NESTED SUPERPOSITION PRINCIPLE FOR RANDOM MEASURES AND THE GEOMETRY OF THE WASSERSTEIN ON WASSERSTEIN SPACE

ALESSANDRO PINZI AND GIUSEPPE SAVARÉ

ABSTRACT. We study the geometric structure of the space of random measures $\mathcal{P}_p(\mathcal{P}_p(X))$, endowed with the Wasserstein on Wasserstein metric, where (X, d) is a complete separable metric space. In this setting, we prove a metric superposition principle, in the spirit of [Lis07], that will allow us to recover important geometric features of the space.
When X is $\mathbb{R}^d$, we study the differential structure of $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ in analogy with the simpler Wasserstein space $\mathcal{P}_p(\mathbb{R}^d)$. We show that continuity equations for random measures involving the abstract concept of derivation acting on cylinder functions can be more conveniently described by suitable non-local vector fields $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$. In this way, we can

- characterize the absolutely continuous curves on the Wasserstein on Wasserstein space;
- define and characterize its tangent bundle;
- prove a superposition principle for the solutions to the standard non-local continuity equation in terms of solutions of interacting particle systems.

Keywords: random measures, Wasserstein space, superposition principle, continuity equation.
2020 MSC: 49Q22, 35R06, 60G57, 53C23.

CONTENTS

Date: December 23, 2025.

1. Introduction

The study of measure-valued solutions to the continuity equation in Euclidean spaces,

$$\partial_t \mu_t + \mathrm{div}(v_t \mu_t) = 0, \quad v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d,\ \mu_t \in \mathcal{P}(\mathbb{R}^d), \tag{1.1}$$

has become of central interest in the last decades, in particular for its connections to optimal transport [BB00] and to Wasserstein gradient flows [JKO98], [AGS08]. A crucial role is played by the so-called superposition principle [Amb04] (see also [AGS08], [AC14]), that represents solutions of the continuity equation as time marginals of a probability measure concentrated on the associated characteristics system of ODEs, that is

$$\lambda \in \mathcal{P}(C([0,T], \mathbb{R}^d)) \ \text{ s.t. } \ \lambda\text{-a.e. } \gamma \text{ is in } AC([0,T], \mathbb{R}^d) \text{ and solves } \dot\gamma(t) = v_t(\gamma(t)). \tag{1.2}$$

It was first proved by L. Ambrosio for studying well-posedness of a Lagrangian system for ODE under non-smooth assumptions of the vector field, started by the seminal work [DL89], and then extended from Sobolev to BV in [Amb04]; we also refer the reader to the subsequent works [CD08; BCD21; BCK24].

The superposition principle actually holds in much greater generality, substituting the ground space $\mathbb{R}^d$ with a general (complete, separable) metric space X. A first result in this setting was proved by [Lis07] for absolutely continuous curves of measures in $\mathcal{P}_p(X)$, the space of probability measures on X endowed with the p-Wasserstein metric: such evolution can be represented in terms of measures over p-absolutely continuous curves in X. This result was useful to prove the equivalence of definitions for Sobolev spaces on metric measure space, in particular for the one using test-plans, that were introduced in [AGS13; AGS14]. For a further understanding of the topic, we refer the reader to [AILP24].
The superposition result was then refined by taking into account the non-smooth 'differential structure' of the space: thanks to the concept of derivations, introduced in [Wea00] (see also [Di 14]), we may give meaning to (1.1) and (1.2), and the superposition result still holds. These kinds of results can be found in [AT14], in which they work on a metric measure space $(X, d, \mathbf{m})$, and in [ST17], working on general metric spaces, also comparing it with a Smirnov-type decomposition for normal metric currents, [Smi93; PS12; PS13]. It is worth citing also these kinds of results obtained on Wiener spaces [AF09; Tre15], or in a stochastic setting [Fig08; Tre16], that aimed to prove existence and uniqueness for Lagrangian flows associated to Sobolev and BV coefficients.

One of the main application of these results is to study the geometry of the Wasserstein space. In [AGS08, Chapters 6-8], the authors extensively studied the geometry of the Wasserstein space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$: the (static) optimal transport problem, with generic convex costs; the characterization of absolutely continuous curves and geodesics on the Wasserstein space; the characterization of the OT problem in terms of the continuity equation (the so-called Benamou-Brenier formulation); the characterization of its tangent bundle; etc...

In this paper, one of the main goals is to reproduce this theory for the continuity equation for measures on $\mathcal{P}_p(\mathbb{R}^d)$ and for corresponding evolutions in the *Wasserstein on Wasserstein space* $(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)), \mathcal{W}_p)$. For a generic metric space (X, d), the set $\mathcal{P}_p(\mathcal{P}_p(X))$ is defined as the collection of random measures $M \in \mathcal{P}(\mathcal{P}(X))$ that satisfy

$$\int_{\mathcal{P}(X)} W_p^p(\mu, \delta_{x_0}) dM(\mu) = \int_{\mathcal{P}(X)} \int_X d^p(x, x_0) d\mu(x) dM(\mu) < +\infty,$$

for some (and then all) $x_0 \in X$. Then, the Wasserstein on Wasserstein distance is defined as

$$\mathcal{W}_p^p(M, N) := \inf \left\{ \int_{\mathcal{P}(X)\times\mathcal{P}(X)} W_p^p(\mu, \nu) d\Pi(\mu, \nu) \, : \, \Pi \in \Gamma(M, N) \right\}, \tag{1.3}$$

for all $M, N \in \mathcal{P}_p(\mathcal{P}_p(X))$, where $\Gamma(M, N) \subset \mathcal{P}(\mathcal{P}(X) \times \mathcal{P}(X))$ is the set of couplings between M and N. In [PS25] (see also [EP25; BPS25]), we study the structure of solutions to the static problem (1.3) for $p = 2$ and X Hilbert space and we show its link with the theory of *totally convex functionals*, that allows (surprisingly) to recover many features of convex analysis in this infinite dimensional and non-linear setting.

