Lecture 18: An Introduction to Otto's Calculus

1 Otto's Calculus

In the seminal papers [94, 95], Otto introduced a formal interpretation of the Wasserstein distance as a Riemannian distance in $\mathscr{P}_2(\mathbb{R}^n)$, using this interpretation as a guide to rigorous results on the large time asymptotics of solutions to porous medium equations. To simplify the presentation, let us work in the subspace $\mathscr{P}_2^a(\mathbb{R}^n)$ of absolutely continuous measures μ, identified with their densities ϱ.

According to Otto's calculus, elements s of the tangent space $T_\varrho \mathscr{P}_2^a(\mathbb{R}^n)$ are thought as functions with null mean and gradient tangent vectors $v = \nabla\phi$ are coupled to tangent vectors s by solving the degenerate elliptic PDE

$$-\operatorname{div}\big((\nabla\phi)\varrho\big) = s.$$

Notice that this picture is fully consistent with the structure of the continuity equation, that we extensively discussed in the previous sections. Otto's metric tensor is then

$$\langle s, s'\rangle_\varrho := \int_{\mathbb{R}^n} \langle \nabla\phi, \nabla\phi'\rangle \varrho \,\mathrm{d}x \quad \text{whenever} \quad -\operatorname{div}\big((\nabla\phi)\varrho\big) = s,\ -\operatorname{div}\big((\nabla\phi')\varrho\big) = s'. \tag{18.1}$$

Notice that the vector fields v are *gradient* vector fields and that, when $\varrho \sim 1$, this metric is reminiscent of the flat H^{-1} metric on tangent vectors. The restriction to gradient vector fields can be understood, for instance, on the basis of the fact that optimal transport maps are gradients. In addition, the same holds for the natural velocity field $v_t = t^{-1}\nabla\psi_t$ attached to geodesics (see Sect. 2 in Lecture 16).

L. Ambrosio et al., *Lectures on Optimal Transport*, UNITEXT 130,
https://doi.org/10.1007/978-3-030-72162-6_18

In this perspective, the Benamou–Brenier formula can be precisely understood as the fact that W_2 is the Riemannian distance associated to this metric tensor, as $\int_{\mathbb{R}^n} |\nabla\phi_t|^2 \varrho_t \,\mathrm{d}x$ is precisely the Riemannian action of the curve ϱ_t solving a continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho_t + \mathrm{div}\,\big((\nabla\phi_t)\varrho_t\big) = 0.$$

More generally, the close link between solutions of the continuity equations and absolutely continuous curves in $\mathscr{P}_2(\mathbb{R}^n)$, proved in Proposition 17.9, shows that this interpretation of $\mathscr{P}_2(\mathbb{R}^n)$ as a kind of infinite-dimensional Riemannian manifold is fully consistent and goes beyond the class of absolutely continuous measures.

2 Formal Interpretation of Some Evolution Equations as Wasserstein Gradient Flows

This section is devoted to discuss at a formal level the possibility of interpreting some evolution equations as gradient flows of energy functionals with respect to the Wasserstein distance, relying on Otto's calculus and computing the "Wasserstein gradient" $\nabla^W\mathcal{E}$. In many cases the metric theory can then be used, among other things, to make this interpretation fully rigorous and in Sect. 3 we will perform this task for the heat equation.

Assume that an energy $\mathcal{E} : \mathscr{P}_2(\mathbb{R}^n) \to (-\infty, \infty]$ is given. We recall that any gradient flow μ_t must be in particular an absolutely continuous curve with metric derivative locally in L^2. At this stage, thanks to Theorem 17.10 we can say that μ_t is a solution of the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \mathrm{div}(v_t\mu_t) = 0, \tag{18.2}$$

for some velocity field v_t such that $\|v_t\|_{L^2(\mu_t)} \in L^1_{\mathrm{loc}}(0,\infty)$. Thus, if we are able to identify the Wasserstein gradient of $\mathcal{E}$, the condition

$$v_t = -\nabla^W\mathcal{E}(\mu_t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,\infty) \tag{18.3}$$

turns the continuity equation (18.2) into

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t = \mathrm{div}(\nabla^W\mathcal{E}(\mu_t)\mu_t). \tag{18.4}$$

Let us compute then the Wasserstein gradients of the main examples of energy functionals we introduced so far: internal energy, potential energy and interaction energy.

As we discussed in Lecture 13, having a metric tensor at our disposal, the problem of computing the Wasserstein gradient of $\mathcal{E}$ at any $\varrho \in \mathscr{P}_2^a(\mathbb{R}^n)$ is reduced to the computation of the action of its differential $\mathrm{d}_\varrho \mathcal{E}$ on any $s \in T_\varrho \mathscr{P}_2^a(\mathbb{R}^n)$, and therefore to the computation of

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathcal{E}(\varrho_t),$$

where ϱ_t is any absolutely continuous curve passing through ϱ with velocity s at time $t = 0$. To this aim we fix $\varphi \in C_c^\infty(\mathbb{R}^n)$ and we consider the vector field $v = \nabla\varphi$. There are two natural choices for a smooth curve in the space of probabilities with velocity v at time 0, namely

$$\text{either} \quad \varrho_t := (\mathrm{id} + tv)_\# \varrho \quad \text{or} \quad \tilde{\varrho}_t := (X_t)_\# \varrho,$$

where X_t is the flow map associated to v. It is easily seen (recall Lecture 16) that the velocity field for the former is $v_t = v \circ (\mathrm{id} + tv)^{-1}$, while the velocity field for the latter is time-independent, and equal to v.

Let us focus on the case of the internal energy functional $\mathbb{U}$ associated to a density U. As a consequence of the previous remarks, defining $s := \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\varrho_t$, we can compute

$$\begin{aligned} \mathrm{d}_\varrho \mathbb{U}(s) &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathbb{U}(\varrho_t) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \left(\int_{\mathbb{R}^n} U(\varrho_t)\,\mathrm{d}x \right) \\ &= -\int_{\mathbb{R}^n} U'(\varrho)\,\mathrm{div}(v\varrho)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^n} \langle \nabla U'(\varrho), v \rangle \varrho\,\mathrm{d}x\,. \end{aligned}$$

This last expression allows us to identify $\nabla^W \mathbb{U}(\varrho)$ with $\nabla U'(\varrho)$, since with this choice we recover

$$\langle \nabla^W \mathbb{U}(\varrho), s \rangle_\varrho = \mathrm{d}_\varrho \mathbb{U}(s),$$

which is the relation that a Riemannian gradient should satisfy.

