Chapter 8

Absolutely Continuous Curves in $\mathscr{P}_p(X)$ and the Continuity Equation

In this chapter we endow $\mathscr{P}_p(X)$, when X is a separable Hilbert space, with a kind of differential structure, consistent with the metric structure introduced in the previous chapter. Our starting point is the analysis of absolutely continuous curves $\mu_t : (a,b) \to \mathscr{P}_p(X)$ and of their metric derivative $|\mu'|(t)$: recall that these concepts depend only on the metric structure of $\mathscr{P}_p(X)$, by Definition 1.1.1 and (1.1.3). We show in Theorem 8.3.1 that for $p > 1$ this class of curves coincides with (distributional, in the duality with smooth cylindrical test functions) solutions of the continuity equation

$$\frac{\partial}{\partial t}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \qquad \text{in } X \times (a,b).$$

More precisely, given an absolutely continuous curve μ_t, one can find a Borel time-dependent velocity field $v_t : X \to X$ such that $\|v_t\|_{L^p(\mu_t)} \le |\mu'|(t)$ for $\mathscr{L}^1$-a.e. $t \in (a,b)$ and the continuity equation holds. Conversely, if μ_t solve the continuity equation for some Borel velocity field v_t with $\int_a^b \|v_t\|_{L^p(\mu_t)}\,dt < +\infty$, then μ_t is an absolutely continuous curve and $\|v_t\|_{L^p(\mu_t)} \ge |\mu'|(t)$ for $\mathscr{L}^1$-a.e. $t \in (a,b)$.

As a consequence of Theorem 8.3.1 we see that among all velocity fields v_t which produce the same flow μ_t, there is a unique optimal one with smallest $L^p(\mu_t; X)$-norm, equal to the metric derivative of μ_t; we view this optimal field as the "tangent" vector field to the curve μ_t. To make this statement more precise, one can show that the minimality of the L^p norm of v_t is characterized by the property

$$v_t \in \overline{\{j_q(\nabla\varphi) : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^p(\mu_t;X)} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b), \qquad (8.0.1)$$

where q is the conjugate exponent of p and $j_q : L^q(\mu; X) \to L^p(\mu; X)$ is the duality map, i.e. $j_q(v) = |v|^{q-2}v$ (here gradients are thought as covectors, and therefore as elements of L^q).

The characterization (8.0.1) of tangent vectors strongly suggests, in the case $p = 2$, to consider the following tangent to $\mathscr{P}_2(X)$

$$\mathrm{Tan}_\mu \mathscr{P}_2(X) := \overline{\{\nabla\varphi : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^2(\mu;X)} \qquad \forall \mu \in \mathscr{P}_2(X), \tag{8.0.2}$$

endowed with the natural L^2 metric. Moreover, as a consequence of the characterization of absolutely continuous curves in $\mathscr{P}_2(X)$, we recover the Benamou–Brenier (see [21], where the formula was introduced for numerical purposes) formula for the Wasserstein distance:

$$W_2^2(\mu_0, \mu_1) = \min\left\{\int_0^1 \|v_t\|^2_{L^2(\mu_t;X)}\,dt : \ \frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0\right\}. \tag{8.0.3}$$

Indeed, for any admissible curve we use the inequality between L^2 norm of v_t and metric derivative to obtain:

$$\int_0^1 \|v_t\|^2_{L^2(\mu_t;X)}\,dt \geq \int_0^1 |\mu'|^2(t)\,dt \geq W_2^2(\mu_0, \mu_1).$$

Conversely, since we know that $\mathscr{P}_2(X)$ is a length space, we can use a geodesic μ_t and its tangent vector field v_t to obtain equality in (8.0.3). Similar arguments work in the case $p > 1$ as well, with the only drawback that a priori the L^p closure of $j_q(\nabla\varphi)$ is not a vector space in general, so we are able only to define a tangent cone. We also show that optimal transport maps belong to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ under quite general conditions.

In this way we recover in a more general framework the *Riemannian interpretation* of the Wasserstein distance developed by Otto in [107] (see also [106], [83]) and used to study the long time behaviour of the porous medium equation. In the original paper [107], (8.0.3) is derived in the case $X = \mathbb{R}^d$ using formally the concept of Riemannian submersion and the family of maps $\phi \mapsto \phi_\# \mu$ (indexed by $\mu \ll \mathscr{L}^d$) from Arnold's space of diffeomorphisms into the Wasserstein space. In Otto's formalism tangent vectors are rather thought as $\boldsymbol{s} = \frac{d}{dt}\mu_t$ and these vectors are identified, via the continuity equation, with $-D\cdot(v_{\boldsymbol{s}}\mu_t)$. Moreover $v_{\boldsymbol{s}}$ is chosen to be the gradient of a function $\psi_{\boldsymbol{s}}$, so that $D\cdot(\nabla\psi_{\boldsymbol{s}}\mu_t) = -\boldsymbol{s}$. Then the metric tensor is induced by the identification $\boldsymbol{s} \mapsto \nabla\phi_{\boldsymbol{s}}$ as follows:

$$\langle \boldsymbol{s}, \boldsymbol{s}' \rangle_{\mu_t} := \int_{\mathbb{R}^d} \langle \nabla\psi_{\boldsymbol{s}}, \nabla\psi_{\boldsymbol{s}'} \rangle \, d\mu_t.$$

As noticed in [107], both the identification between tangent vectors and gradients and the scalar product depend on μ_t, and these facts lead to a non trivial geometry of the Wasserstein space. We prefer instead to consider directly v_t as the tangent vectors, allowing them to be not necessarily gradients: this leads to (8.0.2).

Another consequence of the characterization of absolutely continuous curves is a result, given in Proposition 8.4.6, concerning the infinitesimal behaviour of the Wasserstein distance along absolutely continuous curves μ_t: given the tangent vector field v_t to the curve, we show that

$$\lim_{h\to 0} \frac{W_p(\mu_{t+h}, (\boldsymbol{i} + hv_t)_\# \mu_t)}{|h|} = 0 \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b).$$

Moreover the optimal transport plans between μ_t and μ_{t+h}, rescaled in a suitable way, converge to the transport plan $(\boldsymbol{i} \times v_t)_\# \mu_t$ associated to v_t (see (8.4.6)). This proposition shows that the infinitesimal behaviour of the Wasserstein distance is governed by transport maps even in the situations when globally optimal transport maps fail to exist (recall that the existence of optimal transport maps requires regularity assumptions on the initial measure μ). As a consequence, we will obtain in Theorem 8.4.7 a formula for the derivative of the map $t \mapsto W_p^p(\mu_t, \nu)$.

8.1 The continuity equation in $\mathbb{R}^d$

In this section we collect some results on the continuity equation

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T), \tag{8.1.1}$$

which we will need in the sequel. Here μ_t is a Borel family of probability measures on $\mathbb{R}^d$ defined for t in the open interval $I := (0,T)$, $v : (x,t) \mapsto v_t(x) \in \mathbb{R}^d$ is a Borel velocity field such that

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|\, d\mu_t(x)\, dt < +\infty, \tag{8.1.2}$$

and we suppose that (8.1.1) holds in the sense of distributions, i.e.

$$\begin{gathered}\int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \varphi(x,t) + \langle v_t(x), \nabla_x \varphi(x,t) \rangle \Big)\, d\mu_t(x)\, dt = 0, \\ \forall\, \varphi \in C_c^\infty(\mathbb{R}^d \times (0,T)).\end{gathered} \tag{8.1.3}$$

Remark 8.1.1 (More general test functions). By a simple regularization argument via convolution, it is easy to show that (8.1.3) holds if $\varphi \in C_c^1\big(\mathbb{R}^d \times (0,T)\big)$ as well. Moreover, under condition (8.1.2), we can also consider bounded test functions φ, with bounded gradient, whose support has a compact projection in $(0,T)$ (that is, the support in x need not be compact): it suffices to approximate φ by $\varphi\chi_R$ where $\chi_R \in C_c^\infty(\mathbb{R}^d)$, $0 \le \chi_R \le 1$, $|\nabla \chi_R| \le 2$ and $\chi_R = 1$ on $B_R(0)$. This more general choice of the test functions is consistent with the infinite-dimensional case, where cylindrical test functions will be considered, see Definition 5.1.11 and (8.3.8).

First of all we recall some (technical) preliminaries.

Lemma 8.1.2 (Continuous representative). *Let μ_t be a Borel family of probability measures satisfying* (8.1.3) *for a Borel vector field v_t satisfying* (8.1.2). *Then there exists a* narrowly continuous *curve $t \in [0,T] \mapsto \tilde\mu_t \in \mathscr{P}(\mathbb{R}^d)$ such that $\mu_t = \tilde\mu_t$ for $\mathscr{L}^1$-a.e. $t \in (0,T)$. Moreover, if $\varphi \in C^1_c(\mathbb{R}^d \times [0,T])$ and $t_1 \le t_2 \in [0,T]$ we have*

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(x,t_2)\, d\tilde\mu_{t_2}(x) - \int_{\mathbb{R}^d} \varphi(x,t_1)\, d\tilde\mu_{t_1}(x) \\ = \int_{t_1}^{t_2} \int_{\mathbb{R}^d} \Big(\partial_t\varphi + \langle \nabla\varphi, v_t\rangle\Big)\, d\mu_t(x)\, dt.\end{aligned} \tag{8.1.4}$$

Proof. Let us take $\varphi(x,t) = \eta(t)\zeta(x)$, $\eta \in C^\infty_c(0,T)$ and $\zeta \in C^\infty_c(\mathbb{R}^d)$; we have

$$-\int_0^T \eta'(t)\Big(\int_{\mathbb{R}^d} \zeta(x)\, d\mu_t(x)\Big)\, dt = \int_0^T \eta(t)\Big(\int_{\mathbb{R}^d} \langle \nabla\zeta(x), v_t(x)\rangle\, d\mu_t(x)\Big)\, dt,$$

so that the map

$$t \mapsto \mu_t(\zeta) = \int_{\mathbb{R}^d} \zeta(x)\, d\mu_t(x)$$

belongs to $W^{1,1}(0,T)$ with distributional derivative

$$\dot\mu_t(\zeta) = \int_{\mathbb{R}^d} \langle \nabla\zeta(x), v_t(x)\rangle\, d\mu_t(x) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,T) \tag{8.1.5}$$

with

$$|\dot\mu_t(\zeta)| \le V(t) \sup_{\mathbb{R}^d} |\nabla\zeta|, \quad V(t) := \int_{\mathbb{R}^d} |v_t(x)|\, d\mu_t(x), \quad V \in L^1(0,T). \tag{8.1.6}$$

If L_ζ is the set of its Lebesgue points, we know that $\mathscr{L}^1((0,T)\setminus L_\zeta) = 0$. Let us now take a countable set Z which is dense in $C^1_c(\mathbb{R}^d)$ with respect the usual C^1 norm $\|\zeta\|_{C^1} = \sup_{\mathbb{R}^d}(|\zeta|, |\nabla\zeta|)$ and let us set $L_Z := \cap_{\zeta\in Z} L_\zeta$. The restriction of the curve μ to L_Z provides a uniformly continuous family of bounded functionals on $C^1_c(\mathbb{R}^d)$, since (8.1.6) shows

$$|\mu_t(\zeta) - \mu_s(\zeta)| \le \|\zeta\|_{C^1} \int_s^t V(\lambda)\, d\lambda \quad \forall\, s,t \in L_Z.$$

Therefore, it can be extended in a unique way to a continuous curve $\{\tilde\mu_t\}_{t\in[0,T]}$ in $[C^1_c(\mathbb{R}^d)]'$. If we show that $\{\mu_t\}_{t\in L_Z}$ is also tight, the extension provides a continuous curve in $\mathscr{P}(\mathbb{R}^d)$.

For, let us consider nonnegative, smooth functions $\zeta_k : \mathbb{R}^d \to [0,1]$, $k \in \mathbb{N}$, such that

$$\zeta_k(x) = 1 \quad \text{if } |x| \le k, \quad \zeta_k(x) = 0 \quad \text{if } |x| \ge k+1, \quad |\nabla\zeta_k(x)| \le 2.$$

It is not restrictive to suppose that $\zeta_k \in Z$. Applying the previous formula (8.1.5), for $t,\ s \in L_Z$ we have

$$|\mu_t(\zeta_k) - \mu_s(\zeta_k)| \le a_k := 2\int_0^T \int_{k<|x|<k+1} |v_\lambda(x)|\,d\mu_\lambda(x)\,d\lambda,$$

with $\sum_{k=1}^{+\infty} a_k < +\infty$. For a fixed $s \in L_Z$ and $\varepsilon > 0$, being μ_s tight, we can find $k \in \mathbb{N}$ such that $\mu_s(\zeta_k) > 1 - \varepsilon/2$ and $a_k < \varepsilon/2$. It follows that

$$\mu_t(\overline{B_{k+1}(0)}) \ge \mu_t(\zeta_k) \ge 1 - \varepsilon \quad \forall\, t \in L_Z.$$

Now we show (8.1.4). Let us choose $\varphi \in C_c^1(\mathbb{R}^d \times [0,T])$ and set $\varphi_\varepsilon(x,t) = \eta_\varepsilon(t)\varphi(x,t)$, where $\eta_\varepsilon \in C_c^\infty(t_1,t_2)$ such that

$$0 \le \eta_\varepsilon(t) \le 1, \quad \lim_{\varepsilon\downarrow 0} \eta_\varepsilon(t) = \chi_{(t_1,t_2)}(t)\ \forall\, t \in [0,T], \quad \lim_{\varepsilon\downarrow 0} \eta_\varepsilon' = \delta_{t_1} - \delta_{t_2}$$

in the duality with continuous functions in $[0,T]$. We get

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t(\eta_\varepsilon\varphi) + \langle \nabla_x(\eta_\varepsilon\varphi), v_t\rangle\Big)\,d\mu_t(x)\,dt \\ &= \int_0^T \eta_\varepsilon(t) \int_{\mathbb{R}^d} \Big(\partial_t\varphi(x,t) + \langle v_t(x), \nabla_x\varphi(x,t)\rangle\Big)\,d\mu_t(x)\,dt \\ &\quad + \int_0^T \eta_\varepsilon'(t) \int_{\mathbb{R}^d} \varphi(x,t)\,d\tilde\mu_t(x)\,dt. \end{aligned}$$

Passing to the limit as ε vanishes and invoking the continuity of $\tilde\mu_t$, we get (8.1.4). □

Lemma 8.1.3 (Time rescaling). *Let* $\mathsf{t} : s \in [0,T'] \to \mathsf{t}(s) \in [0,T]$ *be a strictly increasing absolutely continuous map with absolutely continuous inverse* $\mathsf{s} := \mathsf{t}^{-1}$. *Then* (μ_t, v_t) *is a distributional solution of* (8.1.1) *if and only if*

$$\hat\mu := \mu\circ\mathsf{t},\ \hat v := \mathsf{t}' v\circ\mathsf{t}, \qquad \textit{is a distributional solution of } (8.1.1) \textit{ on } (0,T').$$

Proof. By an elementary smoothing argument we can assume that s is continuously differentiable and $\mathsf{s}' > 0$. We choose $\hat\varphi \in C_c^1(\mathbb{R}^d \times (0,T'))$ and let us set $\varphi(x,t) := \hat\varphi(x,\mathsf{s}(t))$; since $\varphi \in C_c^1(\mathbb{R}^d \times (0,T))$ we have

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} \big(\mathsf{s}'(t)\partial_s\hat\varphi(x,\mathsf{s}(t)) + \langle \nabla\hat\varphi(x,\mathsf{s}(t)), \hat v_t(x)\rangle\big)\,d\mu_t(x)\,dt \\ &= \int_0^T \mathsf{s}'(t) \int_{\mathbb{R}^d} \Big(\partial_s\hat\varphi(x,\mathsf{s}(t)) + \langle \nabla_x\hat\varphi(x,\mathsf{s}(t)), \frac{v_t(x)}{\mathsf{s}'(t)}\rangle\Big)\,d\mu_t(x)\,dt \\ &= \int_0^{T'} \int_{\mathbb{R}^d} \Big(\partial_s\hat\varphi(x,s) + \langle \nabla_x\hat\varphi(x,s), \mathsf{t}'(s)v_{\mathsf{t}(s)}(x)\rangle\Big)\,d\hat\mu_s(x)\,ds. \end{aligned}$$

□

When the velocity field v_t is more regular, the classical method of characteristics provides an explicit solution of (8.1.1).