In the present paper, we focus on the *dynamic aspects* of the Wasserstein on Wasserstein space:

(1) its absolutely continuous curves and geodesics (Section 3), under the solely general assumption that (X, d) is a complete and separable metric space;
(2) its 'differential structure', i.e. the description of absolutely continuous curves in terms of an abstract continuity equation for random measures, Sections 4 and 5, when $X = \mathbb{R}^d$;
(3) the description of its tangent bundle $\operatorname{Tan} \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$, Section 4.3, again when $X = \mathbb{R}^d$.

One of the main tools we develop to pursue these objectives is a *nested superposition principle*: it comes both in the metric form (in the spirit of [Lis07]) and its differential form (in the spirit of [AGS08, Theorem 8.2.1]). The term *nested*, comes from the particular structure of the space considered. Indeed, we will prove that starting from an absolutely continuous curve of random measures $(M_t)_{t\in[0,T]} \in C([0, T], \mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ (possibly arising as a solution of a continuity equation), we can lift it to two "dynamic" measures: the first one, $\Lambda \in \mathcal{P}(C([0, T], \mathcal{P}(\mathbb{R}^d)))$, is a measure on curves of probability measures on $\mathcal{P}(\mathbb{R}^d)$, the second one, $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C([0, T], \mathbb{R}^d)))$, is a law on random curves in $\mathbb{R}^d$.

We describe such a nested superposition principle in both the metric and the differential setting, in the symplified case of of the evolution of a particle system.

Interacting N-particle systems. Consider an interacting system of $N \in \mathbb{N}$ particles in $X := \mathbb{R}^d$, described by a vector $\boldsymbol{x}(t) = (x_1(t), \cdots, x_N(t)) \in X^N$; the velocity of each particle $x_i(t)$ is expressed by $v(t, x_i(t), \boldsymbol{x}(t))$, where $v : [0, T] \times X \times X^N \to X$ is a smooth (for the easy of presentation) vector field symmetric in the last N components, i.e. denoting by S_N the usual symmetric group of permutations of $\{1, \ldots, N\}$, it holds

$$v(t, x, \boldsymbol{x}) = v(t, x, \sigma(\boldsymbol{x})) \quad \forall \sigma \in S_N, \quad \sigma(x_1, \ldots, x_N) := (x_{\sigma(1)}, \ldots, x_{\sigma(N)}).$$

Given any initial position $\boldsymbol{x}_0 = (x_{1,0}, \dots, x_{N,0}) \in X^N$ of the N particles, there exists a unique solution $\boldsymbol{x} : [0,T] \to X^N$ whose components solve

$$\begin{cases} \dot{x}_i(t) = v(t, x_i(t), \boldsymbol{x}(t)) \\ x_i(0) = x_{i,0} \end{cases} \tag{1.4}$$

The evolution can also be described either by time-dependent flow map $\boldsymbol{X} : [0,T] \times X^N \to X^N$ such that $X(t, \boldsymbol{x}_0) = \boldsymbol{x}(t) = (x_1(t), \dots, x_N(t))$, or by the point-to-curve evolution map $\Gamma : X^N \to C([0,T], X^N)$ such that $\Gamma(\boldsymbol{x}_0) = \boldsymbol{x}(\cdot) = \boldsymbol{X}(\cdot, \boldsymbol{x}_0)$. Both these maps are invariant with respect permutations of the initial distribution of particles.

Suppose that we describe a distribution on the initial configurations of particles by assigning a symmetric probability measure $m \in \mathcal{P}(X^N)$, thus satisfying $\sigma_\# m = m$ for all permutations $\sigma \in S_N$. The above maps $\boldsymbol{X}, \Gamma$, can be used to describe the evolution of m driven by the system (1.4):

(1) for all $t \in [0,T]$, $m_t := (X(t, \cdot))_\# m \in \mathcal{P}(X^N)$ is a curve of probability measures that solves the continuity equation

$$\partial_t m_t + \mathrm{div}_{\boldsymbol{x}}(\boldsymbol{v}_t m_t) = 0, \tag{1.5}$$

where $\boldsymbol{v} : [0,T] \times X^N \to X^N$ whose components are $\boldsymbol{v}_i(t, \boldsymbol{x}) = v(t, x_i, \boldsymbol{x})$;

(2) $\boldsymbol{\eta} := \Gamma_\# m \in \mathcal{P}(C([0,T], X^N))$ is a probability measure over curves of X^N, concentrated over curves $\boldsymbol{x}(\cdot)$ that solve the system (1.4);

(3) the canonical isomorphism between $C([0,T], X^N)$ and $C([0,T], X)^N$ allows us to define a probability measure $\boldsymbol{\theta} \in \mathcal{P}(C([0,T], X)^N)$.