With analogous computations one can formally compute the Wasserstein gradients also for the potential energy $\mathbb{V}$ (induced by a density V) and for the interaction energy $\mathbb{W}$ (induced by a density W), obtaining

$$\nabla^W \mathbb{V}(\varrho) = \nabla V \qquad \text{and} \qquad \nabla^W \mathbb{W}(\varrho) = \nabla\,(W * \varrho)\,(= (\nabla W) * \varrho)\,.$$

Combining all these ingredients one can consider the general case of an energy

$$\mathcal{E}(\varrho) = \int_{\mathbb{R}^n} (U(\varrho) + V\varrho + (W * \varrho)\,\varrho)\,\mathrm{d}x$$

and the discussion above allows us to formally interpret the gradient flow of $\mathcal{E}$ in $(\mathscr{P}_2(\mathbb{R}^n), W_2)$ as a solution of the following evolution equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho_t = \operatorname{div}\big((\nabla U'(\varrho_t) + \nabla V + \nabla(W * \varrho_t))\varrho_t\big). \tag{18.5}$$

We observe that the right hand side of (18.5) is given by a combination of a diffusion term (induced by the internal energy), a transport term (associated to the potential energy) and an interaction term (induced by the interaction energy).

The rest of this section is devoted to the discussion of a few explicit examples: the heat equation, the Fokker–Planck equation, the porous medium equation and the system of interacting particles.

Example 18.1 (The Heat Equation) If we consider the Wasserstein gradient flow associated to the Shannon-Boltzmann logarithmic entropy Ent (i.e. the internal energy associated to the density $U(\varrho) = \varrho \ln(\varrho)$) we can observe that

$$U'(\varrho) = 1 + \ln(\varrho) \quad \text{and} \quad \nabla U'(\varrho) = \frac{\nabla\varrho}{\varrho}.$$

Thus, the formal computation of the Wasserstein gradient we presented above allows us to interpret (up to the usual identification between densities ϱ and measures μ) the heat flow as gradient flow of Ent, since

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho_t = \Delta\varrho_t = \operatorname{div}(\nabla\varrho_t) = \operatorname{div}(\frac{\nabla\varrho_t}{\varrho_t}\varrho_t) = \operatorname{div}(\nabla^W \mathrm{Ent}(\varrho_t)\varrho_t).$$

Let us briefly recapitulate the interpretations of the heat equation as a gradient flow we introduced so far:

(i) in Lecture 13 we proved that the heat flow can be interpreted as gradient flow of the Dirichlet energy $\int_{\mathbb{R}^n} |\nabla\varrho|^2$ with respect to the L^2 metric;
(ii) in the same Lecture we proved that it can also be interpreted as gradient flow of the energy $\int_{\mathbb{R}^n} \varrho^2$ with respect to the H^{-1} metric;
(iii) we just proved (formally, for the moment) that the heat flow admits an interpretation as gradient flow of Ent with respect to the Wasserstein metric.

Example 18.2 (The Fokker–Planck Equation) Let us consider the energy $\mathcal{E}$ defined by $\mathcal{E}(\varrho) := \mathrm{Ent}(\varrho) + \int_{\mathbb{R}^n} V\varrho\,\mathrm{d}x$, where V is any smooth potential such that $\int_{\mathbb{R}^n} e^{-V} < \infty$. Then, thanks to the discussion above,

$$\nabla^W \mathcal{E}(\varrho) = \frac{\nabla\varrho}{\varrho} + \nabla V,$$

thus we can say that any gradient flow ϱ_t of $\mathcal{E}$ in $\mathscr{P}$ with respect to the Wasserstein metric is a solution of the Fokker–Planck equation we introduced in Lecture 13 (see Eq. (13.3)). Indeed

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho_t = \operatorname{div}\big((\frac{\nabla\varrho_t}{\varrho_t} + \nabla V)\varrho_t\big) = \Delta\varrho_t + \operatorname{div}\big((\nabla V)\varrho_t\big).$$

To let the picture be more complete we recall that we already introduced an interpretation of the Fokker–Planck equation as a gradient flow in Lecture 13. We also remark that, according to (15.5) in Proposition 15.6, we can interpret $\mathcal{E}$ as the relative entropy Ent_γ, where $\gamma = e^{-V}\mathscr{L}^n$, hence we can see (FP) as the gradient flow of Ent_γ with respect to the Wasserstein metric (still at a formal level).

At this point of the discussion some questions arise: why are these new interpretations relevant? And can we get anything out of them?

First of all we remark that the interpretations presented above and their many variants provide new contractivity estimates and rates of convergence for many partial differential equations (starting from the seminal paper [95], see also [12] for more details about this topic). We also point out that in the Wasserstein interpretation there is a clear separation between metric and measure (i.e. the entropy functional depends only on the reference measure, while the Wasserstein metric depends only on the underlying distance of the ambient space). In the Hilbertian interpretation presented in Lecture 13 instead, both the distance and the reference measure are involved in the definition of the Dirichlet energy.

To conclude this discussion, let us consider the specific case of the Fokker–Planck equation (FP) when γ is the Gaussian measure: solutions converge to the equilibrium γ as $t \to \infty$. But, while the interpretation of (FP) as an Hilbertian gradient flow in $L^2(\gamma)$ does not provide a rate of convergence (because of the lack of uniform convexity in L^2), in the W_2 picture the exponential rate of convergence holds, and it comes as a consequence the 1-convexity of the relative entropy Ent_γ guaranteed by (15.5). In particular, this provides the estimate

$$W_2(w_t\gamma, \gamma) \le e^{-t} W_2(w_0\gamma, \gamma) \qquad \forall t \ge 0.$$

Example 18.3 (The Porous-Medium Equation) Let us consider for any $m \neq 1$ the internal energy $\mathbb{U}(\varrho) := \int_{\mathbb{R}^n} \frac{1}{m-1}\varrho^m$. With this choice we have that

$$U(\varrho) = \frac{\varrho^m}{m-1}, \quad U'(\varrho) = \frac{m}{m-1}\varrho^{m-1} \quad \text{and} \quad \nabla U'(\varrho) = m\varrho^{m-2}\nabla\varrho = \frac{\nabla\varrho^m}{\varrho},$$

therefore we can say that any W_2 gradient flow of $\mathbb{U}$ is indeed a solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho_t = \operatorname{div}(\frac{\nabla{\varrho_t}^m}{\varrho_t}\varrho_t) = \Delta\varrho_t^m,$$

that is the porous-medium equation we already discussed in Lecture 13.