First we recall an elementary result of the theory of ordinary differential equations.

Lemma 8.1.4 (The characteristic system of ODE). *Let v_t be a Borel vector field such that for every compact set $B \subset \mathbb{R}^d$*

$$\int_0^T \Big(\sup_B |v_t| + \mathrm{Lip}(v_t, B) \Big)\, dt < +\infty. \tag{8.1.7}$$

Then, for every $x \in \mathbb{R}^d$ and $s \in [0,T]$ the ODE

$$X_s(x,s) = x, \quad \frac{d}{dt} X_t(x,s) = v_t(X_t(x,s)), \tag{8.1.8}$$

admits a unique maximal solution defined in an interval $I(x,s)$ relatively open in $[0,T]$ and containing s as (relatively) internal point.
Furthermore, if $t \mapsto |X_t(x,s)|$ is bounded in the interior of $I(x,s)$ then $I(x,s) = [0,T]$; finally, if v satisfies the global bounds analogous to (8.1.7)

$$S := \int_0^T \Big(\sup_{\mathbb{R}^d} |v_t| + \mathrm{Lip}(v_t, \mathbb{R}^d) \Big)\, dt < +\infty, \tag{8.1.9}$$

then the flow map X satisfies

$$\int_0^T \sup_{x \in \mathbb{R}^d} |\partial_t X_t(x,s)|\, dt \le S, \qquad \sup_{t,s \in [0,T]} \mathrm{Lip}(X_t(\cdot,s), \mathbb{R}^d) \le e^S. \tag{8.1.10}$$

For simplicity, we set $X_t(x) := X_t(x,0)$ in the particular case $s = 0$ and we denote by $\tau(x) := \sup I(x,0)$ the length of the maximal time domain of the characteristics leaving from x at $t = 0$.

Remark 8.1.5 (The characteristics method for backward first order linear PDE's). Characteristics provide a useful representation formula for classical solutions of the backward equation (formally adjoint to (8.1.1))

$$\partial_t \varphi + \langle v_t, \nabla \varphi \rangle = \psi \quad \text{in } \mathbb{R}^d \times (0,T), \quad \varphi(x,T) = \varphi_T(x) \quad x \in \mathbb{R}^d, \tag{8.1.11}$$

when, e.g., $\psi \in C^1_b(\mathbb{R}^d \times (0,T)), \varphi_T \in C^1_b(\mathbb{R}^d)$ and v satisfies the global bounds (8.1.9), so that maximal solutions are always defined in $[0,T]$. A direct calculation shows that

$$\varphi(x,t) := \varphi_T(X_T(x,t)) - \int_t^T \psi(X_s(x,t), s)\, ds \tag{8.1.12}$$

solve (8.1.11). For $X_s(X_t(x,0),t) = X_s(x,0)$ yields

$$\varphi(X_t(x,0),t) = \varphi_T(X_T(x,0)) - \int_t^T \psi(X_s(x,0), s)\, ds,$$

and differentiating both sides with respect to t we obtain

$$\left[\frac{\partial\varphi}{\partial t} + \langle v_t, \nabla\varphi\rangle\right](X_t(x,0),t) = \psi(X_t(x,0),t).$$

Since x (and then $X_t(x,0)$) is arbitrary we conclude that (8.1.18) is fulfilled.

Now we use characteristics to prove the existence, the uniqueness, and a representation formula of the solution of the continuity equation, under suitable assumption on v.

Lemma 8.1.6. *Let v_t be a Borel velocity field satisfying* (8.1.7), (8.1.2), *let $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$, and let X_t be the maximal solution of the ODE* (8.1.8) *(corresponding to $s = 0$). Suppose that for some $\bar t \in (0,T]$*

$$\tau(x) > \bar t \quad \text{for } \mu_0\text{-a.e. } x \in \mathbb{R}^d. \tag{8.1.13}$$

Then $t \mapsto \mu_t := (X_t)_\# \mu_0$ is a continuous solution of (8.1.1) *in $[0, \bar t]$.*

Proof. The continuity of μ_t follows easily since $\lim_{s\to t} X_s(x) = X_t(x)$ for μ_0-a.e. $x \in \mathbb{R}^d$: thus for every continuous and bounded function $\zeta : \mathbb{R}^d \to \mathbb{R}$ the dominated convergence theorem yields

$$\lim_{s\to t}\int_{\mathbb{R}^d} \zeta\, d\mu_s = \lim_{s\to t}\int_{\mathbb{R}^d} \zeta(X_s(x))\, d\mu_0(x) = \int_{\mathbb{R}^d} \zeta(X_t(x))\, d\mu_0(x) = \int_{\mathbb{R}^d} \zeta\, d\mu_t.$$

For any $\varphi \in C_c^\infty(\mathbb{R}^d \times (0,\bar t))$ and for μ_0-a.e. $x \in \mathbb{R}^d$ the maps $t \mapsto \phi_t(x) := \varphi(X_t(x),t)$ are absolutely continuous in $(0,\bar t)$, with

$$\dot\phi_t(x) = \partial_t\varphi(X_t(x),t) + \langle\nabla\varphi(X_t(x),t), v_t(X_t(x))\rangle = \Lambda(\cdot,t)\circ X_t,$$

where $\Lambda(x,t) := \partial_t\varphi(x,t) + \langle\nabla\varphi(x,t), v_t(x)\rangle$. We thus have

$$\begin{aligned}\int_0^T\int_{\mathbb{R}^d} |\dot\phi_t(x)|\, d\mu_0(x)\, dt &= \int_0^T\int_{\mathbb{R}^d} |\Lambda(X_t(x),t)|\, d\mu_0(x)\, dt \\ &= \int_0^T\int_{\mathbb{R}^d} |\Lambda(x,t)|\, d\mu_t(x)\, dt \\ &\le \mathrm{Lip}(\varphi)\Big(T + \int_0^T\int_{\mathbb{R}^d} |v_t(x)|\, d\mu_t(x)\, dt\Big) < +\infty\end{aligned}$$

and therefore

$$\begin{aligned}0 &= \int_{\mathbb{R}^d} \varphi(x,\bar t)\, d\mu_{\bar t}(x) - \int_{\mathbb{R}^d}\varphi(x,0)\, d\mu_0(x) = \int_{\mathbb{R}^d}\Big(\varphi(X_{\bar t}(x),\bar t) - \varphi(x,0)\Big)\, d\mu_0(x) \\ &= \int_{\mathbb{R}^d}\Big(\int_0^{\bar t}\dot\phi_t(x)\, dt\Big)\, d\mu_0(x) = \int_0^{\bar t}\int_{\mathbb{R}^d}\big(\partial_t\varphi + \langle\nabla\varphi, v_t\rangle\big)\, d\mu_t\, dt,\end{aligned}$$

by a simple application of Fubini's theorem. □

We want to prove that, under reasonable assumptions, in fact *any* solution of (8.1.1) can be represented as in Lemma 8.1.6. The first step is a uniqueness theorem for the continuity equation under minimal regularity assumptions on the velocity field. Notice that the only global information on v_t is (8.1.14). The proof, based on a classical duality argument (see for instance [57, 9]), could be much simplified by the assumption that the velocity field is globally bounded, but we prefer to keep here a version of the lemma stronger than the one actually needed in the proof of Theorem 8.3.1.

Proposition 8.1.7 (Uniqueness and comparison for the continuity equation). *Let σ_t be a narrowly continuous family of signed measures solving $\partial_t\sigma_t + \nabla\cdot(v_t\sigma_t) = 0$ in $\mathbb{R}^d \times (0,T)$, with $\sigma_0 \le 0$,*

$$\int_0^T \int_{\mathbb{R}^d} |v_t|\, d|\sigma_t| dt < +\infty, \tag{8.1.14}$$

and

$$\int_0^T \left(|\sigma_t|(B) + \sup_B |v_t| + \mathrm{Lip}(v_t, B) \right) dt < +\infty$$

for any bounded closed set $B \subset \mathbb{R}^d$. Then $\sigma_t \le 0$ for any $t \in [0,T]$.

Proof. Fix $\psi \in C_c^\infty(\mathbb{R}^d \times (0,T))$ with $0 \le \psi \le 1$, $R > 0$, and a smooth cut-off function

$$\begin{gathered}\chi_R(\cdot) = \chi(\cdot/R) \in C_c^\infty(\mathbb{R}^d) \quad \text{such that } 0 \le \chi_R \le 1,\ |\nabla\chi_R| \le 2/R,\\ \chi_R \equiv 1 \text{ on } B_R(0), \text{ and } \chi_R \equiv 0 \text{ on } \mathbb{R}^d \setminus B_{2R}(0).\end{gathered} \tag{8.1.15}$$

We define w_t so that $w_t = v_t$ on $B_{2R}(0) \times (0,T)$, $w_t = 0$ if $t \notin [0,T]$ and

$$\sup_{\mathbb{R}^d} |w_t| + \mathrm{Lip}(w_t, \mathbb{R}^d) \le \sup_{B_{2R}(0)} |v_t| + \mathrm{Lip}(v_t, B_{2R}(0)) \quad \forall\, t \in [0,T]. \tag{8.1.16}$$

Let w_t^ε be obtained from w_t by a double mollification with respect to the space and time variables: notice that w_t^ε satisfy

$$\sup_{\varepsilon\in(0,1)} \int_0^T \left(\sup_{\mathbb{R}^d} |w_t^\varepsilon| + \mathrm{Lip}(w_t^\varepsilon, \mathbb{R}^d) \right) dt < +\infty. \tag{8.1.17}$$

We now build, by the method of characteristics described in Remark 8.1.5, a smooth solution $\varphi^\varepsilon : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ of the PDE

$$\frac{\partial\varphi^\varepsilon}{\partial t} + \langle w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle = \psi \quad \text{in } \mathbb{R}^d \times (0,T), \quad \varphi^\varepsilon(x,T) = 0 \quad x \in \mathbb{R}^d. \tag{8.1.18}$$

Combining the representation formula (8.1.12), the uniform bound (8.1.17), and the estimate (8.1.10), it is easy to check that $0 \ge \varphi^\varepsilon \ge -T$ and $|\nabla\varphi^\varepsilon|$ is uniformly bounded with respect to ε, t and x.

We insert now the test function $\varphi^\varepsilon\chi_R$ in the continuity equation and take into account that $\sigma_0 \le 0$ and $\varphi^\varepsilon \le 0$ to obtain

$$\begin{aligned} 0 \quad &\ge \quad -\int_{\mathbb{R}^d} \varphi^\varepsilon \chi_R \, d\sigma_0 = \int_0^T \int_{\mathbb{R}^d} \chi_R \frac{\partial \varphi^\varepsilon}{\partial t} + \langle v_t, \chi_R \nabla\varphi^\varepsilon + \varphi^\varepsilon \nabla\chi_R\rangle \, d\sigma_t dt \\ &= \int_0^T \int_{\mathbb{R}^d} \chi_R(\psi + \langle v_t - w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle) \, d\sigma_t dt + \int_0^T \int_{\mathbb{R}^d} \varphi^\varepsilon \langle \nabla\chi_R, v_t\rangle \, d\sigma_t dt \\ &\ge \int_0^T \int_{\mathbb{R}^d} \chi_R(\psi + \langle v_t - w_t^\varepsilon, \nabla\varphi^\varepsilon\rangle) \, d\sigma_t dt - \int_0^T \int_{\mathbb{R}^d} |\nabla\chi_R||v_t| \, d|\sigma_t| \, dt. \end{aligned}$$

Letting $\varepsilon \downarrow 0$ and using the uniform bound on $|\nabla\varphi^\varepsilon|$ and the fact that $w_t = v_t$ on $\operatorname{supp}\chi_R \times [0,T]$, we get

$$\int_0^T \int_{\mathbb{R}^d} \chi_R \psi \, d\sigma_t \, dt \le \int_0^T \int_{\mathbb{R}^d} |\nabla\chi_R||v_t| \, d|\sigma_t| \, dt \le \frac{2}{R} \int_0^T \int_{R \le |x| \le 2R} |v_t| \, d|\sigma_t| \, dt.$$

Eventually letting $R \to \infty$ we obtain that $\int_0^T \int_{\mathbb{R}^d} \psi \, d\sigma_t dt \le 0$. Since ψ is arbitrary the proof is achieved. □

Proposition 8.1.8 (Representation formula for the continuity equation). *Let μ_t, $t \in [0,T]$, be a narrowly continuous family of Borel probability measures solving the continuity equation* (8.1.1) *w.r.t. a Borel vector field v_t satisfying* (8.1.7) *and* (8.1.2). *Then for μ_0-a.e. $x \in \mathbb{R}^d$ the characteristic system* (8.1.8) *admits a globally defined solution $X_t(x)$ in $[0,T]$ and*

$$\mu_t = (X_t)_\# \mu_0 \quad \forall\, t \in [0,T]. \tag{8.1.19}$$

Moreover, if

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \, dt < +\infty \quad \textit{for some } p > 1, \tag{8.1.20}$$

then the velocity field v_t is the time derivative of X_t in the L^p-sense

$$\lim_{h\downarrow 0} \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} - v_t(X_t(x)) \right|^p d\mu_0(x) \, dt = 0, \tag{8.1.21}$$

$$\lim_{h\to 0} \frac{X_{t+h}(x,t) - x}{h} = v_t(x) \quad \textit{in } L^p(\mu_t;\mathbb{R}^d) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t \in (0,T). \tag{8.1.22}$$

Proof. Let $E_s = \{\tau > s\}$ and let us use the fact that, proved in Lemma 8.1.6, that $t \mapsto X_{t\#}(\chi_{E_s}\mu_0)$ is the solution of (8.1.1) in $[0,s]$. By Proposition 8.1.7 we get also

$$X_{t\#}(\chi_{E_s}\mu_0) \le \mu_t \qquad \text{whenever } 0 \le t \le s.$$

Using the previous inequality with $s = t$ we can estimate:

$$\begin{aligned}
\int_{\mathbb{R}^d} \sup_{(0,\tau(x))} |X_t(x) - x| \, d\mu_0(x) &\leq \int_{\mathbb{R}^d} \int_0^{\tau(x)} |\dot{X}_t(x)| \, d\mu_0(x) \\
&= \int_{\mathbb{R}^d} \int_0^{\tau(x)} |v_t(X_t(x))| \, d\mu_0(x) \\
&= \int_0^T \int_{E_t} |v_t(X_t(x))| \, d\mu_0(x) \, dt \\
&\leq \int_0^T \int_{\mathbb{R}^d} |v_t| \, d\mu_t \, dt.
\end{aligned}$$

It follows that $X_t(x)$ is bounded on $(0, \tau(x))$ for μ_0-a.e. $x \in \mathbb{R}^d$ and therefore X_t is globally defined in $[0, T]$ for μ_0-a.e. in $\mathbb{R}^d$. Applying Lemma 8.1.6 and Proposition 8.1.7 we obtain (8.1.19).

