohn's model of quantum fields*. Ideas and methods in quantum and statistical physics (Oslo, 1988), Cambridge Univ. Press, Cambridge (2002), 301-330.

[2] L. AMBROSIO - N. FUSCO - D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press (2000).

[3] L. AMBROSIO - N. GIGLI - G. SAVARÉ, *Gradient flows in metric spaces and in the spaces of probability measures*. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, (2005).

[4] L. AMBROSIO - G. SAVARÉ, *Gradient flows in spaces of probability measures.* Handbook of Differential Equations. Evolutionary equations III, North Holland 2007.

[5] L. AMBROSIO - G. SAVARÉ - L. ZAMBOTTI, *Existence and Stability for Fokker-Planck equations with log-concave reference measure.* (2007), Submitted paper.

[6] P. BÉNILAN, *Solutions intégrales d'équations d'évolution dans un espace de Banach.* C.R.Acad.Sci. Paris Sér. A-B, **274**, (1972).

[7] V. I. BOGACHEV, *Gaussian measures.* Mathematical Surveys and Monographs, **62**, AMS (1998).

[8] C. BORELL, *Convex set functions in d-space.* Period. Math. Hungar., **6** (1975), 111-136.

[9] H. BRÉZIS, *Opérateurs maximaux monotones.* North-Holland, Amsterdam, (1973).

[10] J. A. CARRILLO - R. MCCANN - C. VILLANI, *Contraction in the 2-Wasserstein length space and thermalization of granular media.* Arch. Rational Mech. Anal., **179** (2006), 217-263.

[11] E. CÉPA, *Problème de Skorohod multivoque*, Annals of Probability, **26** no. 2 (1998), 500-532.

[12] A. DEBUSSCHE - L. ZAMBOTTI, *Conservative Stochastic Cahn-Hilliard equation with reflection*, to appear in Annals of Probability (2007).

[13] G. DAL MASO, *An introduction to Γ-convergence.* Birkhauser (1993).

[14] G. DA PRATO - J. ZABCZYK, *Second order partial differential equations in Hilbert spaces*, London Mathematical Society Lecture Notes Series, **293**, Cambridge University Press (2002).

[15] G. DA PRATO - J. ZABCZYK, *Ergodicity for Infinite Dimensional Systems*, London Mathematical Society Lecture Notes, n.229, Cambridge University Press (1996).

[16] E. DE GIORGI - A. MARINO & M. TOSQUES, *Problems of evolution in metric spaces and maximal decreasing curves.* Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., (8) **68** (1980), 180-187.

[17] S. FANG, *Wasserstein space over the Wiener space.* Preprint, 2007.

[18] D. FEYEL - A. S. USTÜNEL, *Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space*, Probab. Theory Relat. Fields, **128** (2004), 347-385.

[19] M. FUKUSHIMA - Y. OSHIMA - M. TAKEDA, *Dirichlet Forms and Symmetric Markov Processes.* Walter de Gruyter, Berlin-New York (1994).

[20] T. FUNAKI, *Stochastic Interface Models*. In: Lectures on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003 (ed. J. Picard), 103-274, Lect. Notes Math., **1869**, Springer (2005).

[21] T. FUNAKI - S. OLLA, *Fluctuations for $\nabla\phi$ interface model on a wall.* Stoch. Proc. and Appl, **94** (2001), 1-27.

[22] T. FUNAKI - H. SPOHN, *Motion by mean curvature from the Ginzburg-Landau $\nabla\phi$ interface model.* Comm. Math. Phys. **185** (1997), 1-36.

[23] G. GIACOMIN - S. OLLA - H. SPOHN, *Equilibrium fluctuations for $\nabla\phi$ interface model*, Ann. Probab. **29** (2001), 1138-1172.

[24] F. JOHN, *Partial differential equations*. Springer (4th. ed.) (1970).

[25] R. JORDAN - D. KINDERLEHRER - F. OTTO, *The variational formulation of the Fokker-Planck equation.* SIAM J. Math. Anal. **29** (1998), 1-17.

[26] A. LYTCHAK, *Open map theorem for metric spaces.* St. Petersburg Math. J., **17** (2006), 477-491.

[27] Z. M. MA - M. RÖCKNER, *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms.* Universitext, Springer-Verlag (1992).

[28] D. NUALART - E. PARDOUX, *White noise driven quasilinear SPDEs with reflection*, Prob. Theory and Rel. Fields, **93** (1992), 77-89.

[29] R. J. MCCANN, *A convexity principle for interacting gases.* Adv. Math., **128**, (1997) 153-179.

[30] F. OTTO, *The geometry of dissipative evolution equations: the porous medium equation.* Comm. PDE, **26** (2001), 101-174.

[31] G. PERELMAN & A. PETRUNIN, *Quasigeodesics and gradient curves in Alexandrov spaces.* Unpublished preprint (1994).

[32] D. REVUZ - M. YOR, *Continuous Martingales and Brownian Motion*, Springer Verlag (1991).

[33] G. SAVARÉ, *Gradient flows and diffusion semigroups in metric spaces and lower curvature bounds.* CRAS note (2007), in press.

[34] S. SHEFFIELD, *Random Surfaces*, Asterisque, No. 304 (2005).

[35] A. V. SKOROHOD, *Stochastic equations for diffusions in a bounded region*, Theory Probab. Appl. **6** (1961), 264-274.

[36] H. SPOHN, *Interface motion in models with stochastic dynamics*, J. Stat. Phys. **71** (1993), 1081-1132.

[37] D. W. STROOCK - S. R. S. VARADHAN, *Multidimensional diffusion processes.* Springer Verlag, second ed (1997).

[38] H. TANAKA, *Stochastic differential equations with reflecting boundary condition in convex regions*, Hiroshima Math. J. **9** (1979), 163-177.

[39] C. VILLANI, *Topics in optimal transportation.* Graduate Studies in Mathematics, **58** (2003), AMS.

[40] L. ZAMBOTTI, *Integration by parts formulae on convex sets of paths and applications to SPDEs with reflection*, Probab. Theory Related Fields, **123** no. 4 (2002), 579-600.

[41] L. Zambotti, *Integration by parts on δ-Bessel Bridges, $\delta > 3$, and related SPDEs*, Annals of Probability, **31** no. 1 (2003), 323-348.
[42] L. Zambotti, *Fluctuations for a $\nabla\varphi$ interface model with repulsion from a wall*, Prob. Theory and Rel. Fields, **129** no. 3 (2004), 315-339.
[43] L. Zambotti, *Convergence of approximations of monotone gradient systems*, Journal of Evolution Equations, **6** no. 4 (2006), 601-619.

Luigi Ambrosio, Scuola Normale Superiore, Pisa
e-mail: l.ambrosio@sns.it

Received December 17, 2007