Example 18.4 (A System of Interacting Particles) We are given a system of N particles $x_1, \dots, x_N \in \mathbb{R}^n$ with weights $m_1, \dots, m_N$ such that $\sum m_i = 1$ and a (smooth) symmetric potential $W : \mathbb{R}^n \to \mathbb{R}$. Assume for simplicity that $\nabla W(0) = 0$ (even though this assumption is not realistic for many applications) and that the particles evolve according to the following system of first order ordinary differential equations:

$$x_i'(t) = \sum_{j \neq i} m_j \nabla W(x_i(t) - x_j(t)) \tag{18.6}$$

$$= \sum_j m_j \nabla W(x_i(t) - x_j(t)) \qquad i = 1, \dots, N. \tag{18.7}$$

If we introduce the time-dependent empirical measures

$$\mu_t := \sum_{i=1}^N m_i \delta_{x_i(t)},$$

then (18.6) can be rephrased in the following terms

$$x_i'(t) = ((\nabla W) * \mu_t)\,(x_i(t)) \qquad i = 1, \dots, N. \tag{18.8}$$

Thanks to the linearity with respect to μ of the continuity equation and the duality results between solutions of (ODE) and (CE) we explored in Lecture 16, we can conclude that the empirical measures μ_t solve the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t = \operatorname{div}(((\nabla W) * \mu_t)\,\mu_t)$$

and finally that the system can be interpreted as a gradient flow of the interaction energy $\mathbb{W}$ associated to the potential W.

Moving our attention to the second order problem

$$m_i x_i''(t) = \sum_{i \neq j} m_i m_j \nabla W(x_i(t) - x_j(t)) \qquad i = 1, \dots, N, \tag{18.9}$$

after the introduction of the empirical measure

$$\tilde{\mu}_t = \sum_{i=1}^N m_i \delta_{(x_i(t), x_i'(t))}$$

(that is a measure in the phase space with variables (x, p) this time), we can see that also this system evolves according to the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mu}_t = \mathrm{div}(b_t\tilde{\mu}_t),$$

where now the velocity field b_t is given by

$$b_t(x, p) := (p, \nabla W * \mu_t(x)).$$

In this Hamiltonian framework it is no more possible to get a gradient flow interpretation of (18.9), since the vector field b is not a gradient.

3 Rigorous Interpretation of the Heat Equation as a Wasserstein Gradient Flow

In this section we provide a rigorous interpretation of the heat equation on $\mathbb{R}^n$ as gradient flow of the entropy functional in the Wasserstein space $(\mathscr{P}_2(\mathbb{R}^n), W_2)$. In this connection, it is worth to remember the paper [72] which provides another justification of this fact based on the implicit Euler scheme.

We already know from Theorem 15.16 that Ent is geodesically convex on $\mathscr{P}_2(\mathbb{R}^n)$, thanks to the fact that the energy density $U(\varrho) = \varrho \ln \varrho$ satisfies McCann's condition (MC). Observe also that, as we already pointed out, to give a meaning to the EVI formulation of gradient flows we just need the metric structure of the underlying space: the next results show exactly that the heat flow is an EVI gradient flow of the entropy functional.

Let us recall that, for the heat equation in $\mathbb{R}^n$ with initial datum $\bar{\varrho} \in L^1(\mathbb{R}^n)$, one has the explicit expression for the solution

$$\varrho_t(x) = \int_{\mathbb{R}^n} p_t(x, y)\bar{\varrho}(y)\,\mathrm{d}y, \tag{18.10}$$

where

$$p_t(x, y) := (4\pi t)^{-\frac{n}{2}} \exp(-\frac{1}{4t}|x - y|^2) \tag{18.11}$$

is the so called heat kernel. Assuming that $\bar{\varrho}$ is a probability density with finite quadratic moments, the aim of this section is to check that ϱ_t, seen as a curve $\mu_t = \varrho_t \mathscr{L}^n$ in $\mathscr{P}_2(\mathbb{R}^n)$, is an EVI gradient flow. We will use in part the explicit expression (18.10) of ϱ_t to justify some estimates and computations. First of all, one can use the explicit expression to show that the properties of $\bar{\varrho}$, namely being a probability density with finite quadratic moments, are preserved by the heat flow. Therefore it does make sense to consider the curve $\mu_t = \varrho_t \mathscr{L}^n$ as a curve in $\mathscr{P}_2(\mathbb{R}^n)$.

Theorem 18.5 *Let $\bar{\varrho} \in L^1(\mathbb{R}^n)$ be nonnegative, such that*

$$\int_{\mathbb{R}^n} \bar{\varrho}(x)\,\mathrm{d}x = 1, \qquad \int_{\mathbb{R}^n} |x|^2 \bar{\varrho}(x)\,\mathrm{d}x < \infty \quad \textit{and} \quad \mathrm{Ent}(\bar{\varrho}) < \infty.$$

Then $\mu_t = \varrho_t \mathscr{L}^n$ solves EVI *in* $(\mathscr{P}_2(\mathbb{R}^n), W_2)$.

The proof of Theorem 18.5 will come as a consequence of some intermediate results. First of all, in order to give sense to the EVI formulation of gradient flows, we need to check that μ_t is (locally) absolutely continuous with respect to the Wasserstein metric.

Proposition 18.6 *Under the same assumptions of Theorem 18.5 the curve $t \mapsto \mu_t$ belongs to $AC^2_{\mathrm{loc}}((0,\infty); \mathscr{P}_2(\mathbb{R}^n))$.*

Proof We start from the observation that, as a consequence of the explicit expression of ϱ_t, we can say that $\varrho_t \in C^\infty(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ for any $p \in [1,\infty]$, that $\varrho_t > 0$ and that the map $t \mapsto \varrho_t$ is continuously differentiable in $(0,\infty)$, as a $L^2(\mathbb{R}^n)$-valued map, with

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho_t = \Delta \varrho_t \qquad \forall t \in (0,\infty). \tag{18.12}$$

Thus, if we define

$$v_t := -\frac{\nabla \varrho_t}{\varrho_t}$$

it follows that the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \mathrm{div}(v_t \mu_t) = 0 \tag{18.13}$$

holds in the sense of distributions.