Now we observe that the differential quotient $D_h(x, t) := h^{-1}(X_{t+h}(x) - X_t(x))$ can be bounded in $L^p(\mu_0 \times \mathscr{L}^1)$ by

$$\begin{aligned}
&\int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} \right|^p d\mu_0(x) \, dt \\
&= \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{1}{h} \int_0^h v_{t+s}(X_{t+s}(x)) \, ds \right|^p d\mu_0(x) \, dt \\
&\leq \int_0^{T-h} \int_{\mathbb{R}^d} \frac{1}{h} \int_0^h |v_{t+s}(X_{t+s}(x))|^p \, ds \, d\mu_0(x) \, dt \\
&\leq \int_0^T \int_{\mathbb{R}^d} |v_t(X_t(x))|^p \, d\mu_0(x) \, dt < +\infty.
\end{aligned}$$

Since we already know that D_h is pointwise converging to $v_t \circ X_t$ $\mu_0 \times \mathscr{L}^1$-a.e. in $\mathbb{R}^d \times (0, T)$, we obtain the strong convergence in $L^p(\mu_0 \times \mathscr{L}^1)$, i.e. (8.1.21).

Finally, we can consider $t \mapsto X_t(\cdot)$ and $t \mapsto v_t(X_t(\cdot)$ as maps from $(0, T)$ to $L^p(\mu_0; \mathbb{R}^d)$; (8.1.21) is then equivalent to

$$\lim_{h \downarrow 0} \int_0^{T-h} \left\| \frac{X_{t+h} - X_t}{h} - v_t(X_t) \right\|^p_{L^p(\mu_0; \mathbb{R}^d)} dt = 0,$$

and it shows that $t \mapsto X_t(\cdot)$ belongs to $AC^p(0, T; L^p(\mu_0; \mathbb{R}^d))$. General results for absolutely continuous maps in reflexive Banach spaces (see 1.1.3) yield that X_t is differentiable $\mathscr{L}^1$-a.e. in $(0, T)$, so that

$$\lim_{h \to 0} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} - v_t(X_t(x)) \right|^p d\mu_0(x) = 0 \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0, T).$$

Since $X_{t+h}(x) = X_h(X_t(x), t)$, we obtain (8.1.22). □

Now we state an approximation result for general solution of (8.1.1) with more regular ones, satisfying the conditions of the previous Proposition 8.1.8.

Lemma 8.1.9 (Approximation by regular curves). *Let $p \geq 1$ and let μ_t be a time-continuous solution of* (8.1.1) *w.r.t. a velocity field satisfying the p-integrability condition*

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \, dt < +\infty. \tag{8.1.23}$$

Let $(\rho_\varepsilon) \subset C^\infty(\mathbb{R}^d)$ be a family of strictly positive *mollifiers in the x variable, (e.g. $\rho_\varepsilon(x) = (2\pi\varepsilon)^{-d/2} \exp(-|x|^2/2\varepsilon)$), and set*

$$\mu_t^\varepsilon := \mu_t * \rho_\varepsilon, \quad E_t^\varepsilon := (v_t \mu_t) * \rho_\varepsilon, \quad v_t^\varepsilon := \frac{E_t^\varepsilon}{\mu_t^\varepsilon}. \tag{8.1.24}$$

Then μ_t^ε is a continuous solution of (8.1.1) *w.r.t. v_t^ε, which satisfy the local regularity assumptions* (8.1.7) *and the uniform integrability bounds*

$$\int_{\mathbb{R}^d} |v_t^\varepsilon(x)|^p \, d\mu_t^\varepsilon(x) \leq \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \quad \forall \, t \in (0,T). \tag{8.1.25}$$

Moreover, $E_t^\varepsilon \to v_t \mu_t$ narrowly and

$$\lim_{\varepsilon \downarrow 0} \|v_t^\varepsilon\|_{L^p(\mu_t^\varepsilon;\mathbb{R}^d)} = \|v_t\|_{L^p(\mu_t;\mathbb{R}^d)} \qquad \forall t \in (0,T). \tag{8.1.26}$$

Proof. With a slight abuse of notation, we are denoting the measure μ_t^ε and its density w.r.t. $\mathscr{L}^d$ by the same symbol. Notice first that $|E^\varepsilon|(t,\cdot)$ and its spatial gradient are uniformly bounded in space by the product of $\|v_t\|_{L^1(\mu_t)}$ with a constant depending on ε, and the first quantity is integrable in time. Analogously, $|\mu_t^\varepsilon|(t,\cdot)$ and its spatial gradient are uniformly bounded in space by a constant depending on ε. Therefore, as $v_t^\varepsilon = E_t^\varepsilon / \mu_t^\varepsilon$, the local regularity assumptions (8.1.7) is fulfilled if

$$\inf_{|x| \leq R, \, t \in [0,T]} \mu_t^\varepsilon(x) > 0 \qquad \text{for any } \varepsilon > 0, \ R > 0.$$

This property is immediate, since μ_t^ε are continuous w.r.t. t and equi-continuous w.r.t. x, and therefore continuous in both variables.

Lemma 8.1.10 shows that (8.1.25) holds. Notice also that μ_t^ε solve the continuity equation

$$\partial_t \mu_t^\varepsilon + \nabla \cdot (v_t^\varepsilon \mu_t^\varepsilon) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T), \tag{8.1.27}$$

because, by construction, $\nabla \cdot (v_t^\varepsilon \, \mu_t^\varepsilon) = \nabla \cdot ((v_t \mu_t) * \rho_\varepsilon) = (\nabla \cdot (v_t \mu_t)) * \rho_\varepsilon$. Finally, general lower semicontinuity results on integral functionals defined on measures of the form

$$(E, \mu) \mapsto \int_{\mathbb{R}^d} \left| \frac{E}{\mu} \right|^p \, d\mu$$

(see for instance Theorem 2.34 and Example 2.36 in [11]) provide (8.1.26). □

Lemma 8.1.10. *Let $p \geq 1$, $\mu \in \mathscr{P}(\mathbb{R}^d)$ and let E be a $\mathbb{R}^m$-valued measure in $\mathbb{R}^d$ with finite total variation and absolutely continuous with respect to μ. Then*

$$\int_{\mathbb{R}^d} \left|\frac{E*\rho}{\mu*\rho}\right|^p \mu*\rho\,dx \leq \int_{\mathbb{R}^d} \left|\frac{E}{\mu}\right|^p d\mu$$

for any convolution kernel ρ.

Proof. We use Jensen inequality in the following form: if $\Phi : \mathbb{R}^{m+1} \to [0,+\infty]$ is convex, l.s.c. and positively 1-homogeneous, then

$$\Phi\left(\int_{\mathbb{R}^d} \psi(x)\,d\theta(x)\right) \leq \int_{\mathbb{R}^d} \Phi(\psi(x))\,d\theta(x)$$

for any Borel map $\psi : \mathbb{R}^d \to \mathbb{R}^{m+1}$ and any positive and finite measure θ in $\mathbb{R}^d$ (by rescaling θ to be a probability measure and looking at the image measure $\psi_\#\theta$ the formula reduces to the standard Jensen inequality). Fix $x \in \mathbb{R}^d$ and apply the inequality above with $\psi := (E/\mu, 1)$, $\theta := \rho(x-\cdot)\mu$ and

$$\Phi(z,t) := \begin{cases} \dfrac{|z|^p}{t^{p-1}} & \text{if } t > 0 \\ 0 & \text{if } (z,t) = (0,0) \\ +\infty & \text{if either } t < 0 \text{ or } t = 0, z \neq 0, \end{cases}$$

to obtain

$$\begin{aligned} \left|\frac{E*\rho(x)}{\mu*\rho(x)}\right|^p \mu*\rho(x) &= \Phi\left(\int_{\mathbb{R}^d} \frac{E}{\mu}(y)\rho(x-y)\,d\mu(y), \int \rho(x-y)d\mu(y)\right) \\ &\leq \int_{\mathbb{R}^d} \Phi(\frac{E}{\mu}(y),1)\rho(x-y)\,d\mu(y) \\ &= \int_{\mathbb{R}^d} \left|\frac{E}{\mu}\right|^p (y)\rho(x-y)\,d\mu(y). \end{aligned}$$

An integration with respect to x leads to the desired inequality. □

8.2 A probabilistic representation of solutions of the continuity equation

In this section we extend Proposition 8.1.8 to the case when the vector field fails to satisfy (8.1.7) and is in particular not Lipschitz w.r.t. x. Of course in this situation we have to take into account that characteristics are not unique, and we do that by considering suitable probability measures in the space Γ_T of continuous maps from $[0,T]$ into $\mathbb{R}^d$, endowed with the sup norm. The results presented here are not used in the rest of the book, but we believe that they can have an independent interest. Indeed, this kind of notion plays an important role in the uniqueness and stability of Lagrangian flows in [10] and provides an alternative way to the approach of [57].

Our basic representation formula for solutions $\mu_t^{\boldsymbol{\eta}}$ of the continuity equation (8.1.1) is given by

$$\int_{\mathbb{R}^d} \varphi \, d\mu_t^{\boldsymbol{\eta}} := \int_{\mathbb{R}^d \times \Gamma_T} \varphi(\gamma(t)) \, d\boldsymbol{\eta}(x, \gamma) \qquad \forall \varphi \in C_b^0(\mathbb{R}^d), \ \ t \in [0, T] \tag{8.2.1}$$

where $\boldsymbol{\eta}$ is a probability measure in $\mathbb{R}^d \times \Gamma_T$. In the case when $\boldsymbol{\eta}$ is the push forward under $x \mapsto (x, X_\cdot(x))$ of μ_0 (here we are considering X as a function mapping $x \in \mathbb{R}^d$ into the solution curve $t \mapsto X_t(x)$ in Γ_T) we see that the measures $\mu_t^{\boldsymbol{\eta}}$ implicitly defined by (8.2.1) simply reduce to the standard ones considered in Proposition 8.1.8, i.e. $\mu_t^{\boldsymbol{\eta}} = X_t(\cdot)_\# \mu_0$.

By introducing the evaluation maps

$$\mathsf{e}_t : (x, \gamma) \in \mathbb{R}^d \times \Gamma_T \mapsto \gamma(t) \in \mathbb{R}^d, \quad \text{for } t \in [0, T], \tag{8.2.2}$$

(8.2.1) can also be written as

$$\mu_t^{\boldsymbol{\eta}} = (\mathsf{e}_t)_\# \boldsymbol{\eta}. \tag{8.2.3}$$

Theorem 8.2.1 (Probabilistic representation). *Let $\mu_t : [0, T] \to \mathscr{P}(\mathbb{R}^d)$ be a narrowly continuous solution of the continuity equation (8.1.1) for a suitable Borel vector field $v(t, x) = v_t(x)$ satisfying (8.1.20) for some $p > 1$. Then there exists a probability measure $\boldsymbol{\eta}$ in $\mathbb{R}^d \times \Gamma_T$ such that*

(i) *$\boldsymbol{\eta}$ is concentrated on the set of pairs (x, γ) such that $\gamma \in AC^p(0, T; \mathbb{R}^d)$ is a solution of the ODE $\dot\gamma(t) = v_t(\gamma(t))$ for $\mathscr{L}^1$-a.e. $t \in (0, T)$, with $\gamma(0) = x$;*

(ii) *$\mu_t = \mu_t^{\boldsymbol{\eta}}$ for any $t \in [0, T]$, with $\mu_t^{\boldsymbol{\eta}}$ defined as in (8.2.1).*

Conversely, any $\boldsymbol{\eta}$ satisfying (i) and

$$\int_0^T \int_{\mathbb{R}^d \times \Gamma_T} |v_t(\gamma(t))| \, d\boldsymbol{\eta}(x, \gamma) \, dt < +\infty, \tag{8.2.4}$$

induces via (8.2.1) a solution of the continuity equation, with $\mu_0 = \gamma(0)_\# \boldsymbol{\eta}$.

Proof. We first prove the converse implication, since its proof is much simpler. Indeed, notice that due to assumption (i) the set F of all (t, x, γ) such that either $\dot\gamma(t)$ does not exist or it is different from $v_t(\gamma(t))$ is $\mathscr{L}^1 \times \boldsymbol{\eta}$-negligible. As a consequence, we have

$$\dot\gamma(t) = v_t(\gamma(t)) \quad \boldsymbol{\eta}\text{-a.e., for } \mathscr{L}^1\text{-a.e. } t \in (0, T).$$

It is immediate to check using (8.2.1) that $t \mapsto \mu_t^{\boldsymbol{\eta}}$ is narrowly continuous. Now we check that $t \mapsto \int \zeta \, d\mu_t^{\boldsymbol{\eta}}$ is absolutely continuous for $\zeta \in C^1(\mathbb{R}^d)$ bounded and with a bounded gradient. Indeed, for $s < t$ in I we have

$$\begin{aligned}\left|\int_{\mathbb{R}^d}\zeta\,d\mu_s^{\boldsymbol{\eta}} - \int_{\mathbb{R}^d}\zeta\,d\mu_t^{\boldsymbol{\eta}}\right| &\le \int_s^t\int_{\mathbb{R}^d\times\Gamma_T}|\langle\nabla\zeta(\gamma(\tau)),\dot\gamma(\tau)\rangle|\,d\boldsymbol{\eta}\,d\tau\\ &\le \|\nabla\zeta\|_\infty\int_s^t\int_{\mathbb{R}^d\times\Gamma_T}|v_\tau(\gamma(\tau))|\,d\boldsymbol{\eta}\,d\tau.\end{aligned}$$

By (8.2.4) this inequality immediately gives the absolute continuity of the map. We have also

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^d}\zeta\,d\mu_t^{\boldsymbol{\eta}} &= \frac{d}{dt}\int_{\mathbb{R}^d\times\Gamma_T}\zeta(\gamma(t))\,d\boldsymbol{\eta}\\ &= \int_{\mathbb{R}^d\times\Gamma_T}\langle\nabla\zeta(\gamma),\dot\gamma(t))\rangle\,d\boldsymbol{\eta} = \int_{\mathbb{R}^d}\langle\nabla\zeta,v_t\rangle\,d\mu_t^{\boldsymbol{\eta}}\end{aligned}$$

for $\mathscr{L}^1$-a.e. $t\in(0,T)$. Since this pointwise derivative is also a distributional one, this proves that (8.1.4) holds for test function φ of the form $\zeta(x)\psi(t)$ and therefore for all test functions.