Thanks to the invariance with respect to permuations, it is natural to consider the projected evolution in the quotient space X^N / S_N, that we can identify with the space of uniform discrete measures of the form $\frac{1}{N} \sum_{i=1}^N \delta_{x_i}$. To this aim, for every space $\mathcal{X}$, we introduce the function

$$\mathcal{J} : \mathcal{X}^N \to \mathcal{P}(\mathcal{X}) \quad \mathcal{J}(x_1, \dots, x_N) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \tag{1.6}$$

and we transform the measures $(m_t), \boldsymbol{\eta}, \boldsymbol{\theta}$ under the action of the corresponding versions of $\mathcal{J}$:

(1) starting from the curve (m_t) of measures in X^N we obtain the curve of probability measures over probability measures (that we call (law of) *random measures*)

$$M_t := \mathcal{J}_\# m_t \in \mathcal{P}(\mathcal{P}(X)), \quad t \in [0,T]; \tag{1.7}$$

(2) starting from $\boldsymbol{\eta}$ and using the map $\mathcal{J}' : C([0,T]; X^N) \to C([0,T]; \mathcal{P}(X))$ defined as

$$\mathcal{J}'[\boldsymbol{x}](t) := \mathcal{J}(\boldsymbol{x}(t)) \quad \text{for every } \boldsymbol{x} \in C([0,T]; X^N),$$

we get $\Lambda := \mathcal{J}'_\# \boldsymbol{\eta} \in \mathcal{P}(C([0,T], \mathcal{P}(\mathbb{R}^d)))$;

(3) starting from $\boldsymbol{\theta}$ and using the map $\mathcal{J}$ in the space $\mathcal{X} := C([0,T], X)$ we obtain $\mathfrak{L} := \mathcal{J}_\# \boldsymbol{\theta} \in \mathcal{P}(\mathcal{P}(C([0,T], \mathbb{R}^d)))$.

Thanks to the symmetry assumption, the original structure that we had between the measures $(m_t)_{t \in [0,T]}$, $\boldsymbol{\eta}$, and $\boldsymbol{\theta}$ is maintained, and no information have been lost about the evolution. Moreover, interpreting $v(t, x, \boldsymbol{x})$ as the restriction to discrete probability measures of a nonlocal vector field $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ via the formula

$$v(t, x, \boldsymbol{x}) = b(t, x, \mathcal{J}(\boldsymbol{x})), \tag{1.8}$$

we can provide intrinsic differential characterizations of $M, \Lambda, \mathfrak{L}$ as follows:

(1) the curve of random measures $(M_t)_{t\in[0,T]}$ solves an abstract continuity equation (see Example 4.15)

$$\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0, \tag{1.9}$$

in duality with smooth cylinder functions (see (1.15) and Section 4). The latter are functions $F : \mathcal{P}(X) \to \mathbb{R}$ of the form $F(\mu) = \Psi(\int_X \phi_1 d\mu, \dots, \int_X \phi_k d\mu)$ and admit a Wasserstein gradient defined as

$$\nabla_W F(x,\mu) := \sum_{j=1}^{k} \partial_j \Psi \left(\int_X \phi_1 d\mu, \dots, \int_X \phi_k d\mu \right) \nabla \phi_j(x) \quad \forall (x,\mu) \in X \times \mathcal{P}(X).$$

(2) the probability measure $\Lambda \in \mathcal{P}(C([0,T], \mathcal{P}(X)))$ is concentrated over curves $(\mu_t)_{t\in[0,T]}$ solutions of the non-local continuity equation on X given by

$$\partial_t \mu_t + \mathrm{div}_X(b_t(\cdot, \mu_t)\mu_t) = 0. \tag{1.10}$$

We can think that each μ_t is an empirical measure associated to a group of N particles following the flow given by b.

(3) the probability measure $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C([0,T], X)))$ is concentrated over dynamic measures $\lambda \in \mathcal{P}(C([0,T], X))$ that, in turn, are concentrated over solutions of

$$\dot\gamma(t) = b(t, \gamma(t), (e_t)_\#\lambda),$$

where $(e_t)_\#\lambda$ is the marginal at time t of λ. This means that λ is the collective distribution of the trajectories of single particles evolving according to the system (1.4).

Nested superposition principles. Our goal is to recover the same structure for general evolution of laws of random measures, also including the general setting of a complete and separable metric space (X, d) for the metric-variational aspects (we use the shorthand $C_T(\mathcal{X})$ for $C([0,T], \mathcal{X})$ for any space $\mathcal{X}$):

- a (absolutely) continuous curve of random measures $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(X)))$ (see §2.2.3);
- a probability measure over (absolutely) continuous curves of measure $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(X)))$;
- a law of random measures over (absolutely) continuous curves $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(X)))$.

There is a natural hierarchy "$\mathfrak{L} \implies \Lambda \implies M_t$" between these objects, in the sense that any measure $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(X)))$ naturally induces a $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(X)))$, which in turn induces a curve $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$. This relation can be described from the standard viewpoint of Bayesian sampling schemes. Recall that, for a general random probability measure $M \in \mathcal{P}(\mathcal{P}(X))$, the associated sampling procedure is: first draw $\mu \in \mathcal{P}(X)$ with law M; then, conditional on μ, draw $x \in X$ with law μ.