Therefore, in order to prove the W_2-absolute continuity of the curve μ_t, it will be sufficient to apply Proposition 17.9, after proving that $\|v_t\|^2_{L^2(\mu_t)}$ is locally integrable in $(0,\infty)$. Using the explicit expression

$$v_t(x) = -\frac{\int_{\mathbb{R}^n} \nabla_x p_t(x,y)\bar{\varrho}(y)\,\mathrm{d}y}{\int_{\mathbb{R}^n} p_t(x,y)\bar{\varrho}(y)\,\mathrm{d}y}$$

and arguing as in (17.6) one can prove that

$$\int_{\mathbb{R}^n} |v_t|^2(x)\,\mathrm{d}\mu_t(x) \le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|\nabla_x p_t(x,y)|^2}{p_t(x,y)}\bar{\varrho}(y)\,\mathrm{d}x\,\mathrm{d}y$$

so that $\|v_t\|_{L^2(\mu_t)} \in L^\infty(\epsilon, \infty)$ for any $\epsilon > 0$. However, we prefer to provide another proof of integrability, inspired by Ambrosio et al. [14] (dealing with the theory of heat flow in metric measure spaces) which provides a space-time sharp estimate of $\|v_t\|^2_{L^2(\mu_t)}$ very much in the spirit of the EDI theory of gradient flows.

To this aim, for $\epsilon \in (0, e^{-1})$, in order to take care of the singularity at $r = 0$ of the density $e(r) = r \ln r$, we define the regularized energy densities $e_\epsilon(r)$ as follows:

$$\begin{cases} e_\epsilon(r) = (1 + \ln \epsilon)r & \text{in } [0, \epsilon] \\ e_\epsilon(r) = r \ln r + \epsilon & \text{in } [\epsilon, \infty). \end{cases}$$

Observe that $e(r) \le e_\epsilon(r) \le \frac{1}{2}r^2$, $e_\epsilon(r) \downarrow e(r)$ as ϵ goes to 0 and $e_\epsilon \in C^{1,1}(0, \infty)$. Then extend e_ϵ to $C^{1,1}(\mathbb{R})$ convex densities $\tilde{e}_\epsilon$ writing $\tilde{e}_\epsilon(r) = e_\epsilon(r) - (1 + \ln \epsilon)r$ for $r \ge 0$ and $\tilde{e}_\epsilon(r) = 0$ for $r < 0$. Observe that $\tilde{e}_\epsilon(\bar{\varrho}) \in L^1(\mathbb{R}^n)$ and one can prove that the map

$$t \mapsto \int_{\mathbb{R}^n} \tilde{e}_\epsilon(\varrho_t(x))\, \mathrm{d}x$$

is locally Lipschitz on $(0, \infty)$, with the explicit expression for its derivative provided by (18.12) and an integration by parts:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^n} \tilde{e}_\epsilon(\varrho_t(x))\, \mathrm{d}x = \int_{\mathbb{R}^n} \tilde{e}'_\epsilon(\varrho_t)\Delta\varrho_t\, \mathrm{d}x = -\int_{\mathbb{R}^n} \tilde{e}''_\epsilon(\varrho_t(x))|\nabla\varrho_t(x)|^2\, \mathrm{d}x. \tag{18.14}$$

Recalling that the total mass is preserved by the heat flow, an integration in time of (18.14) yields to

$$\int_{\mathbb{R}^n} e_\epsilon(\varrho_T(x))\, \mathrm{d}x + \int_0^T \left(\int_{\{\varrho_t > \epsilon\}} \frac{|\nabla\varrho_t(x)|^2}{\varrho_t(x)}\, \mathrm{d}x \right) \mathrm{d}t = \int_{\mathbb{R}^n} e_\epsilon(\bar{\varrho}(x))\, \mathrm{d}x. \tag{18.15}$$

Passing to the limit as $\epsilon \to 0$ in (18.15) we conclude that

$$\mathrm{Ent}(\mu_T) + \int_0^T \left(\int_{\{\varrho_t > 0\}} \frac{|\nabla\varrho_t(x)|^2}{\varrho_t(x)}\, \mathrm{d}x \right) \mathrm{d}t = \mathrm{Ent}(\bar{\mu}) \qquad \forall T \ge 0,$$

yielding in particular the stated local integrability in time of $\|v_t\|^2_{L^2(\mu_t)}$, and concluding the proof. □

Remark 18.7 (The Heat Equation as an EDE *Gradient Flow)* Notice that, as stated in Theorem 15.26, computations analogous to those in Theorem 15.25 (dealing with the slightly simpler case of the relative entropy with respect to a Gaussian) show that $\int_{\{\varrho_t>0\}} \frac{|\nabla\varrho_t(x)|^2}{\varrho_t(x)}\,\mathrm{d}x$ coincides with the slope of the Entropy at μ_t. In addition, the same quantity provides an upper bound for the metric derivative of the curve μ_t, thanks to Proposition 17.9. Therefore (18.15) implies

$$\mathrm{Ent}(\mu_T) + \int_0^T \frac{1}{2}|\mu_t'|^2 + \frac{1}{2}|\nabla^-\mathrm{Ent}|^2(\mu_t)\,\mathrm{d}t \le \mathrm{Ent}(\bar\mu) \qquad \forall T \ge 0,$$

namely the EDI formulation of gradient flows. However, our goal in this section is to achieve the stronger EVI property.

The forthcoming Lemmas 18.8 and 18.10 encode the information about the "Riemannian-like" behaviour of the Wasserstein distance and the geodesic convexity of the entropy functional, respectively.

Lemma 18.8 *If $\nu \in \mathscr{P}_2(\mathbb{R}^n)$ has compact support then*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\mu_t,\nu) = \int_{\mathbb{R}^n} \langle T_{\mu_t}^\nu - \mathrm{id}, \nabla\varrho_t\rangle\,\mathrm{d}x \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (0,\infty), \tag{18.16}$$

where $T_{\mu_t}^\nu$ is the unique optimal transport map from μ_t to ν (with quadratic cost), whose existence follows from Theorem 5.2.