Conversely, let μ_t, v_t be given as in the statement of the theorem and let us apply the regularization Lemma 8.1.9, finding approximations μ_t^ε, v_t^ε satisfying the continuity equation, the uniform integrability condition (8.1.2) and the local regularity assumptions (8.1.7). Therefore, we can apply Proposition 8.1.8, obtaining the representation formula $\mu_t^\varepsilon = (X_t^\varepsilon)_\#\mu_0^\varepsilon$, where X_t^ε is the maximal solution of the ODE $\dot X_t^\varepsilon = v_t^\varepsilon(X_t^\varepsilon)$ with the initial condition $X_0^\varepsilon = x$ (see Lemma 8.1.4). Thinking X^ε as a map from $\mathbb{R}^d$ to Γ_T, we thus define

$$\boldsymbol{\eta}^\varepsilon := (\boldsymbol{i}\times X_\cdot^\varepsilon)_\#\mu_0^\varepsilon \in \mathscr{P}(\mathbb{R}^d\times\Gamma_T).$$

Now we claim that the family $\boldsymbol{\eta}^\varepsilon$ is tight as $\varepsilon\downarrow 0$ and that any limit point $\boldsymbol{\eta}$ fulfills (i) and (ii). The tightness of the family can be obtained from Lemma 5.2.2, by choosing the maps $\boldsymbol{r}^1,\boldsymbol{r}^2$ defined in $\mathbb{R}^d\times\Gamma_T$

$$\boldsymbol{r}^1:(x,\gamma)\mapsto x\in\mathbb{R}^d,\quad \boldsymbol{r}^2:(x,\gamma)\mapsto\gamma-x\in\Gamma_T, \tag{8.2.5}$$

and noticing that $\boldsymbol{r}:\boldsymbol{r}^1\times\boldsymbol{r}^2:\mathbb{R}^d\times\Gamma_T\to\mathbb{R}^d\times\Gamma_T$ is proper, the family $\boldsymbol{r}^1_\#\boldsymbol{\eta}^\varepsilon$ is given by the first marginals μ_0^ε which are tight (indeed, they narrowly converge to μ^0), while $\boldsymbol{\beta}^\varepsilon := \boldsymbol{r}^2_\#\boldsymbol{\eta}^\varepsilon$ satisfy

$$\begin{aligned}\int_{\Gamma_T}\int_0^T|\dot\gamma|^p\,dt\,d\boldsymbol{\beta}^\varepsilon &= \int_{\mathbb{R}^d}\int_0^T|\dot X_t^\varepsilon(x)|^p\,dt\,d\mu_0^\varepsilon(x)\\ &= \int_{\mathbb{R}^d}\int_0^T|v_t^\varepsilon(X_t^\varepsilon)|^p\,dt\,d\mu_0^\varepsilon(x) = \int_0^T\int_{\mathbb{R}^d}|v_t^\varepsilon(x)|^p\,d\mu_t^\varepsilon(x)\,dt\\ &\le \int_0^T\int_{\mathbb{R}^d}|v_t(x)|^p\,d\mu_t(x)\,dt.\end{aligned}$$

Since for $p>1$ the functional $\gamma\mapsto\int_0^T|\dot\gamma|^p\,dt$ (set to $+\infty$ if $\gamma\notin AC^p((0,T);\mathbb{R}^d)$ or $\gamma(0)\neq 0$) has compact sublevel sets in Γ_T, also $\boldsymbol{\beta}^\varepsilon$ is tight, due to Remark 5.1.5.

Let now $\boldsymbol{\eta}$ be a narrow limit point of $\boldsymbol{\eta}^\varepsilon$, along some infinitesimal sequence ε_i. Since

$$\int_{\mathbb{R}^d}\varphi\,d\mu_t^{\boldsymbol{\eta}^{\varepsilon_i}}=\int_{\mathbb{R}^d\times\Gamma_T}\varphi(\gamma(t))\,d\boldsymbol{\eta}^{\varepsilon_i}=\int_{\mathbb{R}^d}\varphi(X_t^{\varepsilon_i})\,d\mu_0^{\varepsilon_i}=\int_{\mathbb{R}^d}\varphi\,d\mu_t^{\varepsilon_i}$$

for any $\varphi\in C_b^0(\mathbb{R}^d)$, we can pass to the limit as $i\to\infty$ to obtain that $\mu_t^{\boldsymbol{\eta}}=\mu_t$, so that condition (ii) holds.

Finally we check condition (i). Let $w(t,x)=w_t(x)$ be a bounded uniformly continuous function, and let us prove the estimate

$$\int_{\mathbb{R}^d\times\Gamma_T}\left|\gamma(t)-x-\int_0^t w_\tau(\gamma(\tau))\,d\tau\right|^p\,d\boldsymbol{\eta}(x,\gamma)\le(2T)^{p-1}\int_0^T\int_{\mathbb{R}^d}|v_\tau-w_\tau|^p\,d\mu_\tau\,d\tau. \tag{8.2.6}$$

Indeed, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\Gamma_T}\left|\gamma(t)-x-\int_0^t w_\tau(\gamma(\tau))\,d\tau\right|^p\,d\boldsymbol{\eta}^\varepsilon(x,\gamma)\\
=\;&\int_{\mathbb{R}^d}\left|X_t^\varepsilon(x)-x-\int_0^t w_\tau(X_\tau^\varepsilon(x))\,d\tau\right|^p\,d\mu^0(x)\\
=\;&\int_{\mathbb{R}^d}\left|\int_0^t(v_\tau^\varepsilon-w_\tau)(X_\tau^\varepsilon(x))\,d\tau\right|^p\,d\mu^0(x)\le t^{p-1}\int_0^t\int_{\mathbb{R}^d}|v_\tau^\varepsilon-w_\tau|^p\,d\mu_t^\varepsilon\,d\tau\\
\le\;&(2t)^{p-1}\int_0^t\int_{\mathbb{R}^d}|v_\tau^\varepsilon-w_\tau^\varepsilon|^p\,d\mu_t^\varepsilon\,d\tau+(2t)^{p-1}\int_0^t\int_{\mathbb{R}^d}|w_\tau^\varepsilon-w_\tau|^p\,d\mu_t^\varepsilon\,d\tau\\
\le\;&(2T)^{p-1}\int_0^T\int_{\mathbb{R}^d}|v_\tau-w_\tau|^p\,d\mu_\tau\,d\tau+(2T)^{p-1}\int_0^T\sup_{x\in\mathbb{R}^d}|w_\tau^\varepsilon(x)-w_\tau(x)|^p\,d\tau,
\end{aligned}$$

where in the last two inequalities we have added and subtracted $w_\tau^\varepsilon:=w_\tau*\rho_\varepsilon$ and then used Lemma 8.1.10. Setting $\varepsilon=\varepsilon_i$ and passing to the limit as $i\to\infty$ we recover (8.2.6), since the function under the integral is a continuous and nonnegative test function in $\mathbb{R}^d\times\Gamma_T$.

Now let $\mu:=\int_0^T\mu_t\,d\mathscr{L}^1(t)$ the Borel measure on $\mathbb{R}^d\times(0,T)$ whose disintegration with respect to $\mathscr{L}^1$ is $\{\mu_t\}_{t\in(0,T)}$ and let $w^n\in C_c^0(\mathbb{R}^d\times(0,T);\mathbb{R}^d)$ be continuous functions with compact support converging to v in $L^p(\mu;\mathbb{R}^d)$. Using the fact that $\mu_t=\mu_t^{\boldsymbol{\eta}}$ we have

$$\int_{\mathbb{R}^d\times\Gamma_T}\int_0^T|w_\tau^n(\gamma(\tau))-v_\tau(\gamma(\tau))|^p\,d\tau\,d\boldsymbol{\eta}=\int_0^T\int_{\mathbb{R}^d}|w_\tau^n-v_\tau|^p\,d\mu_\tau\,d\tau\to 0,$$

as $n\to\infty$ so that, using the triangular inequality in $L^p(\boldsymbol{\eta})$, we can pass to the limit as $n\to\infty$ in (8.2.6) with $w=w^n$ to obtain

$$\int_{\mathbb{R}^d\times\Gamma_T}\left|\gamma(t)-x-\int_0^t v_\tau(\gamma(\tau))\,d\tau\right|^p\,d\boldsymbol{\eta}(x,\gamma)=0\quad\forall\,t\in[0,T], \tag{8.2.7}$$

and therefore

$$\gamma(t) - x - \int_0^t v_\tau(\gamma(\tau))\,d\tau = 0 \qquad \text{for } \boldsymbol{\eta}\text{-a.e. } (x,\gamma)$$

for any $t \in [0,T]$. Choosing all t's in $(0,T)\cap\mathbb{Q}$ we obtain an exceptional $\boldsymbol{\eta}$-negligible set that does not depend on t and use the continuity of γ to show that the identity is fulfilled for any $t \in [0,T]$. □

Notice that due to condition (i) the measure $\boldsymbol{\eta}$ in the previous theorem can also be identified with a measure $\boldsymbol{\sigma}$ in Γ_T whose projection on $\mathbb{R}^d$ via the map $\mathsf{e}_0 : \gamma \mapsto \gamma(0)$ is μ_0 and whose corresponding disintegration $\boldsymbol{\sigma} = \int_{\mathbb{R}^d} \boldsymbol{\sigma}_x\,d\mu_0(x)$ is made by probability measures $\boldsymbol{\sigma}_x$ concentrated on solutions of the ODE starting from x at $t = 0$. In this case (8.2.1) takes the simpler equivalent form

$$\int_{\mathbb{R}^d} \varphi\,d\mu_t^{\boldsymbol{\sigma}} := \int_{\Gamma_T} \varphi(\gamma(t))\,d\boldsymbol{\sigma}(\gamma) \qquad \forall \varphi \in C_b^0(\mathbb{R}^d),\ \ t \in [0,T]. \tag{8.2.8}$$

Finally we notice that the results of this section could be easily be extended to the case when $\mathbb{R}^d$ is replaced by a separable Hilbert space, using a finite dimensional projection argument (see in particular the last part of the proof of Theorem 8.3.1).

8.3 Absolutely continuous curves in $\mathscr{P}_p(X)$

In this section we show that the continuity equation characterizes the class of absolutely continuous curves in $\mathscr{P}_p(X)$, with $p > 1$ and X separable Hilbert space (see [9] for a discussion of the degenerate case $p = 1$ when $X = \mathbb{R}^d$).

Let us first recall that the map $j_p : L^p(\mu;X) \to L^q(\mu;X)$ defined by (here $q = p'$ is the conjugate exponent of p)

$$v \mapsto j_p(v) := \begin{cases} |v|^{p-2}v & \text{if } v \neq 0, \\ 0 & \text{if } v = 0, \end{cases} \tag{8.3.1}$$

provides the differential of the convex functional

$$v \in L^p(\mu;X) \mapsto \frac{1}{p}\int_X |v(x)|^p\,d\mu(x), \tag{8.3.2}$$

for every measure $\mu \in \mathscr{P}(X)$; in particular it satisfies

$$\|j_p(v)\|^q_{L^q(\mu,X)} = \|v\|^p_{L^p(\mu,X)} = \int_X \langle j_p(v), v\rangle\,d\mu(x), \tag{8.3.3}$$

$$w = j_p(v) \quad \Longleftrightarrow \quad v = j_q(w), \tag{8.3.4}$$

$$\frac{1}{p}\|v\|^p_{L^p(\mu;X)} - \frac{1}{p}\|w\|^p_{L^p(\mu;X)} \geq \langle j_p(w), v-w\rangle \quad \forall\, v,w \in L^p(\mu;X). \tag{8.3.5}$$

Recall that the space of smooth cylindrical functions $\mathrm{Cyl}(X)$ has been introduced in Definition 5.1.11; the space $\mathrm{Cyl}(X\times I)$, $I=(a,b)$ being an open interval, is defined analogously considering functions $\psi\in C^\infty_c(\mathbb{R}^d\times I)$ and functions $\varphi(x,t)=\psi(\pi(x),t)$.

Theorem 8.3.1 (Absolutely continuous curves and the continuity equation). *Let I be an open interval in $\mathbb{R}$, let $\mu_t : I\to\mathscr{P}_p(X)$ be an absolutely continuous curve and let $|\mu'|\in L^1(I)$ be its metric derivative, given by Theorem 1.1.2. Then there exists a Borel vector field $v:(x,t)\mapsto v_t(x)$ such that*

$$v_t\in L^p(\mu_t;X), \qquad \|v_t\|_{L^p(\mu_t;X)}\leq |\mu'|(t) \qquad \textit{for } \mathscr{L}^1\textit{-a.e. } t\in I, \tag{8.3.6}$$

and the continuity equation

$$\partial_t\,\mu_t + \nabla\cdot(v_t\mu_t)=0 \qquad \textit{in } X\times I \tag{8.3.7}$$

holds in the sense of distributions, i.e.

$$\int_I\int_X \Big(\partial_t\varphi(x,t)+\langle v_t(x),\nabla_x\varphi(x,t)\rangle\Big)\,d\mu_t(x)\,dt = 0 \quad \forall\,\varphi\in\mathrm{Cyl}(X\times I). \tag{8.3.8}$$

Moreover, for $\mathscr{L}^1$-a.e. $t\in I$ $j_p(v_t)$ belongs to the closure in $L^q(\mu_t,X)$ of the subspace generated by the gradients $\nabla\varphi$ with $\varphi\in\mathrm{Cyl}(X)$.
Conversely, if a narrowly continuous curve $\mu_t : I\to\mathscr{P}_p(X)$ satisfies the continuity equation for some Borel velocity field v_t with $\|v_t\|_{L^p(\mu_t;X)}\in L^1(I)$ then $\mu_t : I\to\mathscr{P}_p(X)$ is absolutely continuous and $|\mu'|(t)\leq\|v_t\|_{L^p(\mu_t;X)}$ for $\mathscr{L}^1$-a.e. $t\in I$.

Proof. Taking into account Lemma 1.1.4 and Lemma 8.1.3, we will assume with no loss of generality that $|\mu'|\in L^\infty(I)$ in the proof of the first statement. To fix the ideas, we also assume that $I=(0,1)$.

First of all we show that for every $\varphi\in\mathrm{Cyl}(X)$ the function $t\mapsto\mu_t(\varphi)$ is absolutely continuous, and its derivative can be estimated with the metric derivative of μ_t. Indeed, for s, $t\in I$ we have, for $\boldsymbol{\mu}_{st}\in\Gamma_o(\mu_s,\mu_t)$ and using the Hölder inequality,

$$|\mu_t(\varphi)-\mu_s(\varphi)| = \left|\int_{X\times X}\big(\varphi(y)-\varphi(x)\big)\,d\boldsymbol{\mu}_{st}\right| \leq \mathrm{Lip}(\varphi)W_p(\mu_s,\mu_t),$$

whence the absolute continuity follows. In order to estimate more precisely the derivative of $\mu_t(\varphi)$ we introduce the upper semicontinuous and bounded map

$$H(x,y):=\begin{cases}|\nabla\varphi(x)| & \text{if } x=y,\\[2ex] \dfrac{|\varphi(x)-\varphi(y)|}{|x-y|} & \text{if } x\neq y,\end{cases}$$

and notice that, setting $\boldsymbol{\mu}_h = \boldsymbol{\mu}_{(s+h)s}$, we have

$$\begin{aligned} \frac{|\mu_{s+h}(\varphi) - \mu_s(\varphi)|}{|h|} &\leq \frac{1}{|h|} \int_{X\times X} |x-y| H(x,y)\, d\boldsymbol{\mu}_h \\ &\leq \frac{W_p(\mu_{s+h},\mu_s)}{|h|} \left(\int_{X\times X} H^q(x,y)\, d\boldsymbol{\mu}_h \right)^{1/q}, \end{aligned}$$

where q is the conjugate exponent of p. If t is a point where $s \mapsto \mu_s$ is metrically differentiable, using the fact that $\boldsymbol{\mu}_h \to (x,x)_\# \mu_t$ narrowly (because their marginals are narrowly converging, any limit point belongs to $\Gamma_o(\mu_t,\mu_t)$ and is concentrated on the diagonal of $X \times X$) we obtain

$$\limsup_{h\to 0} \frac{|\mu_{t+h}(\varphi) - \mu_t(\varphi)|}{|h|} \leq |\mu'|(t) \left(\int_X |H|^q(x,x)\, d\mu_t \right)^{1/q} = |\mu'|(t) \|\nabla\varphi\|_{L^q(\mu_t;X)}. \tag{8.3.9}$$

Set $Q = X \times I$ and let $\mu = \int \mu_t\, dt \in \mathscr{P}(Q)$ be the measure whose disintegration is $\{\mu_t\}_{t\in I}$. For any $\varphi \in \mathrm{Cyl}(Q)$ we have

$$\begin{aligned} \int_Q \partial_s \varphi(x,s)\, d\mu(x,s) &= \lim_{h\downarrow 0} \int_Q \frac{\varphi(x,s) - \varphi(x,s-h)}{h}\, d\mu(x,s) \\ &= \lim_{h\downarrow 0} \int_I \frac{1}{h} \Big(\int_X \varphi(x,s)\, d\mu_s(x) - \int_X \varphi(x,s)\, d\mu_{s+h}(x) \Big)\, ds. \end{aligned}$$