In particular, we will apply this scheme to the three objects introduced above. Given $t \in [0,T]$ we have:

$$\mu \sim M_t \qquad\qquad \boldsymbol{\mu} \sim \Lambda \qquad\qquad \lambda \sim \mathfrak{L}$$

and, conditionally,

$$x \mid \mu \sim \mu \qquad\qquad y \mid \boldsymbol{\mu} \sim \mu_t \qquad\qquad \gamma \mid \lambda \sim \lambda$$

where $\mu \in \mathcal{P}(X)$, $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(X))$, $\lambda \in \mathcal{P}(C_T(X))$, $\gamma \in C_T(X)$ and $x, y \in X$. Therefore,

(i) given a sample (λ, γ) from the third model, a sample from the second model is

$$\boldsymbol{\mu} := E\big(\lambda\big), \quad y := e_t(\gamma) = [\gamma](t), \tag{1.11}$$

where $E : \mathcal{P}(C_T(X)) \to C_T(\mathcal{P}(X))$ is defined as $E(\lambda) := \big((e_t)_\# \lambda\big)_{t\in[0,T]}$. In other words, $\Lambda := E_\# \mathfrak{L}$;

(ii) given a sample $(\boldsymbol{\mu}, y)$ from the second model, we recover a sample from the first model as

$$\mu := \mu_t, \quad x := y.$$

In other words, $M_t := (\mathfrak{e}_t)_\# \Lambda$, where $\mathfrak{e}_t(\boldsymbol{\mu}) := \mu_t$;

(iii) Clearly, we could also pass directly from $\mathfrak{L}$ to M_t, by defining

$$\mu := \mathfrak{e}_t \circ E(\lambda), \quad x := \gamma(t).$$

Notice also that $E_t := \mathfrak{e}_t \circ E : \mathcal{P}(C_T(X)) \to \mathcal{P}(X)$ coincides with the push forward of the evaluation map at time t, $(e_t)_\#$. Then, $M_t := (E_t)_\# \mathfrak{L} = ((e_t)_\sharp)_\sharp \mathfrak{L}$, is obtained by a nested push-forward construction.

In Section 3, we show that such a hierarchy can be reversed, preserving relevant regularity and structural properties. More precisely, given a curve of random measure $(M_t)_{t\in[0,T]} \in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(X)))$, $p > 1$, we show that there exist $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(X)))$ and $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(X)))$ consistent with the hierarchical structure above and satisfying the additional properties that are summarized in the following theorem, whose complete proof is a byproduct of Sections 3.2—3.5.

Theorem 1.1 (Nested metric superposition and minimal energy liftings)**.** *Let (X, d) be a complete and separable metric space and $p > 1$. Let $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(X)))$. Then, there exist $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(X)))$ and $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(X)))$ satisfying:*

(1) $(\mathfrak{e}_t)_\# \Lambda = M_t$ for all $t \in [0, T]$, Λ-a.e. $\boldsymbol{\mu}$ is in $AC^p_T(\mathcal{P}_p(X))$, and

$$\int_0^T |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t)dt = \int \int_0^T |\dot{\boldsymbol{\mu}}|^p_{W_p}(t)dt\, d\Lambda(\boldsymbol{\mu}); \tag{1.12}$$

(2) $(E_t)_\# \mathfrak{L} = M_t$ for all $t \in [0, T]$, $\mathfrak{L}$-a.e. $\lambda \in \mathcal{P}(C_T(X))$ is concentrated over $AC^p_T(X)$, and

$$\int_0^T |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t)dt = \int \int \int_0^T |\dot{\boldsymbol{\gamma}}|^p\, dt\, d\lambda(\boldsymbol{\gamma})\, d\mathfrak{L}(\lambda); \tag{1.13}$$

(3) $\Lambda = E_\# \mathfrak{L}$ and there exists a Λ-measurable map $G : C_T(\mathcal{P}(\mathbb{R}^d)) \to \mathcal{P}(C_T(\mathbb{R}^d))$ such that $\mathfrak{L} = G_\# \Lambda$ and $E(G(\boldsymbol{\mu})) = \boldsymbol{\mu}$ for Λ-a.e. $\boldsymbol{\mu}$.

In the particular case of the Euclidean setting, we can further leverage the fine-grained information encoded in the continuity equation (1.9). The corresponding main result when $X = \mathbb{R}^d$ is a byproduct of Sections 4.1 and 5 and can be summarized as follows (we keep the notation for the evaluations maps $e_t, E, E_t, \mathfrak{e}_t$ introduced above).

Theorem 1.2 (Nested superposition principle for the random continuity equation in $\mathbb{R}^d$)**.** *Let $X = \mathbb{R}^d$ and let $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(X)))$ and $b : [0, T] \times X \times \mathcal{P}(X) \to X$ be a Borel non-local vector field satisfying the integrability condition*

$$\int_0^T \int \int |b(t, x, \mu)| d\mu(x) dM_t(\mu) dt < +\infty.$$

Assume that $(\boldsymbol{M}, b)$ satisfy the continuity equation (1.9)

$$\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0 \quad t \in (0, T), \tag{1.14}$$

is satisfied in duality with cylinder functions (see Definition 4.1), in the sense that for any $F \in \mathrm{Cyl}^1_c(\mathcal{P}(X))$ *it holds*

$$\frac{d}{dt}\int_{\mathcal{P}(X)} F(\mu)dM_t(\mu) = \int_{\mathcal{P}(X)}\int_X b_t(x,\mu)\cdot\nabla_W F(x,\mu)d\mu(x)dM_t(\mu) \quad \text{in } \mathscr{D}'(0,T). \tag{1.15}$$