Proof Thanks to Proposition 18.6 we know that μ_t is locally absolutely continuous on $(0,\infty)$ with respect to the W_2 distance. It follows in particular that the map

$$t \mapsto \frac{1}{2}W_2^2(\mu_t,\nu)$$

is differentiable $\mathscr{L}^1$-a.e. on $(0,\infty)$. From now on we fix such a differentiability point t and we use the fact, already pointed out in the proof of Proposition 18.6, that (18.12) holds in the strong L^2 sense. The differentiability of $s \mapsto W_2^2(\mu_s,\nu)$ at $s = t$ justifies the following expansion:

$$\frac{1}{2}W_2^2(\mu_{t+h},\nu) - \frac{1}{2}W_2^2(\mu_t,\nu) = \frac{h}{2}\frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=t} W_2^2(\mu_s,\nu) + o(h) \quad (h \to 0)\,. \tag{18.17}$$

Thanks to the compactness assumption on ν we can find a Lipschitz Kantorovich potential φ from μ_t to ν. From the optimality of φ we deduce that

$$\frac{1}{2}W_2^2(\mu_t,\nu) = \int_{\mathbb{R}^n} \varphi\,\mathrm{d}\mu_t + \int_{\mathbb{R}^n} Q_1(-\varphi)\,\mathrm{d}\nu,$$

while

$$\frac{1}{2}W_2^2(\mu_{t+h},\nu) \ge \int_{\mathbb{R}^n} \varphi\,\mathrm{d}\mu_{t+h} + \int_{\mathbb{R}^n} Q_1(-\varphi)\,\mathrm{d}\nu.$$

Hence

$$\begin{aligned}\frac{1}{2}W_2^2(\mu_{t+h},\nu)-\frac{1}{2}W_2^2(\mu_t,\nu) &\geq \int_{\mathbb{R}^n}\varphi\,\mathrm{d}(\mu_{t+h}-\mu_t)\\ &= h\int_{\mathbb{R}^n}\varphi\Delta\varrho_t\,\mathrm{d}x+o(h)\quad (h\to 0) \qquad (18.18)\\ &= -h\int_{\mathbb{R}^n}\langle\nabla\varphi,\nabla\varrho_t\rangle\,\mathrm{d}x+o(h)\quad (h\to 0),\end{aligned}$$

where in the last two passages we used (18.12) and integrated by parts, respectively. Notice that the integration by parts is justified by the boundedness of $\nabla\varphi$ and by the fact that all derivatives of ϱ_t are rapidly decreasing at infinity.

By comparing (18.18) with (18.17) we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=t} W_2^2(\mu_s,\nu) = -\int_{\mathbb{R}^n}\langle\nabla\varphi,\nabla\varrho_t\rangle\,\mathrm{d}x$$

and, after recalling the explicit expression $T_{\mu_t}^{\nu}=\mathrm{id}-\nabla\varphi$ of the optimal transport map in terms of the Kantorovich potential φ, we get the desired conclusion. □

Remark 18.9 The heuristic interpretation of Lemma 18.8 is that if you want to differentiate the square of the distance form a fixed point q along a curve σ on a Riemannian manifold (at a differentiability point $p=\sigma(t)$) you just need to couple with the metric tensor the speed of the curve $\dot\sigma(t)$ and the speed $\dot\gamma(0)$ of a geodesic γ such that $\gamma(0)=p$ and $\gamma(1)=q$. In the case of our interest $T_{\mu_t}^{\nu}-\mathrm{id}$ is the speed at time 0 of the geodesic joining μ_t with ν, while the speed of the curve μ_t should equal $\frac{\nabla\varrho_t}{\varrho_t}$ if we expect the heat flow to be a gradient flow of the entropy functional. Thus at the right hand-side of (18.16) we recover exactly their scalar product with respect to the "Riemannian metric" of $\mathscr{P}_2(\mathbb{R}^n)$ at the point μ_t.

Lemma 18.10 *If $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^n)$ are such that ν has compact support and $\mu=\varrho\mathscr{L}^n$ with $\varrho\in C^\infty(\mathbb{R}^n)$, bounded and with a bounded integrable gradient, then*

$$\mathrm{Ent}(\nu)\geq \mathrm{Ent}(\mu)+\int_{\mathbb{R}^n}\langle T_\mu^\nu-\mathrm{id},\nabla\varrho\rangle\,\mathrm{d}x. \qquad (18.19)$$

In particular this holds with $\varrho=\varrho_t$ for all $t>0$.

Proof Consider the unique constant speed W_2 geodesic μ_s joining $\mu_0=\mu$ with $\mu_1=\nu$ and recall that, defining $T:=T_\mu^\nu$ and $T_s:=(1-s)T+s\mathrm{id}$, we have $\mu_s=(T_s)_\#\mu$. Thanks to Theorem 15.16, which yields the geodesic convexity of Ent, we have that

$$\mathrm{Ent}(\nu)-\mathrm{Ent}(\mu)\geq\frac{\mathrm{Ent}(\mu_s)-\mathrm{Ent}(\mu)}{s}\quad \text{for any } s\in(0,1],$$

thus, in order to get the desired conclusion, it suffices to prove that

$$\liminf_{s\to 0^+} \frac{\mathrm{Ent}(\mu_s) - \mathrm{Ent}(\mu)}{s} \geq \int_{\mathbb{R}^n} \langle T - \mathrm{id}, \nabla \varrho \rangle \, \mathrm{d}x.$$

To this aim we recall that Proposition 15.14 provides the explicit expression for the density ϱ_s of the interpolating measure μ_s, namely

$$\varrho_s = \left(\frac{\varrho}{\det \nabla T_s} \right) \circ T_s^{-1},$$

so that an application of the change of variables formula yields

$$\mathrm{Ent}(\mu_s) = \int_{\mathbb{R}^n} \ln \left(\frac{\varrho}{\det \nabla T_s} \right) \varrho \, \mathrm{d}x. \tag{18.20}$$

Thanks to Brenier's theorem and the compactness of the support of ν, T is given by ∇f, where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex Lipschitz function. Moreover, by Alexandrov's Theorem 6.4, f is twice differentiable $\mathscr{L}^n$-a.e. (and then μ-a.e.), and also in the sense of distributions. Denoting by div the pointwise divergence and by Div the distributional divergence operators, as in the proof of the isoperimetric inequality we use the inequality between measures

$$\mathrm{Div}\, T = \mathrm{tr}\, D^2 f \geq \mathrm{tr}\, \nabla^2 f \mathscr{L}^n = \mathrm{div}\, T \mathscr{L}^n .$$

The inequality above comes from Theorem 6.4, since the nonnegativity of $D^2 f$ guarantees also the nonnegativity of the singular part of $D^2 f$ with respect to the Lebesgue measure, whose trace is the difference $\mathrm{tr}\, D^2 f - \mathrm{tr}\, \nabla^2 f \mathscr{L}^n$. We conclude that, in the sense of distributions,

$$\mathrm{Div}(T - \mathrm{id}) \geq \mathrm{div}(T - \mathrm{id}) \mathscr{L}^n . \tag{18.21}$$

Alexandrov's theorem ensures us also that the first order expansion

$$\det \nabla T_s(x) = 1 + s\, \mathrm{div}(T - \mathrm{id})(x) + o(s) \quad (s \to 0) \tag{18.22}$$

holds for μ-a.e. $x \in \mathbb{R}^n$.