Taking into account (8.3.9), Fatou's Lemma yields

$$\begin{aligned} \left| \int_Q \partial_s \varphi(x,s)\, d\mu(x,s) \right| &\leq \int_J |\mu'|(s) \Big(\int_X |\nabla\varphi(x,s)|^q\, d\mu_s(x) \Big)^{1/q} ds \\ &\leq \Big(\int_J |\mu'|^p(s)\, ds \Big)^{1/p} \Big(\int_Q |\nabla\varphi(x,s)|^q\, d\mu(x,s) \Big)^{1/q}, \end{aligned} \tag{8.3.10}$$

where $J \subset I$ is any interval such that $\operatorname{supp}\varphi \subset J \times X$. If $\mathscr{V}$ denotes the closure in $L^q(\mu;X)$ of the subspace $V := \Big\{ \nabla\varphi, \quad \varphi \in \mathrm{Cyl}(Q) \Big\}$, the previous formula says that the linear functional $L : V \to \mathbb{R}$ defined by

$$L(\nabla\varphi) := -\int_Q \partial_s \varphi(x,s)\, d\mu(x,s)$$

can be uniquely extended to a bounded functional on $\mathscr{V}$. Therefore the minimum problem

$$\min \left\{ \frac{1}{q} \int_Q |w(x,s)|^q\, d\mu(x,s) - L(w) : \; w \in \mathscr{V} \right\} \tag{8.3.11}$$

admits a unique solution $w \in \mathscr{V}$ such that $v := j_q(w)$ satisfies

$$\int_Q \langle v(x,s), \nabla\varphi(x,s) \rangle\, d\mu(x,s) = \langle L, \nabla\varphi \rangle \qquad \forall \varphi \in \mathrm{Cyl}(Q). \tag{8.3.12}$$

Setting $v_t(x) = v(x,t)$ and using the definition of L we obtain (8.3.8). Moreover, choosing a sequence $(\nabla\varphi_n) \subset V$ converging to w in $L^q(\mu;X)$, it is easy to show that for $\mathscr{L}^1$-a.e. $t \in I$ there exists a subsequence $n(i)$ (possibly depending on t) such that $\nabla\varphi_{n(i)}(\cdot,t) \in \mathrm{Cyl}(X)$ converge in $L^q(\mu_t;X)$ to $w(\cdot,t) = j_p(v(\cdot,t))$.

Finally, choosing an interval $J \subset I$ and $\eta \in C_c^\infty(J)$ with $0 \le \eta \le 1$, (8.3.12) and (8.3.10) yield

$$\begin{aligned}
&\int_Q \eta(s)|v(x,s)|^p\,d\mu(x,s) = \int_Q \eta\langle v,w\rangle\,d\mu = \lim_{n\to\infty}\int_Q \eta\langle v,\nabla\varphi_n\rangle\,d\mu\\
&= \lim_{n\to\infty}\langle L,\nabla(\eta\varphi_n)\rangle \le \Big(\int_J |\mu'|^p(s)\,ds\Big)^{1/p}\lim_{n\to\infty}\Big(\int_{X\times J}|\nabla\varphi_n|^q\,d\mu\Big)^{1/q}\\
&= \Big(\int_J |\mu'|^p(s)\,ds\Big)^{1/p}\Big(\int_{X\times J}|w|^q\,d\mu\Big)^{1/q} = \Big(\int_J |\mu'|^p(s)\,ds\Big)^{1/p}\Big(\int_{X\times J}|v|^p\,d\mu\Big)^{1/q}.
\end{aligned}$$

Taking a sequence of smooth approximations of the characteristic function of J we obtain

$$\int_J\int_X |v_s(x)|^p\,d\mu_s(x)\,ds \le \int_J |\mu'|^p(s)\,ds, \tag{8.3.13}$$

and therefore

$$\|v_t\|_{L^p(\mu_t,X)} \le |\mu'|(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t\in I.$$

Now we show the converse implication, assuming first that $X = \mathbb{R}^d$. We apply the regularization Lemma 8.1.9, finding approximations μ_t^ε, v_t^ε satisfying the continuity equation, the uniform integrability condition (8.1.2) and the local regularity assumptions (8.1.7). Therefore, we can apply Proposition 8.1.8, obtaining the representation formula $\mu_t^\varepsilon = (T_t^\varepsilon)_\#\mu_0^\varepsilon$, where T_t^ε is the maximal solution of the ODE $\dot T_t^\varepsilon = v_t^\varepsilon(T_t^\varepsilon)$ with the initial condition $T_0^\varepsilon = x$ (see Lemma 8.1.4).

Now, taking into account Lemma 8.1.10, we estimate

$$\begin{aligned}
\int_{\mathbb{R}^d}|T_{t_2}^\varepsilon(x)-T_{t_1}^\varepsilon(x)|^p\,d\mu_0^\varepsilon &\le (t_2-t_1)^{p-1}\int_{\mathbb{R}^d}\int_{t_1}^{t_2}|\dot T_t^\varepsilon(x)|^p\,dt\,d\mu_0^\varepsilon \qquad (8.3.14)\\
&= (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d}|v_t^\varepsilon(x)|^p\,d\mu_t^\varepsilon\,dt\\
&\le (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d}|v_t|^p\,d\mu_t dt,
\end{aligned}$$

therefore the transport plan $\boldsymbol{\gamma}^\varepsilon := (T_{t_1}^\varepsilon\times T_{t_2}^\varepsilon)_\#\mu_0^\varepsilon$ satisfies

$$W_p^p(\mu_{t_1}^\varepsilon,\mu_{t_2}^\varepsilon) \le \int_{\mathbb{R}^{2d}}|x-y|^p\,d\boldsymbol{\gamma}^\varepsilon \le (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d}|v_t|^p\,d\mu_t\,dt.$$

Since for every $t\in I$ μ_t^ε converges narrowly to μ_t as $\varepsilon\to 0$, Lemma 7.1.3 shows that for any limit point $\boldsymbol{\gamma}$ of $\boldsymbol{\gamma}^\varepsilon$ we have

$$W_p^p(\mu_{t_1},\mu_{t_2}) \le \int_{\mathbb{R}^{2d}}|x-y|^p\,d\boldsymbol{\gamma} \le (t_2-t_1)^{p-1}\int_{t_1}^{t_2}\int_{\mathbb{R}^d}|v_t|^p\,d\mu_t dt.$$

Since t_1 and t_2 are arbitrary this implies that μ_t is absolutely continuous and that its metric derivative is less than $\|v_t\|_{L^p(\mu_t;X)}$ for $\mathscr{L}^1$-a.e. $t \in I$.

We conclude the proof considering the general infinite-dimensional case and following a typical reduction argument, by projecting measures on finite dimensional subspaces. Let $\pi^d : X \to \mathbb{R}^d$ be the canonical maps, given by (5.1.28) for an orthonormal basis (e_n) of X, let $\mu_t^d := \pi^d_\# \mu_t \in \mathscr{P}(\mathbb{R}^d)$, and let $\{\mu_{ty}\}_{y\in\mathbb{R}^d}$ be the disintegration of μ_t with respect to μ_t^d as in Theorem 5.3.1. Notice that considering test functions $\varphi = \psi \circ \pi^d$ in (8.1.3), with $\nabla\varphi = (\pi^d)^* \circ \nabla\psi \circ \pi^d$, gives

$$
\begin{aligned}
\frac{d}{dt}\int_X \varphi\, d\mu_t(x) &= \int_X \langle \pi^d(v_t), \nabla\psi \circ \pi^d\rangle\, d\mu_t(x)\\
&= \int_{\mathbb{R}^d}\Big(\int_{(\pi^d)^{-1}(y)} \langle \pi^d(v_t), \nabla\psi\circ\pi^d\rangle\, d\mu_{ty}(x)\Big)\, d\mu_t^d(y)\\
&= \int_{\mathbb{R}^d}\langle \int_{(\pi^d)^{-1}(y)} \pi^d\big(v_t(x)\big)\, d\mu_{ty}(x), \nabla\psi(y)\rangle\, d\mu_t^d(y) = \int_{\mathbb{R}^d}\langle v_t^d(y), \nabla\psi(y)\rangle\, d\mu_t^d(y),
\end{aligned}
$$

with $v_t^d(y) := \int_{(\pi^d)^{-1}(y)} \pi^d\big(v_t^d(x)\big)\, d\mu_{ty}(x)$, and therefore

$$
\partial_t\, \mu_t^d + \nabla\cdot(v_t^d \mu_t^d) = 0 \qquad \text{in } \mathbb{R}^d \times I.
$$

Notice also that, by similar calculations,

$$
\begin{aligned}
\left|\int_{\mathbb{R}^d}\langle v_t^d(y), \chi(y)\rangle\, d\mu_t^d(y)\right| &= \left|\int_X \langle \pi^d(v_t(x)), \chi(\pi^d(x))\rangle\, d\mu_t\right|\\
&\le \|v_t\|_{L^p(\mu_t;X)}\|\chi\|_{L^q(\mu_t^d;\mathbb{R}^d)}
\end{aligned}
$$

for any $\chi \in L^\infty(\mu_t^d;\mathbb{R}^d)$, hence $\|v_t^d\|_{L^p(\mu_t^d;\mathbb{R}^d)} \le \|v_t\|_{L^p(\mu_t;X)}$. Therefore $t \mapsto \mu_t^d$ is an absolutely continuous curve in $\mathscr{P}_p(\mathbb{R}^d)$ and

$$
W_p(\mu_{t_1}^d, \mu_{t_2}^d) \le \int_{t_1}^{t_2} \|v_t^d\|_{L^p(\mu_t^d;\mathbb{R}^d)}\, dt \le \int_{t_1}^{t_2} \|v_t\|_{L^p(\mu_t;X)}\, dt \quad \forall t_1,\, t_2 \in I,\ t_1 \le t_2.
$$

Let now

$$
\hat\mu_t^d = (\pi^d)^*_\# \mu_t^d = \hat\pi^d_\# \mu_t,
$$

be the image of the measures μ_t^d under the isometries $(\pi^d)^* : y \mapsto \sum_1^d y_i e_i$. Passing to the limit as $d \to \infty$ and using the narrow convergence of $\hat\mu_t^d$ to μ_t and (7.1.11) we obtain

$$
W_p(\mu_{t_1}, \mu_{t_2}) \le \int_{t_1}^{t_2} \|v_t\|_{L^p(\mu_t,X)}\, dt \qquad \forall t_1,\, t_2 \in I,\ t_1 \le t_2.
$$

This proves that μ_t is absolutely continuous and that its metric derivative can be estimated with $\|v_t\|_{L^p(\mu_t;X)}$. □

In the case when the measures are constant in time, by combining the previous finite dimensional projection argument and the smoothing technique of Lemma 8.1.9, one obtains an important approximation property. Let us first collect some preliminary useful properties of orthogonal projections of measures and vector fields, some of which we already proved in the last part of the above proof.

Lemma 8.3.2 (Finite dimensional projection of vector fields). *Let $\mu \in \mathscr{P}_p(X)$, $v \in L^p(\mu; X)$, and let $\{e_n\}_{n=1}^\infty$ be a complete orthonormal system of X, with the associated canonical maps $\pi^d, (\pi^d)^*, \hat\pi^d$ given by (5.1.28), (5.1.29), and (5.1.30). We consider the finite dimensional subspaces $X^d := \mathrm{span}(e_1, \ldots, e_d)$, the measures $\hat\mu^d := \hat\pi^d_\# \mu$, the disintegration $\{\mu_x\}_{x \in X^d}$ of μ w.r.t. $\hat\mu^d$ given by Theorem 5.3.1, and the vector field*

$$\hat v^d(x) := \int_{(\hat\pi^d)^{-1}(x)} \hat\pi^d(v(y))\, d\mu_x(y) \quad \text{for } \hat\mu^d\text{-a.e. } x \in X^d. \tag{8.3.15}$$

The following properties hold:

(i) $\mathrm{supp}\,\hat\mu^d \subset X^d$, $\hat\mu^d \to \mu$ *in $\mathscr{P}_p(X)$ as $d \to \infty$. If μ is regular then also $\hat\mu^d|_{X^d}$ is regular;*

(ii) $\hat v^d \in L^p(\hat\mu^d; X^d)$ *with*

$$\|\hat v^d\|_{L^p(\hat\mu^d;X^d)} \le \|v\|_{L^p(\mu;X)}; \tag{8.3.16}$$

(iii) $\hat v^d$ *is characterized by the following identity*

$$\int_X \langle \boldsymbol{\zeta}(x), \hat v^d(x)\rangle\, d\hat\mu^d(x) = \int_X \langle \hat\pi^d \boldsymbol{\zeta}(\hat\pi^d(x)), v(x)\rangle\, d\mu(x), \tag{8.3.17}$$

for every bounded Borel vector field $\boldsymbol{\zeta} : X \to X$;

(iv) *If $\nabla \cdot (v\mu) = 0$ (in the duality with smooth cylindrical maps), then also $\nabla \cdot (\hat v^d \hat\mu^d) = 0$;*

(v) *for every continuous function $f : X \times X \to \mathbb{R}$ with p-growth according to (5.1.21) we have*

$$\lim_{d\to\infty} \int_{X\times X} f(x, \hat v^d(x))\, d\hat\mu^d(x) = \int_{X\times X} f(x, v(x))\, d\mu(x). \tag{8.3.18}$$

In particular, $\hat v^d \hat\mu^d \to v\mu$ in the duality with $C^0_b(X;X)$ and

$$\lim_{d\to\infty} \|\hat v^d\|_{L^p(\mu^d;X)} = \|v\|_{L^p(\mu;X)}. \tag{8.3.19}$$

Proof. (i) is immediate and we have seen in the previous proof that (ii) is a direct consequence of (iii); in order to check this point we simply use the Definition

(8.3.15) of $\hat v^d$ obtaining

$$\begin{aligned}\int_X \langle \boldsymbol\zeta(x), \hat v^d(x)\rangle\, d\hat\mu^d(x) &= \int_X \langle \boldsymbol\zeta(x), \int_{(\hat\pi^d)^{-1}(x)} \hat\pi^d v(y)\, d\mu_x(y)\rangle\, d\hat\mu^d(x)\\ &= \int_X \int_{(\hat\pi^d)^{-1}(x)} \langle \boldsymbol\zeta(\hat\pi^d(y)), \hat\pi^d v(y)\rangle\, d\mu_x(y)\, d\hat\mu^d(x)\\ &= \int_X \langle \boldsymbol\zeta(\hat\pi^d(x)), \hat\pi^d v(x)\rangle\, d\mu(x) = \int_X \langle \hat\pi^d \boldsymbol\zeta(x), v(x)\rangle\, d\mu(x).\end{aligned}$$

(iv) follows by (iii) simply choosing $\boldsymbol\zeta := \nabla(\hat\chi^d_R\varphi)$, for $\varphi \in \mathrm{Cyl}(X)$ and $\hat\chi^d_R := \chi_R \circ \pi^d$ as in (8.1.15), and observing that

$$\hat\pi^d\big(\nabla(\hat\chi^d_R(\hat\pi^d)\varphi(\hat\pi^d))\big) = \nabla\big((\hat\chi^d_R\varphi)\circ\hat\pi^d\big), \quad (\hat\chi^d_R\varphi)\circ\hat\pi^d \in \mathrm{Cyl}(X).$$