Then, there exists $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(X)))$ *and* $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(X)))$ *such that:*

(1) $(\mathfrak{e}_t)_\#\Lambda = M_t$ *for any* $t \in [0,T]$*, and* Λ*-a.e.* $\boldsymbol{\mu}$ *belongs to* $AC_T(\mathcal{P}(X))$ *and solves*

$$\partial_t\mu_t + \mathrm{div}(b_t(\cdot,\mu_t)\mu_t) = 0 \quad \text{in } \mathscr{D}'((0,T)\times X)$$

(2) $(E_t)_\#\mathfrak{L} = M_t$ *and* $\mathfrak{L}$*-a.e.* $\lambda \in \mathcal{P}(C_T(X))$ *is concentrated over absolutely continuous curves* γ *that are solutions of*

$$\dot\gamma(t) = b(t,\gamma_t,(e_t)_\#\lambda) \quad \text{in } (0,T).$$

(3) $\Lambda = E_\#\mathfrak{L}$.

Links with the Wasserstein on Wasserstein geometry. As a consequence of Theorems 1.1 and 1.2, we obtain important geometric properties of $(\mathcal{P}_p(\mathcal{P}_p(X)),\mathcal{W}_p)$. In particular:

(1) When (X,d) is a complete, separable, and *geodesic* we characterize geodesics in $(\mathcal{P}_p(\mathcal{P}_p(X)),\mathcal{W}_p)$ as superposition of laws of random geodesics of X, see Section 3.6. The geodesics are also related to optimal couplings and optimal random couplings presented in [PS25] (in the Hilbertian case with $p=2$) and in Section 3.1;

(2) Absolutely continuous curves of random measures $(M_t)_{t\in[0,T]}$ in $AC^p_T(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$ can be represented as solutions to the continuity equation (1.14) driven by a unique non-local vector field $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}^d$ satisfying suitable variational and integrability conditions, see Section 4.2.

(3) We can fully justify the definition of the cotangent space of $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ at M as the closure of the Wasserstein gradient of cylinder functions (see (4.36) and Section 4.3):

$$\mathrm{CoTan}_M\,\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)) := \overline{\left\{\nabla_W F \;:\; F\in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d))\right\}}^{L^{p'}(\widetilde{M};\mathbb{R}^d)}, \tag{1.16}$$

where the measure $\widetilde{M} \in \mathcal{P}(\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d))$ is defined as

$$\widetilde{M} := \int_{\mathcal{P}(\mathbb{R}^d)} \mu\otimes\delta_\mu\,dM(\mu), \qquad \widetilde{M}(A\times B) = \int_B \mu(A)\,dM(\mu)$$

for all Borel sets $A\subseteq\mathbb{R}^d$ and $B\subseteq\mathcal{P}(\mathbb{R}^d)$. In fact, its corresponding tangent space in $L^p(\widetilde{M};\mathbb{R}^d)$ obtained by the duality map from $L^{p'}$ to L^p is generated by all the non-local vector fields of minimal velocity, thus representing the infinitesimal behaviour of all the absolutely continuous curves according to the previous point (2). In this way, we reproduce the results in [AGS08, Chapter 8.4] at the level of random measures.

A remarkable corollary of the above results is the *Benamou-Brenier like formula* for the p-Wasserstein distance on random measures (see also [HM25] for a similar setting):

Theorem 1.3 (Benamou-Brenier formula)**.** *Let* $p>1$*. For all* $M_0, M_1 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ *it holds*

$$\mathcal{W}_p^p(M_0,M_1) = \min\Big\{\int_0^1\int_{\mathcal{P}}\int_{\mathbb{R}^d}|b_t(x,\mu)|^p d\mu(x)dM_t(\mu)dt \;:\; M\in AC^p(0,1;\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))), \\ M(0)=M_0,\ M(1)=M_1,\ \partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_tM_t) = 0\Big\}. \tag{1.17}$$

Other results and literature. The Wasserstein on Wasserstein metric has been studied in recent years, mostly to quantify convergence properties for non-parametric statistical problems [Ngu16; CL24]. In the recent paper [BVK25], the authors use the gradient flow theory on the L^2-Wasserstein on Wasserstein space to solve learning tasks (e.g. the multi-classification problem) through the minimization of suitable functions defined over random measures. For this aim, it was crucial to define a notion of tangent space, a continuity equation, and their link with absolutely continuous curves. In this paper, we completely characterize these objects, extending some of their results, even in the case $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ with $p > 1$.

In [AKPR25] the authors prove a metric superposition principle in the spirit of [Lis07] for absolutely continuous curves of stochastic processes with respect to the adapted Wasserstein metric [BBP24]. The strong relation between the iterated Wasserstein space and the one of filtered processes endowed with the adapted Wasserstein metric has been highlighted in [BPS25], where they study a Monge-Brenier theorem for the static iterated optimal transport in the N-iterated 2-Wasserstein space. We plan to further develop the techniques presented in this paper for the study of the geometry of the N-iterated p-Wasserstein space, with possible applications to the space of filtered processes.