Applying Fatou's lemma and taking into account (18.22), (18.20) and (18.21), we get

$$\begin{aligned} \liminf_{s\to 0^+} \frac{\mathrm{Ent}(\mu_s) - \mathrm{Ent}(\mu)}{s} &\geq -\int_{\mathbb{R}^n} \mathrm{div}(T - \mathrm{id}) \varrho \, \mathrm{d}x \\ &\geq \langle \mathrm{Div}(T - \mathrm{id}), \varrho \rangle \\ &= \int_{\mathbb{R}^n} \langle T - \mathrm{id}, \nabla \varrho \rangle \, \mathrm{d}x. \end{aligned}$$

Notice that the first inequality above is justified by boundedness and continuity of ϱ. The second one can be justified by approximating ϱ with the family $\varrho_R = \varrho\psi(\cdot/R)$, where $\psi \in C_c^\infty(\mathbb{R}^n)$ is identically equal to 1 on the unit ball. □

Remark 18.11 As for Lemma 18.8, also Lemma 18.10 admits a Riemannian interpretation. Indeed, for a geodesically convex (smooth) function f on a Riemannian manifold M, one has

$$f(q) \geq f(p) + \langle \nabla f(p), \dot{\gamma}(0) \rangle,$$

where $p, q \in M$ and $\gamma : [0, 1] \to M$ is a geodesic joining p to q. As before, at the right hand-side of (18.19), we can identify all the ingredients which appear in the smooth context, since $\nabla^W \mathrm{Ent}(\mu) = \frac{\nabla \varrho}{\varrho}$ and we already observed that $T_\mu^\nu - \mathrm{id}$ is the speed at time 0 of the geodesic joining μ to ν.

Proof of Theorem 18.5 Applying Lemmas 18.8 and 18.10, we obtain that

$$\mathrm{Ent}(\nu) \geq \mathrm{Ent}(\mu_t) + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\mu_t, \nu)$$

for any $\nu \in \mathscr{P}_2(\mathbb{R}^n)$ with compact support. By a simple approximation procedure we can drop the compactness assumption on the support of ν and conclude that μ_t is the EVI gradient flow of Ent starting from $\bar{\mu}$. □

4 More Recent Ideas and Developments

Here and throughout the last lecture we will briefly present some more recent ideas and developments about Optimal Transport.

Contractivity via Action Estimates We first want to illustrate the idea, appeared in [97] and then refined in [44], that one can obtain contractivity estimates on a semigroup S via action estimates (see also [24] for similar ideas in the context of semigroups induced by conservation laws, where contractivity occurs with respect to L^1-like distances).

We are given a length space (X, d) and a semigroup S_t, $t \geq 0$, on X. We recall the definition of quadratic action of an absolutely continuous curve $\gamma : [0, 1] \to X$

$$\mathcal{A}(\gamma) = \frac{1}{2}\int_0^1 |\gamma'(s)|^2 \,\mathrm{d}s.$$

If we assume that the action estimate

$$\mathcal{A}(\gamma^t) \leq e^{-2kt}\mathcal{A}(\gamma) \qquad t \geq 0 \tag{18.23}$$

holds, where γ^t is the deformation of the curve γ induced by S_t, namely $\gamma^t(s) := S_t\gamma(s)$, then one can prove that S is k-contractive, namely

$$d(S_t x, S_t y) \le e^{-kt} d(x, y) \qquad \forall x,\ y \in X \quad \text{and} \quad \forall t \ge 0. \tag{18.24}$$

Indeed, for any absolutely continuous curve γ joining $\gamma(0) = x$ to $\gamma(1) = y$, the curve γ^t provides an admissible curve between $S_t x$ and $S_t y$, thus

$$\frac{1}{2} d^2(S_t x, S_t y) \le \mathcal{A}(\gamma^t) \le e^{-2kt}\mathcal{A}(\gamma).$$

Therefore, since (X, d) is a length space, passing to the infimum among all the admissible curves γ joining x to y, and taking the square roots of both sides, we conclude that (18.24) holds. Notice also that, conversely, contractivity immediately implies the action estimate (18.23), so that the two properties are equivalent.

We illustrate the validity of the method in the particular case of the heat semigroup in $\mathbb{R}^n$, with $(X, d) = (\mathscr{P}_2(\mathbb{R}^n), W_2)$. This example is also propedeutic to the next lecture, where we are going to investigate the deep relations between contractivity of the heat flow in the space of probability measures and many other geometric and analytic concepts.

Let $p_t(x, y)$ be the Euclidean heat kernel in (18.11) and, as in Sect. 3, define the heat flow starting from $f \in L^1(\mathbb{R}^n)$ with the classical formula:

$$P_t f(x) := \int_{\mathbb{R}^n} p_t(x, y) f(y)\, \mathrm{d}y \qquad t > 0. \tag{18.25}$$

We already remarked in that Section that, since sign and total mass are preserved by the evolution, the heat flow can also be seen as an evolution problem in the space of absolutely continuous probability measures with finite quadratic moments, when identified with their densities.