Therefore we get

$$\int_X \langle \nabla\varphi, \hat v^d\rangle\, d\hat\mu^d = \lim_{R\uparrow+\infty} \int_X \langle \nabla(\hat\chi^d_R\varphi), \hat v^d\rangle\, d\hat\mu^d = \lim_{R\uparrow+\infty} \int_X \langle \nabla\big((\hat\chi^d_r\varphi)\circ\hat\pi^d\big), v\rangle\, d\mu = 0.$$

Finally, (8.3.17) easily yields

$$\lim_{d\to\infty} \int_X \langle \boldsymbol\zeta, \hat v^d\rangle\, d\hat\mu^d = \int_X \langle \boldsymbol\zeta, v\rangle\, d\mu \quad \forall\, \boldsymbol\zeta \in C^0_b(X;X); \tag{8.3.20}$$

taking into account of (8.3.16), of Definition 5.4.3, and of Theorem 5.4.4, we conclude. □

Proposition 8.3.3 (Approximation by regular measures). *For any $\mu \in \mathscr{P}_p(X)$, any $v \in L^p(\mu;X)$ such that $\nabla\cdot(v\mu) = 0$ (in the duality with smooth cylindrical functions), and any complete orthonormal system $\{e_n\}_{n\ge1}$, there exist measures $\mu_h \in \mathscr{P}_p(X)$ and vectors $v_h \in L^p(\mu_h;X)$, $h\in\mathbb{N}$, such that*

i. $\operatorname{supp}\mu_h \subset X_h := \operatorname{span}(e_1,\dots,e_h)$ *(in the finite dimensional case we simply set $X_h = X$),*

ii. $\mu_h|_{X_h} \in \mathscr{P}^r_p(X_h)$,

iii. $v_h(x) \in X_h(x) \quad \forall\, x\in X, \qquad \nabla\cdot(v_h\mu_h) = 0,$

iv. $\mu_h \to \mu$ *in $\mathscr{P}_p(X)$ as $h\to\infty$,*

v. *for every continuous function $f : X\times X \to \mathbb{R}$ with p-growth according to* (5.1.21) *we have*

$$\lim_{h\to\infty} \int_{X\times X} f(x, v_h(x))\, d\mu_h(x) = \int_{X\times X} f(x, v(x))\, dx. \tag{8.3.21}$$

In particular, $v_h\mu_h \to v\mu$ in the duality with $C^0_b(X;X)$ and

$$\lim_{h\to\infty} \|v_h\|_{L^p(\mu_h;X)} = \|v\|_{L^p(\mu;X)}.$$

Proof. To each finite dimensional measure and vector field provided by Lemma 8.3.2 we apply the smoothing argument of Lemma 8.1.9; the proof is achieved by a simple diagonal argument. □

8.4 The tangent bundle to $\mathscr{P}_p(X)$

Notice that the continuity equation (8.3.7) involves only the action of v_t on $\nabla\varphi$ with $\varphi \in \mathrm{Cyl}(X)$. Moreover, Theorem 8.3.1 shows that the minimal norm among all possible velocity fields v_t is the metric derivative and that $j_p(v_t)$ belongs to the L^q closure of gradients of functions in $\mathrm{Cyl}(X)$. These facts suggest a "canonical" choice of v_t and the following definition of tangent bundle to $\mathscr{P}_p(X)$.

Definition 8.4.1 (Tangent bundle). *Let $\mu \in \mathscr{P}_p(X)$. We define*

$$\mathrm{Tan}_\mu \mathscr{P}_p(X) := \overline{\{j_q(\nabla\varphi) : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^p(\mu;X)},$$

where $j_q : L^q(\mu; X) \to L^p(\mu; X)$ is the duality map defined in (8.3.1) .

Notice also that $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ can be equivalently defined as the image under j_q of the L^q closure of gradients of smooth cylindrical functions in X. The choice of $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ is motivated by the following variational selection principle (nonlinear in the case $p \neq 2$):

Lemma 8.4.2 (Variational selection of the tangent vectors). *A vector $v \in L^p(\mu; X)$ belongs to the tangent cone $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ iff*

$$\|v + w\|_{L^p(\mu;X)} \geq \|v\|_{L^p(\mu;X)} \quad \forall\, w \in L^p(\mu; X) \text{ such that } \nabla \cdot (w\mu) = 0. \tag{8.4.1}$$

In particular, for every $v \in L^p(\mu; X)$ there exists a unique $\Pi(v) \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ in the equivalence class of v modulo divergence-free vector fields, $\Pi(v)$ is the element of minimal L^p-norm in this class, and

$$\int_X \langle j_p(v), w - \Pi(w)\rangle \, d\mu(x) = 0 \quad \forall\, v \in \mathrm{Tan}_\mu \mathscr{P}_p(X), \ w \in L^p(\mu; X). \tag{8.4.2}$$

Proof. By the convexity of the L^p norm, (8.4.1) holds iff

$$\int_X \langle j_p(v), w\rangle \, d\mu = 0 \quad \text{for any } w \in L^p(\mu; X) \text{ s.t. } \nabla \cdot (w\mu) = 0 \tag{8.4.3}$$

(here the divergence is understood making the duality with smooth cylindrical test functions) and this is true iff $j_p(v)$ belongs to the L^q closure of $\{\nabla\phi : \ \phi \in \mathrm{Cyl}(X)\}$. Therefore $v = j_q(j_p(v))$ belongs to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$. (8.4.2) follows from (8.4.3) since $w - \Pi(w)$ is divergence free. □

Observe that the projection Π is linear and $\mathrm{Tan}_\mu\mathscr{P}_p(X)$ is a vector space only in the Hilbertian case $p = q = 2$.

The remarks above lead also to the following characterization of divergence-free vector fields:

Proposition 8.4.3. *Let $w \in L^p(\mu; X)$. Then $\nabla \cdot (w\mu) = 0$ iff $\|v - w\|_{L^p(\mu;X)} \geq \|v\|_{L^p(\mu;X)}$ for any $v \in \mathrm{Tan}_\mu\mathscr{P}_p(X)$. Moreover equality holds for some v iff $w = 0$.*

Proof. We already proved that $\nabla \cdot (w\mu) = 0$ implies $\|v - w\|_{L^p(\mu;X)} \geq \|v\|_{L^p(\mu;X)}$ for any $v \in \mathrm{Tan}_\mu\mathscr{P}_p(X)$. Let us prove now the opposite implication. Indeed, being $\mathrm{Tan}_\mu\mathscr{P}_p(X)$ a cone, a differentiation yields

$$\int_X \langle j_p(v), w\rangle \, d\mu = 0 \qquad \forall v \in \mathrm{Tan}_\mu\mathscr{P}_p(X),$$

and choosing $v = j_q(\nabla\varphi)$, with $\varphi \in \mathrm{Cyl}(X)$, we obtain $\int_X \langle \nabla\varphi, w\rangle \, d\mu = 0$ for any $\varphi \in \mathrm{Cyl}(X)$.

We give now an elementary proof of the fact that if equality holds for some v, then $w = 0$. If equality holds for some v the convexity of the L^p norm gives $\|v + tw\|_{L^p(\mu;X)} = \|v\|_{L^p(\mu;X)}$ for any $t \in [0, 1]$, and differentiation with respect to t gives

$$\int_X |v + tw|^{p-2}\langle v + tw, w\rangle \, d\mu = 0 \qquad \forall t \in (0, 1).$$

Differentiating once more (and using the monotone convergence theorem and the convexity of the map $t \mapsto |a + tb|^p$) we eventually obtain

$$\int_X |v + tw|^{p-2} \left[|w|^2 + (p-2)\frac{(\langle v + tw, w\rangle)^2}{|v + tw|^2} \right] d\mu = 0 \quad \forall t \in (0, 1).$$

Since the integrand is nonnegative it immediately follows that $w = 0$. □

In the particular case $p = 2$ the map j_2 is the identity and (8.4.3) gives

$$\mathrm{Tan}_\mu^\perp\mathscr{P}_2(X) = \left\{ v \in L^2(\mu, X) : \ \nabla \cdot (v\mu) = 0 \right\}. \tag{8.4.4}$$

Remark 8.4.4 (Cotangent space, duality, and quotients). Since tangent vectors acts naturally only on gradient vector fields, one could also define the *cotangent space* as

$$\mathrm{CoTan}_\mu\mathscr{P}_p(X) := \overline{\{\nabla\varphi : \ \varphi \in \mathrm{Cyl}(X)\}}^{L^q(\mu;X)}, \tag{8.4.5}$$

and therefore the tangent space by duality. If $\sim$ denotes the equivalence relation which identifies two vector fields in $L^p(\mu; X)$ if their difference is divergence free, the tangent space could be identified with the quotient space $L^p(\mu; X)/\sim$. Definition 8.4.1 and the related lemma 8.4.2 simply operates a canonical (though nonlinear) selection of an element $\Pi(v)$ in the class of v by using the duality map between the Cotangent and the Tangent space. This distinction becomes superfluous in the Hilbertian case $p = q = 2$, since in that case the tangent and the cotangent spaces turn out to be the same, by the usual identification via the Riesz isomorphism.

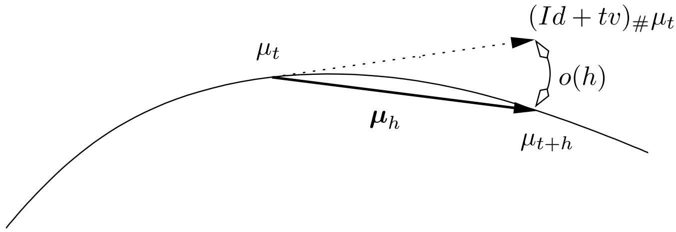
The following two propositions show that the notion of tangent space is consistent with the metric structure, with the continuity equation, and with optimal transport maps (if any).

Proposition 8.4.5 (Tangent vector to a.c. curves). *Let* $\mu_t : I \to \mathscr{P}_p(X)$ *be an absolutely continuous curve and let* $v_t \in L^p(\mu_t; X)$ *be such that* (8.3.7) *holds. Then* v_t *satisfies* (8.3.6) *as well if and only if* $v_t = \Pi(v_t) \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ *for* $\mathscr{L}^1$*-a.e.* $t \in I$*. The vector* v_t *is uniquely determined* $\mathscr{L}^1$*-a.e. in* I *by* (8.3.6) *and* (8.3.7)*.*

Proof. The uniqueness of v_t is a straightforward consequence of the linearity with respect to the velocity field of the continuity equation and of the strict convexity of the L^p norm.

In the proof of Theorem 8.3.1 we built vector fields $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ satisfying (8.3.6) and (8.3.7). By uniqueness, it follows that conditions (8.3.6) and (8.3.7) imply $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ for $\mathscr{L}^1$-a.e. t. □

In the following proposition we recover the tangent vector field to a curve through the infinitesimal behaviour of optimal transport maps, or plans, along the curve. Notice that in the limit we recover a plan $(\boldsymbol{i}\times v_t)_\#\mu_t$ associated to a *classical* transport even in the situation when μ_t are not necessarily absolutely continuous. It is for this reason that we don't need, at least for differential calculus along absolutely continuous curves, the more general notions of tangent space, made by plans instead of maps, discussed in the Appendix.



Proposition 8.4.6 (Optimal plans along a.c. curves). *Let* $\mu_t : I \to \mathscr{P}_p(X)$ *be an absolutely continuous curve and let* $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_p(X)$ *be characterized by Proposition* 8.4.5*. Then, for* $\mathscr{L}^1$*-a.e.* $t \in I$ *the following property holds: for any choice of* $\boldsymbol{\mu}_h \in \Gamma_o(\mu_t, \mu_{t+h})$ *we have*

$$\lim_{h\to 0}\big(\pi^1, \frac{1}{h}(\pi^2 - \pi^1)\big)_\#\boldsymbol{\mu}_h = (\boldsymbol{i} \times v_t)_\#\mu_t \qquad \text{in } \mathscr{P}_p(X \times X) \tag{8.4.6}$$

and

$$\lim_{h\to 0}\frac{W_p(\mu_{t+h}, (\boldsymbol{i} + hv_t)_\#\mu_t)}{|h|} = 0. \tag{8.4.7}$$

In particular, for $\mathscr{L}^1$-a.e. $t \in I$ such that $\mu_t \in \mathscr{P}_p^r(X)$ we have

$$\lim_{h\to 0} \frac{1}{h}(\boldsymbol{t}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{i}) = v_t \qquad \text{in } L^p(\mu_t; X), \tag{8.4.8}$$

where $\boldsymbol{t}_{\mu_t}^{\mu_{t+h}}$ is the unique optimal transport map between μ_t and μ_{t+h}.

Proof. Let $\mathscr{D}_d \subset C_c^\infty(\mathbb{R}^d)$ be a countable set with the following property: for any integer $R > 0$ and any $\psi \in C_c^\infty(\mathbb{R}^d)$ with $\operatorname{supp}\psi \subset B_R$ there exist $(\varphi_n) \subset \mathscr{D}_d$ with $\operatorname{supp}\varphi_n \subset B_R$ and $\varphi_n \to \varphi$ in $C^1(\mathbb{R}^d)$. Let also $\Pi_d \subset \Pi_d(X)$ be a a countable set with the following property: for any $\pi \in \Pi_d(X)$ there exist $\pi_n \in \Pi_d$ such that $\pi_n \to \pi$ uniformly on bounded sets of X (the existence of Π_d follows easily by the separability of X).

We fix $t \in I$ such that $W_p(\mu_{t+h}, \mu_t)/|h| \to |\mu'|(t) = \|v_t\|_{L^p(\mu_t)}$ and

$$\lim_{h\to 0} \frac{\mu_{t+h}(\varphi) - \mu_t(\varphi)}{h} = \int_{\mathbb{R}^d} \langle \nabla\varphi, v_t\rangle \, d\mu_t \qquad \forall \varphi = \psi\circ\pi, \ \psi \in \mathscr{D}_d, \ \pi \in \Pi_d. \tag{8.4.9}$$

Since $\mathscr{D}_d$ and Π_d are countable, the metric differentiation theorem implies that both conditions are fulfilled for $\mathscr{L}^1$-a.e. $t \in I$. Let $\boldsymbol{\mu}_h \in \Gamma_o(\mu_t, \mu_{t+h})$, set

$$\boldsymbol{\nu}_h := \left(\pi^1, \frac{1}{h}(\pi^2 - \pi^1)\right)_\# \boldsymbol{\mu}_h,$$

and fix φ as in (8.4.9) and a limit point $\boldsymbol{\nu}_0 = \int \nu_{0x}\, d\mu_t(x)$ of $\boldsymbol{\nu}_h$ as $h \to 0$ (w.r.t. the narrow convergence). We use the identity

$$\begin{aligned}\frac{\mu_{t+h}(\varphi) - \mu_t(\varphi)}{h} &= \frac{1}{h}\int_{X\times X} \varphi(y) - \varphi(x)\, d\boldsymbol{\mu}_h \\ = \frac{1}{h}\int_{X\times X} \varphi(x + h(y-x)) - \varphi(x)\, d\boldsymbol{\nu}_h &= \int_{X\times X} \langle \nabla\varphi(x), y - x\rangle + \omega_{x,y}(h)\, d\boldsymbol{\nu}_h\end{aligned}$$

with $\omega_{x,y}(h)$ bounded and infinitesimal as $h \to 0$, to obtain

$$\int_X \langle \nabla\varphi, v_t\rangle\, d\mu_t = \int_X \int_X \langle y, \nabla\varphi(x)\rangle \, d\nu_{0x}(y)\, d\mu_t(x).$$

Denoting by $\tilde v_t(x) = \int_X y\, d\nu_{0x}(y)$ the first moment of ν_{0x}, by a density argument it follows that

$$\nabla\cdot((\tilde v_t - v_t)\mu_t) = 0. \tag{8.4.10}$$

We now claim that

$$\int_X \int_X |y|^p \, d\nu_{0x}(y) d\mu_t(x) \le [|\mu'|(t)]^p. \tag{8.4.11}$$

Indeed

$$\begin{aligned}\int_X \int_X |y|^p \, d\nu_{0x}(y) d\mu_t(x) &\leq \liminf_{h \to 0} \int_{X \times X} |y|^p \, d\boldsymbol{\nu}_h \\ &= \liminf_{h \to 0} \frac{1}{h^p} \int_{X \times X} |y - x|^p d\boldsymbol{\mu}_h \\ &= \liminf_{h \to 0} \frac{W_p^p(\mu_{t+h}, \mu_t)}{h^p} = |\mu'|^p(t).\end{aligned}$$

From (8.4.11) we obtain that

$$\|\tilde{v}_t\|^p_{L^p(\mu_t;X)} \leq \int_X \int_X |y|^p \, d\nu_{0x} d\mu_t(x) \leq [|\mu'|(t)]^p = \|v_t\|^p_{L^p(\mu_t;X)}.$$

Therefore Proposition 8.4.3 entails that $\tilde{v}_t = v_t$. Moreover, the first inequality above is strict if ν_{0x} is not a Dirac mass in a set of μ_t-positive measure. Therefore ν_{0x} is a Dirac mass for μ_t-a.e. x and $\boldsymbol{\nu}_0 = (\boldsymbol{i} \times v_t)_\# \mu_t$. This proves the narrow convergence of the measures in (8.4.6). Together with convergence of moments, this gives convergence in the Wasserstein metric.