In [Pin25a], the first author studies the nested superposition principle adding the requirements that all the random measures are absolutely continuous with respect to suitable reference random measures $Q \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ (see also [Del22; Del24]), with applications to the study of the metric measure space $(\mathcal{P}_p(\mathbb{R}^d), W_p, Q)$. The same technique is then applied to the Wasserstein space over a compact Riemannian manifold, and a version of Theorem 1.2 is proved in this setting as well.

In Theorem 1.2, claim (1) can be seen as a particular case of the main result of [ST17], under the hypotheses

$$\int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |b_t(x,\mu)|^p d\mu(x) dM_t(\mu) dt < +\infty, \quad M_0 \in \mathcal{P}_p(\mathbb{R}^d), \tag{1.18}$$

for some $p > 1$. Indeed, as shown in [Sod23], the local Lipschitz constant in the space $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ of a cylinder function $F : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is given by

$$\left(\int_{\mathbb{R}^d} |\nabla_W F(x,\mu)|^p d\mu(x) \right)^{1/p}$$

Exploiting it and our definition of derivation (see Definition 4.3), one can rewrite our setting only using the metric properties needed to apply the results in [ST17]. Anyway, the differences are in the fact that we do not need the additional integrability assumption (actually, in [Pin25b] it is relaxed to the integrability of $\frac{b_t(x,\mu)}{1+|x|}$), and exploiting the particular structure of the space of probability measures, we can perform the other lifting as well, as in Claims (2) and (3).

A similar result to Theorem 1.2 was already obtained in [LSZ22]. In addition to a non-local vector field, they have also two operators $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^{d\times d}$ and $\gamma : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^{d\times d}$, that describe diffusive terms associated with two independent Brownian motions. The novelties in the present paper are various:

- to directly apply the result of [LSZ22], one should ask the p-integrability assumption for the vector field b, as in (1.18);
- the proof of Claim (1) is proved identifying $\mathcal{P}(\mathbb{R}^d)$ with $\mathbb{R}^\infty$ in both cases. On the other hand, to prove Claim (2), in [LSZ22] the authors reproduce the approximation procedure that is commonly used to prove superposition results. Here, we propose a different approach, based on using as a black-box the finite dimensional superposition

principle to perform a measurable selection, that will give us, as a byproduct, Claim (3). In doing this, we need to prove that the set of curves of measures that are solutions of the continuity equation $\partial_t \mu_t + \operatorname{div}(b_t(\cdot, \mu_t)\mu_t) = 0$ is Borel, and similarly for the set of $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ that are concentrated over solutions of $\dot\gamma(t) = b_t(\gamma(t), (e_t)_\# \gamma)$. These measurability results are the main results in Section 5.2 and may have their own independent interest;
- we use this specific setting as a tool to study the geometry of the Wasserstein on Wasserstein space. In particular, Theorem 1.2 put in relation the purely-metric setting of the Wasserstein on Wasserstein space with the non-local continuity equations of the form $\partial_t \mu_t + \operatorname{div}(b_t(\cdot, \mu_t)\mu_t) = 0$, that have been intensively studied in recent years, e.g. [BF21; BF24; CSS23; CSS25a].

Regarding uniqueness, in Section 6 we assume $p \geq 1$, $M_0 \in \mathcal{P}_p(\mathbb{R}^d)$ and b such that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |b(t, x_0, \mu_0) - b(t, x_1, \mu_1)|^p d\pi(x_0, x_1) \leq L(t) W_p^p(\mu_0, \mu_1), \tag{1.19}$$

for all $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$ and some optimal coupling π between μ_0 and μ_1, with $L \in L^1(0, T)$. Under this Lipschitz assumption, we show uniqueness of $\boldsymbol{M}$, Λ and $\mathfrak{L}$ that start from M_0. As already pointed out in [CD18; LSZ22], if the vector field is Lipschitz in x, uniformly with respect to (t, μ), then well-posedness for the interacting particle system easily follows by standard techniques, from which uniqueness of the previous objects follows. On the other hand, this is less trivial assuming only (1.19): our proof actually shows that such a Lipschitz property is rigid enough to imply that the map $\operatorname{supp}(\mu) \ni x \mapsto b_t(x, \mu)$ is $L(t)$-Lipschitz for any $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, and now we can proceed by proving uniqueness of the interacting particle system by the previously cited techniques.

Plan of the paper. In **Section 2**, we recall the main known ingredients that we need to develop our results. In particular, we fix natural (Polish) topologies over the space of probability measures over a Polish space, the space of continuous curves and all their possible combinations, that will be fixed for the rest of the paper.

In **Section 3**, we prove Theorem 1.1, introducing the method used for the proof in Section 3.1, that shows how the optimal transport problem between random measures $M, N \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ can be seen as a minimum problem either over couplings $\Pi \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d))$ or over random couplings $\mathfrak{P} \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d))$. In Section 3.6, we then apply Theorem 1.1 to study the geodesics of the Wasserstein on Wasserstein space $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$.

In **Section 4**, we introduce the continuity equation on random measures, that follows either the dynamics of a family of derivations defined over cylinder functions or the dynamics led by a non-local vector field. In this setting, we prove Claim (1) in Theorem 1.2, following the strategy developed in [AT14]. This result, together with Theorem 1.1, will allow us to characterize absolutely continuous curves of random measures as solutions of a continuity equation, in Section 4.2. In Section 4.3, we define the tangent and cotangent spaces of $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ at a fixed random measure M, and we give a characterization of it in terms of non-local vector fields of minimal energy solving the continuity equation. Finally, we prove a representation result for derivations as non-local vector field, see Theorem 4.23.