In addition, one has

$$p_t(x, \cdot) \text{ is a Gaussian probability density for all } (t, x) \in (0, \infty) \times X \tag{18.26}$$

and it is easily seen that, in measure theoretic terms, (18.25) can be read as

$$P_t f \mathscr{L}^n = \int_{\mathbb{R}^n} \Big(p_t(x, \cdot)\mathscr{L}^n \Big) f(x)\, \mathrm{d}x. \tag{18.27}$$

Lemma 18.12 *Let f, g be probability densities of measures in $\mathscr{P}_2(\mathbb{R}^n)$. Then*

$$W_2(P_t f \mathscr{L}^n, P_t g \mathscr{L}^n) \le W_2(f \mathscr{L}^n, g \mathscr{L}^n). \tag{18.28}$$

Remark 18.13 (Alternative Proofs) We notice that one could try to prove this result directly using the identity (18.27) for $P_t g\mathscr{L}^n$. Indeed, if T is the optimal map from $f\mathscr{L}^n$ to $g\mathscr{L}^n$, one has

$$P_t g\mathscr{L}^n = \int_{\mathbb{R}^n} \Big(p_t(y,\cdot)\mathscr{L}^n\Big) g(y)\,\mathrm{d}y = \int_{\mathbb{R}^n} \Big(p_t(T(x),\cdot)\mathscr{L}^n\Big) f(x)\,\mathrm{d}x.$$

This, together with the observation that $W_2(p_t(x,\cdot)\mathscr{L}^n, p_t(T(x),\cdot)\mathscr{L}^n) = |T(x)-x|$ (since the optimal map is the shift by the vector $T(x)-x$) provides the contractivity property arguing as in (19.22).

Furthermore, since the heat flow satisfies the EVI property (see once more Sect. 3) another, more abstract, proof would follow by the general theory of EVI gradient flows in metric spaces. Indeed, in Lemma 11.13 we proved that EVI_K gradient flows on Hilbert spaces are K-contractive and in the last part of the proof we already pointed out that the conclusion holds even for general metric spaces.

Here we give instead a proof where the monotonicity of the action is employed.

Proof of Lemma 18.12 Let ϱ_s be a geodesic in $\mathscr{P}_2(\mathbb{R}^n)$ from $f\mathscr{L}^n$ to $g\mathscr{L}^n$. If the equation

$$\frac{\mathrm{d}}{\mathrm{d}s}\varrho_s + \mathrm{div}(v_s\varrho_s) = 0,$$

is satisfied for some velocity field v_s then, letting $\varrho_s^t = P_t\varrho_s$ and $\mu_s^t = \varrho_s^t\mathscr{L}^n$ (which is a curve from $P_t f\mathscr{L}^n$ to $P_t g\mathscr{L}^n$), we would like to define a velocity field v_s^t in such a way that

$$\frac{\mathrm{d}}{\mathrm{d}s}\varrho_s^t + \mathrm{div}(v_s^t\varrho_s^t) = 0\,. \tag{18.29}$$

Then Proposition 17.9 would give

$$W_2^2(P_t f\mathscr{L}^n, P_t g\mathscr{L}^n) \le \int_0^1 |(\mu_s^t)'|^2\,\mathrm{d}s \le \int_0^1\int_{\mathbb{R}^n} |v_s^t|^2\varrho_s^t\,\mathrm{d}x\,\mathrm{d}s. \tag{18.30}$$

Moreover, as illustrated in Sect. 4, if our vector field v_s^t satisfies the (pointwise) action monotonicity property

$$\int_{\mathbb{R}^n} |v_s^t|^2\varrho_s^t\,\mathrm{d}x \le \int_{\mathbb{R}^n} |v_s|^2\varrho_s\,\mathrm{d}x, \tag{18.31}$$

for all $s\in(0,1)$, by integration of (18.30) we get

$$W_2^2(P_t f\mathscr{L}^n, P_t g\mathscr{L}^n) \le \int_0^1\int_{\mathbb{R}^n} |v_s|^2\varrho_s\;\mathrm{d}x\,\mathrm{d}s.$$

Finally, choosing the optimal velocity field v_s, see Theorem 17.10, we get (18.28).

Thus, we need to determine a velocity field v_s^t with properties (18.29) and (18.31). Understanding the action of the semigroup on vector fields componentwise, we have

$$-\operatorname{div}(v_s^t \varrho_s^t) = \frac{\mathrm{d}}{\mathrm{d}s}\varrho_s^t = P_t \frac{\mathrm{d}}{\mathrm{d}s}\varrho_s = -P_t \operatorname{div}(v_s \varrho_s) = -\operatorname{div}(P_t(v_s\varrho_s)). \tag{18.32}$$

Therefore any admissible velocity field v_t^t should satisfy

$$v_s^t = \frac{P_t(v_s\varrho_s)}{P_t(\varrho_s)} = \frac{P_t(v_s\varrho_s)}{\varrho_s^t} \tag{18.33}$$

up to a vector field G/ϱ_s^t, with G solenoidal. Choosing v_s^t exactly as in (18.33), we can now check that (18.31) holds. Indeed, let $\psi \in C_c(\mathbb{R}^n;\mathbb{R}^n)$ be a test function. Using the fact that P_t is self-adjoint in $L^2(\mathbb{R}^n)$, we have

$$\begin{aligned}\left|\int_{\mathbb{R}^n}\langle v_s^t, \psi\rangle \varrho_s^t\right| = \left|\int_{\mathbb{R}^n}\langle P_t(v_s\varrho_s), \psi\rangle\right| = \left|\int_{\mathbb{R}^n}\langle v_s, P_t\psi\rangle \varrho_s\right| \\ \le \|v_s\|_{L^2(\varrho_s\mathscr{L}^n)}\|P_t\psi\|_{L^2(\varrho_s\mathscr{L}^n)}.\end{aligned} \tag{18.34}$$

Hence, taking (18.27) into account, by Jensen inequality we obtain

$$\int_{\mathbb{R}^n}(P_t\psi)^2\varrho_s\,\mathrm{d}x \le \int_{\mathbb{R}^n} P_t\psi^2\varrho_s\,\mathrm{d}x = \int_{\mathbb{R}^n}\psi^2\varrho_s^t\,\mathrm{d}x = \|\psi\|^2_{L^2(\varrho_s^t\mathscr{L}^n)}. \tag{18.35}$$

Thus, by (18.34) and (18.35) and by duality, we obtain

$$\|v_s^t\|_{L^2(\varrho_s^t\mathscr{L}^n)} \le \|v_s\|_{L^2(\varrho_s\mathscr{L}^n)}.$$

□

When we move from the Euclidean space to Riemannian manifolds (and even more general spaces) we lose the commutation property between divergence and semigroup crucially used in (18.32). Therefore curvature comes into play and a deeper analysis is necessary: this will be the topic of the last lecture.

Convexity from EVI The following result, taken from [44], conveys the idea that any energy admitting an EVI gradient flow is automatically (geodesically-)convex. We state the result in a simplified form, see the original paper for the more general and useful statement dealing with lower semicontinuous functions with values in $(-\infty, \infty]$.