Now we show (8.4.7). Let $\boldsymbol{\mu}_h = \int_X \mu_{hx} \, d\mu_t(x)$ and let us estimate the distance between μ_{t+h} and $(\boldsymbol{i} + h v_t)_\# \mu_t$ with $\pi^{2,3}_\# \left(\int \delta_{x + h v_t} \times \nu_{hx} \, d\mu_t(x) \right)$. We have then

$$\begin{aligned}\frac{W_p^p(\mu_{t+h}, (\boldsymbol{i} + h v_t)_\# \mu_t)}{h^p} &\leq \int_{X \times X} \frac{1}{h^p} |x + h v_t(x) - y|^p \, d\boldsymbol{\mu}_h \\ &= \int_{X \times X} |v_t(x) - y|^p \, d\boldsymbol{\nu}_h = o(1)\end{aligned}$$

because of (8.4.6).

In the case when $\mu_t \in \mathscr{P}_p^r(X)$, the identity

$$\left(\pi^1, \frac{1}{h}(\pi^2 - \pi^1) \right)_\# \boldsymbol{\mu}_h = \left(\boldsymbol{i} \times \frac{1}{h}(\boldsymbol{t}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{i}) \right)_\# \mu_t$$

and the weak convergence at the level of plans give that $\frac{1}{h}(\boldsymbol{t}_{\mu_t}^{\mu_{t+h}} - \boldsymbol{i})\mu_t$ narrowly converge to $v_t \mu_t$. On the other hand our choice of t ensures that the L^p norms converge to the L^p norm of the limit, therefore the convergence of the densities of these measures w.r.t. μ_t is strong in L^p. □

As an application of (8.4.7) we are now able to show the $\mathscr{L}^1$-a.e. differentiability of $t \mapsto W_p(\mu_t, \sigma)$ along absolutely continuous curves μ_t. Recall that for constant speed geodesics more precise results hold, see Chapter 7.

Theorem 8.4.7 (Generic differentiability of $W_p(\mu_t, \sigma)$). *Let $\mu_t : I \to \mathscr{P}_p(X)$ be an absolutely continuous curve, let $\sigma \in \mathscr{P}_p(X)$ and let $v_t \in \mathrm{Tan}_{\mu_t} \mathscr{P}_p(X)$ be its* *tangent vector field, characterized by Proposition* 8.4.5. *Then*

$$\frac{d}{dt}W_p^p(\mu_t,\sigma) = \int_{X^2} p|x_1-x_2|^{p-2}\langle x_1-x_2, v_t(x_1)\rangle\, d\boldsymbol{\gamma} \qquad \forall \boldsymbol{\gamma}\in\Gamma_o(\mu_t,\sigma) \tag{8.4.12}$$

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

Proof. We show that the stated property is true at any t where (8.4.7) holds and the derivative of $t\mapsto W_p(\mu_t,\sigma)$ exists (recall that this map is absolutely continuous). Due to (8.4.7), we know that the limit

$$L := \lim_{h\to 0}\frac{W_p^p((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)-W_p^p(\mu_t,\sigma)}{h}$$

exists and coincides with $\frac{d}{dt}W_p^p(\mu_t,\sigma)$, and we have to show that it is equal to the left hand side in (8.4.12). Choosing any $\boldsymbol{\gamma}\in\Gamma_o(\mu_t,\sigma)$ we can use the plan $\boldsymbol{\eta} := (\pi^1+hv_t\circ\pi^1,\pi^2)_\#\boldsymbol{\gamma}\in\Gamma((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)$ to estimate from above $W_p^p((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma)$ as follows:

$$\begin{aligned} W_p^p((\boldsymbol{i}+hv_t)_\#\mu_t,\sigma) &\le \int_{X^2}|x_1-x_2|^p\,d\boldsymbol{\eta} = \int_{X^2}|x_1+hv_t(x_1)-x_2|^p\,d\boldsymbol{\gamma}\\ &= W_p^p(\mu_t,\sigma)+h\int_{X^2}p\langle\frac{x_1-x_2}{|x_1-x_2|^{2-p}},v_t(x_1)\rangle\,d\boldsymbol{\gamma}+o(h). \end{aligned}$$

Dividing both sides by h and taking limits as $h\downarrow 0$ or $h\uparrow 0$ we obtain

$$L\le\int_{X^2}p|x_1-x_2|^{p-2}\langle x_1-x_2,v_t(x_1)\rangle\,d\boldsymbol{\gamma}\le L. \qquad \square$$

The argument in the previous proof leads to the so-called superdifferentiability property of the Wasserstein distance, a theme that we will explore more in detail in Chapter 10 (see in particular Theorem 10.2.2).

Remark 8.4.8 (Derivative formula with an arbitrary velocity vector field). In fact, Proposition 8.5.4 will show that formula (8.4.12) holds for *every* Borel velocity vector field v_t satisfying the continuity equation in the distribution sense (8.3.8) and the L^p-estimate $\|v_t\|_{L^p(\mu_t;X)}\in L^1(I)$.

8.5 Tangent space and optimal maps

In this section we compare the tangent space arising from the closure of gradients of smooth cylindrical function with the tangent space built using optimal maps; the latter one is also compared in the Appendix with the geometric tangent space made with plans (see Theorem 12.4.4).

Proposition 8.4.6 suggests another possible definition of tangent cone to a measure in $\mathscr{P}_p(X)$ (see also Section 12.4 in the Appendix): for any $\mu\in\mathscr{P}_p(X)$ we define

$$\mathrm{Tan}^r_\mu\mathscr{P}_p(X) := \overline{\left\{\lambda(\boldsymbol{r}-\boldsymbol{i}) : (\boldsymbol{i}\times\boldsymbol{r})_\#\mu\in\Gamma_o(\mu,\boldsymbol{r}_\#\mu),\ \lambda>0\right\}}^{L^p(\mu;X)}. \tag{8.5.1}$$

The main result of this section shows that the two notions in fact coincide.

Theorem 8.5.1. *For any $p \in (1, +\infty)$ and any $\mu \in \mathscr{P}_p(X)$ we have*

$$\mathrm{Tan}_\mu \mathscr{P}_p(X) = \mathrm{Tan}^r_\mu \mathscr{P}_p(X).$$

We split the (not elementary) proof of this result in various steps, which are of independent interest.

The first step provides an inclusion between the tangent cones when the base measure μ is regular.

Proposition 8.5.2 (Optimal displacement maps are tangent). *If $p \in (1, +\infty)$ and $\mu \in \mathscr{P}^r_p(X)$, then $\mathrm{Tan}^r_\mu \mathscr{P}_p(X) \subset \mathrm{Tan}_\mu \mathscr{P}_p(X)$, i.e. for every measure $\sigma \in \mathscr{P}_p(X)$, if $\boldsymbol{t}_\mu^\sigma$ is the unique optimal transport map between μ and σ given by Theorem* 6.2.4 *and Theorem* 6.2.10*, we have $\boldsymbol{t}_\mu^\sigma - \boldsymbol{i} \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$.*

Proof. Assume first that $\operatorname{supp} \sigma$ is contained in $\overline{B}_R(0)$ for some $R > 0$. Theorem 6.2.4 ensures the representation $\boldsymbol{t}_\mu^\sigma - \boldsymbol{i} = j_q(\nabla\varphi)$, where φ is a locally Lipschitz and $|\cdot|^p$-concave map whose gradient $\nabla\phi = j_p(\boldsymbol{t}_\mu^\sigma - \boldsymbol{i})$ has $(p-1)$-growth (according to (5.1.21)), since $\boldsymbol{t}_\mu^\sigma$ takes its values in a bounded set.

We consider the Euclidean case $X = \mathbb{R}^d$ first and the mollified functions φ_ε. A truncation argument enabling an approximation by gradients with compact support gives that $j_q(\nabla\varphi_\varepsilon)$ belong to $\mathrm{Tan}_\mu \mathscr{P}_p(X)$ (notice also that $\nabla\varphi_\varepsilon$ have still $(p-1)$-growth, uniformly with respect to ϵ). Due to the absolute continuity of μ it is immediate to check using the dominated convergence theorem that $j_q(\nabla\varphi_\varepsilon)$ converge to $j_q(\nabla\varphi)$ in $L^p(\mu; \mathbb{R}^d)$, therefore $j_q(\nabla\varphi) \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ as well.

In the case when X is an infinite dimensional, separable Hilbert case we argue as follows. Let $\pi^d, (\pi^d)^*, \hat{\pi}^d$ be the canonical maps given by (5.1.28), (5.1.29), and (5.1.30) for an orthonormal basis $\{e_n\}_{n\ge 1}$ of X. We set

$$\mu^d := \pi^d_\# \mu,\ \nu^d := \pi^d_\# \nu \in \mathscr{P}(\mathbb{R}^d), \qquad \hat{\mu}^d := \hat{\pi}^d_\# \mu,\ \hat{\nu}^d := \hat{\pi}^d_\# \nu \in \mathscr{P}(X),$$

observing that, by (6.2.1) and (5.2.3), μ^d is absolutely continuous with respect to the d-dimensional Lebesgue measure. Therefore there exists an optimal transportation map $\boldsymbol{r}^d \in L^p(\mu^d; \mathbb{R}^d)$ defined on $\mathbb{R}^d$ such that $\boldsymbol{r}^d_\# \mu^d = \nu^d$ and $\boldsymbol{r}^d - \boldsymbol{i} = j_q(\nabla\psi^d)$ in $\mathbb{R}^d$ for some locally Lipschitz and $|\cdot|^p$-concave map $\psi^d : \mathbb{R}^d \to \mathbb{R}$. By the previous approximation argument, setting $\varphi^d := \psi^d \circ \pi^d$ and

$$\begin{aligned} \hat{\boldsymbol{r}}^d :=& (\pi^d)^* \circ (\boldsymbol{r}^d \circ \pi^d) = (\pi^d)^* \circ \big(j_q(\nabla\psi^d \circ \pi^d) + \pi^d\big) \\ =& j_q\big((\pi^d)^* \circ \nabla\psi^d \circ \pi^d\big) + (\pi^d)^* \circ \pi^d = j_q\big(\nabla\varphi^d\big) + \hat{\pi}^d \end{aligned}$$

(here we used the commutation property $j_q \circ (\pi^d)^* = (\pi^d)^* \circ j_q$), we get $\hat{\boldsymbol{r}}^d - \hat{\pi}^d \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$; moreover, being $(\pi^d)^*$ an isometry, it is immediate to check that $\hat{\boldsymbol{r}}^d$ is an optimal map pushing $\hat{\mu}^d$ on $\hat{\nu}^d$.

Letting $d \to +\infty$, since

$$\lim_{d \uparrow +\infty} \|\hat{\pi}^d - \boldsymbol{i}\|_{L^p(\mu; X)} = 0,$$

we conclude by applying the following Lemma.

Finally, when σ has not a bounded support, we can approximate σ in $\mathscr{P}_p(X)$ by measures σ_n with bounded support and we can apply again the following lemma. The details are left to the reader. □

Lemma 8.5.3. *Let μ, $\nu \in \mathscr{P}_p(X)$ such that $\Gamma_o(\mu,\nu) = \{(\boldsymbol{i}\times\boldsymbol{r})_\#\mu\}$ contains only an optimal transportation map $\boldsymbol{r} \in L^p(\mu;X)$, let $\boldsymbol{t}_n \in L^p(\mu;X)$ be a family of maps converging to the identity in $L^p(\mu;X)$ with $\mu_n := (\boldsymbol{t}_n)_\#\mu$, and let $\nu_n \in \mathscr{P}_p(X)$ be converging to ν as $n \to \infty$. Suppose that $\boldsymbol{r}_n \in L^p(\mu_n;X)$ is an optimal transport map from μ_n to ν_n. Then*

$$\lim_{n\to\infty} \|\boldsymbol{r}_n \circ \boldsymbol{t}_n - \boldsymbol{r}\|_{L^p(\mu,X)} = 0. \tag{8.5.2}$$

Proof. Let $\varphi : X \times X \to \mathbb{R}$ any continuous function with p-growth. Since $W_p^p(\mu_n,\mu) \to 0$, $W_p^p(\nu_n,\nu) \to 0$ as $n \to \infty$, by applying Proposition 7.1.3 and Lemma 5.1.7 we get

$$\begin{aligned}\lim_{n\to\infty}\int_X \varphi(\boldsymbol{t}_n(x),\boldsymbol{r}_n(\boldsymbol{t}_n(x)))\,d\mu(x) &= \lim_{n\to\infty}\int_X \varphi(y,\boldsymbol{r}_n(y))\,d\mu_n(y)\\ &= \int_X \varphi(y,\boldsymbol{r}(y))\,d\mu(y).\end{aligned} \tag{8.5.3}$$

Choosing $\varphi(x_1,x_2) := |x_2|^p$ we get that $\boldsymbol{r}_n \circ \boldsymbol{t}_n$ is bounded in $L^p(\mu;X)$ and its norm converges to the norm of $\boldsymbol{r}$; therefore we can assume that $\boldsymbol{r}_n \circ \boldsymbol{t}_n$ is weakly convergent to some map $\boldsymbol{s} \in L^p(\mu;X)$ and we should prove that $\boldsymbol{s} = \boldsymbol{r}$. Thus we choose $\varphi(x_1,x_2) := \zeta(x_1)\langle x_2, z\rangle$ with ζ continuous and bounded and $z \in X$: (8.5.3) yields

$$\lim_{n\to\infty}\int_X \zeta(\boldsymbol{t}_n(x))\langle z,\boldsymbol{r}_n(\boldsymbol{t}_n(x))\rangle\,d\mu(x) = \int_X \zeta(x)\langle z,\boldsymbol{r}(x)\rangle\,d\mu(x),$$

whereas weak convergence provides

$$\begin{aligned}\lim_{n\to\infty}\int_X \zeta(\boldsymbol{t}_n(x))\langle z,\boldsymbol{r}_n(\boldsymbol{t}_n(x))\rangle\,d\mu(x) &= \lim_{n\to\infty}\int_X \zeta(x)\langle z,\boldsymbol{r}_n(\boldsymbol{t}_n(x))\rangle\,d\mu(x)\\ &= \int_X \zeta(x)\langle z,\boldsymbol{s}(x)\rangle\,d\mu(x).\end{aligned}$$

It follows that $\langle z,\boldsymbol{s}(x)\rangle = \langle z,\boldsymbol{r}(x)\rangle$ for μ-a.e. $x \in X$, $\forall\, z \in X$, and therefore being X separable $\boldsymbol{s} = \boldsymbol{r}$ μ-a.e. in X. □

Proposition 8.5.4. *Let μ, $\nu \in \mathscr{P}_p(X)$ and let $\boldsymbol{\gamma} \in \Gamma_o(\mu,\nu)$. For every divergence-free vector field $w \in L^p(\mu;X)$ we have*

$$\int_{X\times X} \langle j_p(x_2-x_1), w(x_1)\rangle\,d\boldsymbol{\gamma}(x_1,x_2) = 0. \tag{8.5.4}$$

In particular, if $\boldsymbol{r}$ *is an optimal transport map between* μ *and* $\nu = \boldsymbol{r}_{\#}\mu$ *we have*

$$\int_X \langle j_p(\boldsymbol{r}(x) - x), w(x)\rangle \, d\mu(x) = 0 \quad \forall w \in L^p(\mu; X) \ s.t. \ \nabla \cdot \big(w\mu\big) = 0. \tag{8.5.5}$$

Recalling (8.4.3) *we get that* $\boldsymbol{r} - \boldsymbol{i} \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$.