In **Section 5**, we complete the proof of Theorem 1.2, using a similar strategy used for Theorem 1.1. In particular, we first show the measurability of the sets $\mathrm{CE}(b)$ and $\mathrm{SCE}(b)$ (see Definition 5.1) that we exploit to apply a measurable selection argument to define the map G_b.

In **Section 6**, we study the case of a Lipschitz non-local vector field, in the measure variable.

In the appendices, we collect some technical results. In particular: in **Appendix A** we recall the definitions and properties of Lusin and Souslin sets, together with a measurable selection theorem; in **Appendix B** a natural topological-metric structure of $\mathbb{R}^\infty$ is highlighted, and its relation to the space of probability measures as well; in **Appendix C** some results about curves in $\mathcal{P}(\mathbb{R}^\infty)$ and $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ are collected; in **Appendix D** we show some results concerning measurability on the space of probability measures.

Acknowledgements. We wish to thank Anna Korba, Eugenio Regazzini and Luciano Tubaro for stimulating and insightful discussions. We also thank Benoît Bonnet-Weill and Martin Huesmann for their helpful comments on a draft of the present paper.
GS has been supported by the MIUR-PRIN 202244A7YL project Gradient Flows and Non-Smooth Geometric Structures with Applications to Optimization and Machine Learning, by the INDAM project E53C23001740001, and by the Institute for Advanced Study of the Technical University of Munich, funded by the German Excellence Initiative.

2. Preliminaries

Here's a list of the main notations used throughout the paper.

$C_b(X)$, $(C_b(X;\mathbb{R}^n))$	continuous and bounded functions from X to $\mathbb{R}$ (resp. $\mathbb{R}^n$)
$C_c(X)$, $(C_c(X;\mathbb{R}^n))$	cont. functions with compact support from X to $\mathbb{R}$ (resp. $\mathbb{R}^n$)
$C^k_c(\mathbb{R}^d)$, $(C^k_c(\mathbb{R}^d;\mathbb{R}^n))$	k-times differentiable functions in $C_c(\mathbb{R}^d)$ (resp. $C_c(\mathbb{R}^d;\mathbb{R}^n)$)
$L^p(\sigma;\mathbb{R}^d)$,	functions $\mathbb{R}^d$-valued that are p-integrable in a measure space (X,σ)
$\mathcal{M}_+(Y)$	finite positive Borel measures on Polish space Y
$\mathcal{P}(Y)$	Borel probability measures on a Polish space Y
$\mathcal{M}(Y)$	signed Borel measures with finite total variation on Y
$\mathcal{M}(Y;\mathbb{R}^d)$	Borel measures on Y with values in $\mathbb{R}^d$ and finite total variation
$\mathcal{P}_F(Y)$	see Def. 2.13
$\mathcal{L}^1$, $(\mathcal{L}^1_T)$	Lebesgue measure on $\mathbb{R}$ (resp. $[0,T]$)
$C_T(X)$	continuous curves from $[0,T]$ in a topological space X
$AC_T(X)$	absolutely continuous curves from $[0,T]$ to a metric space X
$AC^p_T(X)$	absolutely continuous curves with finite p-energy
D_d	sup distance in $C_T(X)$ w.r.t. the distance d
$a_p, \bar{a}_p$	see Def. 2.3 and (3.1)
$\mathcal{A}_p, \bar{\mathcal{A}}_p$	see Def. 2.8 and (3.3)
$\bar{\mathfrak{A}}_p$	see (3.4)
$\mathcal{P}_p(X)$	prob. measures on a metric space (X,d) with finite p-moment
$W_{p,d}$	p-Wasserstein distance on $\mathcal{P}_p(X)$ built on the distance d
$\hat{W}_{1,d}$	$W_{1,d\wedge 1}$, i.e. 1-Wasserstein distance built on truncated distance
$\mathcal{W}_p$	W_{p,W_p}, distance on random measures (see §2.2.3)
$\hat{\mathcal{W}}_1$	$W_{1,\hat{W}_{1,d}}$, the Wasserstein distance built over $\hat{W}_{1,d}$
e_t	evaluation at time t of a curve
$\mathfrak{e}_t$	the evaluation at time t of a curve of measures $(\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(X))$
E_t	$(e_t)_\sharp$, i.e. push-forward with respect to the map e_t
$\mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$	see Def. 4.1
$\mathrm{Cyl}^1_b(\mathcal{P}(\mathbb{R}^d))$	see Def. 4.1
$\nabla_W F$	see Def. 4.1
$\widetilde{M}$	see Remark 2.6

$M_t \otimes \mathrm{d}t$, $\widetilde{M}_t \otimes \mathrm{d}t$	see §2.3.1
CE(b), SPS(b)	see Def. 5.4

In this section, we introduce the notation about spaces of measures and spaces of curves that we will use in the following. We will reserve the notation (X, d) for a reference complete and separable metric space and we will use the letter Y to denote a generic space which typically arise from suitable, possibly iterated, topological, metric or measure-theoretic constructions starting from (X, d). We take some care to distinguish notions which solely depend on the (Polish) topology τ_Y of Y from concepts that also depend on the choice of a metric d_Y on Y.