Theorem 18.14 (EVI **Implies Convexity**) *Let (X, d) be a geodesic space and let $F : X \to \mathbb{R}$ be a lower semicontinuous energy functional. Assume there exists an* EVI *gradient flow S_t of F starting from any $x \in X$. Then F is convex along all geodesics.*

Sketch of Proof Fix any $\gamma \in \mathrm{Geo}(X)$ and any intermediate time $s \in (0, 1)$. Define, as before, $\gamma^t : [0, 1] \to X$ by $\gamma^t(s) := S_t\gamma(s)$. Here, as in Lemma 11.13, we can use the lower semicontinuity of F to write the EVI differential inequality in a pointwise sense, namely

$$F(w) \geq F(S_r u) + \frac{\mathrm{d}}{\mathrm{d}t}^{+}\Big|_{t=r} d^2(S_t u, w) \qquad \forall r \geq 0,$$

where $\frac{\mathrm{d}}{\mathrm{d}t}^{+}$ denotes the upper right derivative. Applying EVI with $u = \gamma(s)$, $r = 0$ and with test points $w = \gamma(0)$, $w = \gamma(1)$ we obtain that

$$F(\gamma(0)) \geq \frac{\mathrm{d}}{\mathrm{d}t}^{+}\Big|_{t=0} \frac{1}{2} d^2(\gamma^t(s), \gamma(0)) + F(\gamma(s)) \tag{18.36}$$

and

$$F(\gamma(1)) \geq \frac{\mathrm{d}}{\mathrm{d}t}^{+}\Big|_{t=0} \frac{1}{2} d^2(\gamma^t(s), \gamma(1)) + F(\gamma(s)). \tag{18.37}$$

Multiplying (18.36) by $(1 - s)$ and (18.37) by s and adding the two expressions we end up with

$$(1 - s)F(\gamma(0)) + sF(\gamma(1)) - F(\gamma(s)) \geq$$
$$(1 - s)\frac{\mathrm{d}}{\mathrm{d}t}^{+}\Big|_{t=0} \frac{1}{2} d^2(\gamma^t(s), \gamma(0)) + s\frac{\mathrm{d}}{\mathrm{d}t}^{+}\Big|_{t=0} \frac{1}{2} d^2(\gamma^t(s), \gamma(1)).$$

The subadditivity of the upper right derivatives gives

$$(1-s)F(\gamma(0))+sF(\gamma(1))-F(\gamma(s)) \geq \frac{\mathrm{d}}{\mathrm{d}t}^{+}\Big|_{t=0} \Big((1 - s)d^2(\gamma(0), \gamma^t(s)) + sd^2(\gamma^t(s), \gamma(1))\Big). \tag{18.38}$$

Observe now that the triangle inequality gives (see also (16.13))

$$(1 - s)d^2(x, w) + sd^2(w, y) \geq s(1 - s)d^2(x, y),$$

with equality when $w = \gamma(s)$ is the s-intermediate point of γ, therefore the right hand side in (18.38) is nonnegative and we obtain the convexity inequality. □

Monotonicity of Action and Energy Implies EVI The action estimate (18.23) has been cleverly modified in [44] into a differential inequality (with different roles of the deformation parameter t and the interpolation parameter s) involving action and energy which turns out to be sufficient for the validity of the EVI property.

In this case, the deformation scheme used to prove contractivity has to be modified as follows: we keep $v \in X$ fixed and, if $\gamma(s)$ is a continuous curve in $[0, 1]$

connecting v to u, where u plays the role of the "test" point in the EVI formulation, we define

$$\gamma^t(s) := S_{st}(\gamma(s))$$

instead of $S_t(\gamma(s))$. With this notation, the following result (stated for simplicity only in the case EVI = EVI$_0$, i.e. $\lambda = 0$) holds:

Theorem 18.15 (A Differential Criterion for EVI**)** *Assume that a continuous contraction semigroup S in a length space (X, d) and a lower semicontinuous function $F : X \to \mathbb{R}$ satisfy the following assumptions for all $\gamma \in AC^2([0,1]; X)$:*

(i) *$s \mapsto F(\gamma^t_s)$ is absolutely continuous in $[0,1]$ for all $t > 0$;*
(ii) *the maps $r \mapsto \frac{d^+}{dr} d^2(S_r(\gamma(s)), \gamma(s))$, $s \in [0,1]$, are equi-integrable in all intervals $[0, T]$;*
(iii) *for all $t > 0$ one has*

$$\frac{1}{2}\frac{d^+}{dt}|(\gamma^t)'_+(s)|^2 + \frac{d}{ds}F(\gamma^t_s) \le 0 \qquad \text{for } \mathscr{L}^1\text{-a.e. } s \in (0,1), \tag{18.39}$$

where d^+/dt denotes, as usual, the upper right derivative, while $|(\gamma^t)'_+(s)|$ denotes the upper, right metric derivative.

Then $S_t(v)$ satisfies EVI *and therefore it is the gradient flow of F.*

Proof By the semigroup property, it is sufficient to check EVI in the pointwise form (11.13) at $t = 0$. We fix $\gamma \in AC^2((0,1); X)$ connecting v to u. For all $t > 0$, by integration of (18.39) with respect to s we obtain

$$\frac{1}{2}\frac{d^+}{dt}\mathcal{A}(\gamma^t) \le F(\gamma^t_0) - F(\gamma^t_1) = F(v) - F(S_t(u)).$$

Notice that we have been able to pull the upper right derivative out of the integral thanks to Fatou's lemma and to the equi-integrability of the family of maps $s \mapsto |(\gamma^t)'(s)|^2$ in $L^1(0,1)$ (in turn, this equi-integrability is guaranteed by the contractivity of the semigroup S, together with (ii)).

Now, an integration with respect to t gives

$$\begin{aligned}\frac{1}{2}d^2(S_t(u), v) - \frac{1}{2}\mathcal{A}(\gamma^0) &\le \mathcal{A}(\gamma^t) - \mathcal{A}(\gamma^0)\\ &\le tF(v) - \int_0^t F(S_r(u))\,dr.\end{aligned}$$

Since (X, d) is a length space and γ is arbitrary, we can replace $\mathcal{A}(\gamma^0) = \mathcal{A}(\gamma)$ by $d^2(u, v)$ in the inequality. Dividing both sides by t, the lower semicontinuity of F and the continuity of S yield (11.13). □