Proof. We can assume (possibly replacing $\boldsymbol{\gamma}$ by $(\pi_t^{1,1\to 2})_{\#}\boldsymbol{\gamma}$ with t close to 1) that $\boldsymbol{\gamma}$ is the unique optimal transport plan between μ and ν (see Lemma 7.2.1).

By the approximation result stated in Proposition 8.3.3 we can find finite dimensional subspaces X_h, measures $\mu_h \in \mathscr{P}_p(X)$ with support in X_h and regular restriction to X_h converging to μ in $\mathscr{P}_p(X)$, and vectors $w_h \in L^p(\mu_h; X_h)$ such that $\nabla \cdot (w_h\mu_h) = 0$, $(\boldsymbol{i} \times w_h)_{\#}\mu_h \to (\boldsymbol{i} \times w)_{\#}\mu$ in $\mathscr{P}_p(X^2)$. Denoting by $\boldsymbol{t}_h$ the unique optimal transport map between μ_h and $\nu_h := \hat{\pi}^h_{\#}\nu$ (as usual, $\hat{\pi}^h$ is the orthogonal projection of X onto X_h and we identify μ_h and ν_h with their restriction to X_h), we know by Proposition 8.5.2 that $\boldsymbol{t}_h - \boldsymbol{i} \in \mathrm{Tan}^r_{\mu_h}\mathscr{P}_p(X_h)$, and therefore

$$\int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), w_h\rangle \, d\mu_h = 0 \qquad \forall h \in \mathbb{N}.$$

Moreover, the uniqueness of $\boldsymbol{\gamma}$ yields that the transport plans $(\boldsymbol{i} \times \boldsymbol{t}_h)_{\#}\mu_h$ narrowly converge in $\mathscr{P}(X \times X)$ to $\boldsymbol{\gamma}$. Since the marginals of the plans converge in $\mathscr{P}_p(X)$ we have also that the plans are uniformly p-integrable, therefore

$$\begin{aligned} \lim_{h\to\infty} \int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), \tilde{w}\rangle \, d\mu_h &= \lim_{h\to\infty} \int_{X\times X} \langle j_p(x_2 - x_1), \tilde{w}(x_1)\rangle \, d(\boldsymbol{i} \times \boldsymbol{t}_h)_{\#}\mu_h \\ &= \int_{X\times X} \langle j_p(x_2 - x_1), \tilde{w}(x_1)\rangle \, d\boldsymbol{\gamma} \end{aligned}$$

for any continuous function $\tilde{w}$ with linear growth. By Proposition 8.3.3 again (with $f(x_1, x_2) = |x_2 - \tilde{w}(x_1)|^p$) we know that

$$\lim_{\tilde{w}\in C^0_b(X),\ \tilde{w}\to w \text{ in } L^p(\mu;X)} \ \limsup_{h\to\infty} \int_X |w_h - \tilde{w}|^p \, d\mu_h = 0. \tag{8.5.6}$$

Since

$$0 = \int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), \tilde{w}\rangle \, d\mu + \int_X \langle j_p(\boldsymbol{t}_h - \boldsymbol{i}), w_h - \tilde{w}\rangle \, d\mu_h \qquad \text{for any } \tilde{w} \in C^0_b(X),$$

passing to the limit as $h \to \infty$ and using Hölder inequality we obtain

$$\left| \int_X \langle j_p(x_2 - x_1), \tilde{w}(x_1)\rangle \, d\boldsymbol{\gamma} \right| \le \sup_h \|\boldsymbol{t}_h - \boldsymbol{i}\|^{1/q}_{L^p(\mu_h;X)} \limsup_{h\to\infty} \|w_h - \tilde{w}\|_{L^p(\mu_h;X)}.$$

Taking into account (8.5.6) we conclude that $\int_X \langle j_p(x_2 - x_1), w(x_1)\rangle \, d\boldsymbol{\gamma} = 0$. □

The above proposition shows that for general measures $\mu \in \mathscr{P}_p(X)$

$$\mathrm{Tan}^r_\mu \mathscr{P}_p(X) \subset \mathrm{Tan}_\mu \mathscr{P}_p(X). \tag{8.5.7}$$

Now we want to prove the opposite inclusion: let us first mention that the case $p = 2$ is particularly simple.

Corollary 8.5.5. *For any $\mu \in \mathscr{P}_2(X)$ we have* $\mathrm{Tan}_\mu \mathscr{P}_2(X) = \mathrm{Tan}^r_\mu \mathscr{P}_2(X)$.

Proof. We should only check the inclusion $\subset$: if $\varphi \in \mathrm{Cyl}(X)$ it is always possible to choose $\lambda > 0$ such that $x \mapsto \frac{1}{2}|x|^2 + \lambda^{-1}\phi(x)$ is convex. Therefore $\boldsymbol{r} := \boldsymbol{i} + \lambda^{-1}\nabla\varphi$ is cyclically monotone, thus an optimal map between μ and $\boldsymbol{r}_\#\mu$; by (8.5.1) we obtain that $\nabla\phi = \lambda(\boldsymbol{r} - \boldsymbol{i})$ belongs to $\mathrm{Tan}^r_\mu \mathscr{P}_2(X)$. □

In the general case $p \in (1, +\infty)$ the desired inclusion follows by the following characterization:

Proposition 8.5.6. *Let $\mu \in \mathscr{P}_p(X)$, $v \in L^p(\mu; X)$, and $\mu_\varepsilon := (\boldsymbol{i} + \varepsilon v)_\#\mu$ for $\varepsilon > 0$. If $v \in \mathrm{Tan}_\mu \mathscr{P}_p(X)$ then*

$$\lim_{\varepsilon \downarrow 0} \frac{W_p(\mu, \mu_\varepsilon)}{\varepsilon} = \|v\|_{L^p(\mu;X)}, \tag{8.5.8}$$

and denoting by $\boldsymbol{\gamma}_\varepsilon \in \Gamma_o(\mu, \mu_\varepsilon)$ a family of optimal plans, we have

$$\lim_{\varepsilon \downarrow 0} \int_{X \times X} \left| \frac{x_2 - x_1 - \varepsilon v(x_1)}{\varepsilon} \right|^p d\boldsymbol{\gamma}_\varepsilon(x_1, x_2) = 0. \tag{8.5.9}$$

Proof. Let us consider the rescaled plans

$$\boldsymbol{\mu}_\varepsilon := \big(\pi^1, \varepsilon^{-1}(\pi^2 - \pi^1)\big)_\# \boldsymbol{\gamma}_\varepsilon \quad \text{for } \boldsymbol{\gamma}_\varepsilon \in \Gamma_o(\mu, \mu_\varepsilon), \tag{8.5.10}$$

observing that

$$\pi^1_\# \boldsymbol{\mu}_\varepsilon = \mu, \qquad \int_{X^2} |x_2|^p \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2) = \frac{W_p^p(\mu, \mu_\varepsilon)}{\varepsilon^p} \le \|v\|^p_{L^p(\mu;X)}, \tag{8.5.11}$$

$$\int_{X \times X} \left| \frac{x_2 - x_1 - \varepsilon v(x_1)}{\varepsilon} \right|^p d\boldsymbol{\gamma}_\varepsilon(x_1, x_2) = \int_{X \times X} |x_2 - v(x_1)|^p \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2).$$

For every vanishing sequence $\varepsilon_k \to 0$ we can find a subsequence (still denoted by ε_k) and a limit plan $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}_{\varepsilon_k}$ is narrowly converging to $\boldsymbol{\mu}$ in $\mathscr{P}(X \times X_\varpi)$. In particular, for every smooth cylindrical function $\zeta \in \mathrm{Cyl}(X)$ we have

$$\begin{aligned}
\varepsilon^{-1} \int_X \Big(\zeta(x + \varepsilon v(x)) - \zeta(x)\Big) \, d\mu(x) &= \varepsilon^{-1}\Big(\int_X \zeta(x_2) \, d\mu_\varepsilon(x_2) - \int_X \zeta(x_1) \, d\mu(x_1)\Big) \\
= \int_{X \times X} \frac{\zeta(x_2) - \zeta(x_1)}{\varepsilon} \, d\boldsymbol{\gamma}_\varepsilon(x_1, x_2) &= \int_{X \times X} \frac{\zeta(x_1 + \varepsilon x_2) - \zeta(x_1)}{\varepsilon} \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2) \\
= \int_0^1 \int_{X \times X} \langle \nabla\zeta(x_1 + \varepsilon t x_2), x_2 \rangle \, d\boldsymbol{\mu}_\varepsilon(x_1, x_2) \, dt &
\end{aligned} \tag{8.5.12}$$

and

$$\varepsilon^{-1}\int_X \Big(\zeta(x+\varepsilon v(x))-\zeta(x)\Big)\,d\mu(x)=\int_0^1\int_X \langle\nabla\zeta(x+t\varepsilon v(x)),v(x)\rangle\,d\mu(x)\,dt. \tag{8.5.13}$$

Choosing $\varepsilon=\varepsilon_k$ in (8.5.12) and in (8.5.13) and passing to the limit as $k\to\infty$, a repeated application of Lebesgue dominated convergence theorem yields

$$\begin{aligned}\int_X \langle\nabla\zeta(x),v(x)\rangle\,d\mu(x) \qquad\qquad (8.5.14)\\ =\lim_{k\to\infty}\int_0^1\int_X \langle\nabla\zeta(x+t\varepsilon_k v(x)),v(x)\rangle\,d\mu(x)\,dt\\ =\lim_{k\to\infty}\int_0^1\int_{X\times X}\langle\nabla\zeta(x_1+t\varepsilon_k x_2),x_2\rangle\,d\boldsymbol{\mu}_{\varepsilon_k}(x_1,x_2)\,dt\\ =\int_{X\times X}\langle\nabla\zeta(x_1),x_2\rangle\,d\boldsymbol{\mu}(x_1,x_2). \qquad\qquad (8.5.15)\end{aligned}$$

It follows that the limit plan $\boldsymbol{\mu}$ satisfies

$$\int_{X\times X}\langle\nabla\zeta(x_1),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}(x_1,x_2)=0\quad\forall\,\zeta\in\mathrm{Cyl}(X), \tag{8.5.16}$$

and the same relation holds if $\nabla\zeta$ is replaced by any element ξ of the "cotangent space" $\mathrm{CoTan}_\mu\mathscr{P}_p(X)$ (i.e. the closure in $L^q(\mu;X)$ of the gradient vector fields) introduced by (8.4.5).

If $v\in\mathrm{Tan}_\mu\mathscr{P}_p(X)$ and $p\geq 2$, by the p-inequality (10.2.4), we can find a suitable vanishing subsequence $\varepsilon_k\to 0$ and a limit plan $\boldsymbol{\mu}$ such that

$$\begin{aligned}0&\leq c_p\limsup_{\varepsilon\to 0}\int_{X\times X}|x_2-v(x_1)|^p\,d\boldsymbol{\mu}_\varepsilon(x_1,x_2)\\ &\leq\limsup_{\varepsilon\to 0}\int_{X\times X}|x_2|^p-|v(x_1)|^p-p\langle j_p(v(x_1)),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}_\varepsilon(x_1,x_2)\\ &=\lim_{k\to\infty}\frac{W_p^p(\mu,\mu_{\varepsilon_k})}{\varepsilon_k^p}-\|v\|^p_{L^p(\mu;X)}-\int_{X\times X}p\langle j_p(v(x_1)),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}_{\varepsilon_k}(x_1,x_2)\\ &\leq-\int_{X\times X}p\langle j_p(v(x_1)),x_2-v(x_1)\rangle\,d\boldsymbol{\mu}(x_1,x_2)=0\end{aligned}$$

by (8.5.11) and (8.5.16), since $v\in\mathrm{Tan}_\mu\mathscr{P}_p(X)$ is equivalent to $j_p(v)\in\mathrm{CoTan}_\mu\mathscr{P}_p(X)$. The case $p<2$ is completely analogous. □

When μ is regular, the opposite inclusion

$$\mathrm{Tan}_\mu\mathscr{P}_p(X)\subset\mathrm{Tan}^r_\mu\mathscr{P}_p(X),$$

which completes the proof of Theorem 8.5.1, follows easily from the previous proposition: keeping the same notation, we know that $\boldsymbol{\gamma}_\varepsilon$ is induced by an optimal transport map $\boldsymbol{r}_\varepsilon$ so that $\varepsilon^{-1}(\boldsymbol{r}_\varepsilon - \boldsymbol{i}) \in \mathrm{Tan}^r_\mu \mathscr{P}_p(X)$ and (8.5.9) yields

$$\lim_{\varepsilon\to 0} \int_X \left| \frac{\boldsymbol{r}_\varepsilon(x) - x}{\varepsilon} - v(x) \right|^p d\mu(x) = 0. \tag{8.5.17}$$

Therefore v belongs to $\mathrm{Tan}^r_\mu \mathscr{P}_p(X)$.

In the general case, by disintegrating $\boldsymbol{\gamma}_\varepsilon$ with respect to the first variable x_1, a measurable selection theorem [39] allows us to select $\boldsymbol{r}_\varepsilon(x_1)$ such that $\boldsymbol{r}_\varepsilon(x_1) \in \operatorname{supp}(\boldsymbol{\gamma}_\varepsilon)_{x_1}$ and

$$\left| \frac{\boldsymbol{r}_\varepsilon(x_1) - x_1}{\varepsilon} - v(x_1) \right|^p \le 2 \int_X \left| \frac{\boldsymbol{r}_\varepsilon(y) - y}{\varepsilon} - v(y) \right|^p d(\boldsymbol{\gamma}_\varepsilon)_{x_1}(y).$$

Then, since the graph of $\boldsymbol{r}_\varepsilon$ is contained in the support of $\boldsymbol{\gamma}_\varepsilon$, we obtain that $\boldsymbol{r}_\varepsilon$ is $|\cdot|^p$-monotone (so that $\varepsilon^{-1}(\boldsymbol{r}_\varepsilon - \boldsymbol{i}) \in \mathrm{Tan}^r_\mu \mathscr{P}_p(X)$) and (8.5.17) still holds.