2.1. **Spaces of curves.**

2.1.1. *Space of continuous curves.* Let (Y, τ_Y) be a Polish topological space. We will denote with $C_T(Y) := C([0,T], Y)$ the space of continuous curves $\boldsymbol{y} : [0,T] \to Y$, naturally endowed with the *compact-open topology.*

Such a topology can be metrized as well, resulting as Polish (see [Sri08, Theorem 2.4.3] for separability): it is sufficient to choose the usual sup-metric

$$D(\boldsymbol{y}_1, \boldsymbol{y}_2) := \sup_{t\in[0,T]} \delta_Y(\boldsymbol{y}_1(t), \boldsymbol{y}_2(t)) \quad \forall \boldsymbol{y}_1, \boldsymbol{y}_2 \in C_T(Y). \tag{2.1}$$

associated with any metric δ_Y over Y that induces its topology τ_Y. Clearly, by using a bounded metric δ_Y, we may assume that D is bounded as well.

We will denote by $e_t : C_T(Y) \to Y$ the (continuous) evaluation map $e_t(\boldsymbol{y}) := \boldsymbol{y}(t)$.

2.1.2. *Space of absolutely continuous curves in metric spaces.* Let us collect some definitions and results about absolutely continuous curves taking values in the (complete, separable) metric space (X, d_X): note that, in this case, the metric matters, not only its induced topology.

Definition 2.1 (Absolutely continuous curves)**.** *A curve $\boldsymbol{x} : [0,T] \to X$ is said to be absolutely continuous, and we write $\boldsymbol{x} \in AC_T(X)$, if there exists a function $g \in L^1(0,T)$ such that*

$$d_X(\boldsymbol{x}(t), \boldsymbol{x}(s)) \le \int_s^t g(r)dr \quad \textit{whenever} \quad 0 \le s \le t \le T. \tag{2.2}$$

If $g \in L^p(0,T)$, for some $p \in (1, +\infty]$, we say that $\boldsymbol{y} \in AC^p_T(X)$.

The space $AC^p_T(X)$, for $p \in [1, +\infty]$ is a Borel subsets of $C_T(X)$ (see Appendix C).

Proposition 2.2. *Let $\boldsymbol{x} \in AC_T(X)$. Then the limit*

$$\lim_{s\to t} \frac{d_X(\boldsymbol{x}(s), \boldsymbol{x}(t))}{|t-s|} =: |\dot{\boldsymbol{x}}|_{d_X}(t) \tag{2.3}$$

exists for $\mathcal{L}^1$-a.e. $t \in [0,T]$ and it provides the smallest g such that (2.2) is satisfied. We will omit the subscript d_X when it will be clear from the context.

Definition 2.3 (p-action of a curve)**.** *Let $\boldsymbol{x} \in C_T(X)$ and $p \in [1, +\infty)$. The p-action of $\boldsymbol{x}$ is defined as*

$$a_p(\boldsymbol{x}) := \begin{cases} \displaystyle\int_0^T |\dot{\boldsymbol{x}}|^p(t)\,\mathrm{d}t & \textit{if } \boldsymbol{x} \in AC_T(X), \\ +\infty & \textit{otherwise.} \end{cases} \tag{2.4}$$

2.1.3. *Geodesics.* A (minimal, constant speed) geodesic in (X, d_X) is a curve $\boldsymbol{x} \in C([0,1], X)$ that satisfies

$$d(\boldsymbol{x}(t), \boldsymbol{x}(s)) = |t-s| d(\boldsymbol{x}(0), \boldsymbol{x}(1)) \quad \text{for every } s, t \in [0,1]. \tag{2.5}$$

Observe that in (2.5) it is enough to require $\leq$, since any strict inequality somewhere would contradict the triangle inequality.

We denote by $\mathrm{Geo}(X) \subset C([0,1], X)$ the closed (thus Borel) subset of constant speed geodesics.

We say that X is a *geodesic space* if for all $x, y \in X$ there exists $\boldsymbol{x} \in \mathrm{Geo}(X)$ such that $\boldsymbol{x}(0) = x$ and $\boldsymbol{x}(1) = y$. In a geodesic space, using a measurable selection argument (see Theorem A.10), one can always find a Souslin-Borel measurable map

$$\mathrm{geo} : X \times X \to \mathrm{Geo}(X) \tag{2.6}$$

that selects a constant speed geodesic given the starting and ending points. Moreover, if the geodesic property is enforced with uniqueness, i.e. if for all $x, y \in X$ there exists a unique $\boldsymbol{x} \in \mathrm{Geo}(X)$ such that $\boldsymbol{x}(0) = x$ and $\boldsymbol{x}(1) = y$, then the map geo is uniquely defined and Borel measurable (see [Bog07, Lemma 6.7.1]).

Finally, for all $t \in [0,1]$, we denote by $\mathrm{geo}_t : X \times X \to X$ the map $e_t \circ \mathrm{geo}$, that is the evaluation at time t of geo.

2.2. **Spaces of measures.**

2.2.1. *Narrow topology over the spaces of measures.* Let (Y, τ_Y) be a Polish space. We denote by $\mathcal{B}(Y)$ the Borel σ-algebra generated by τ_Y. We denote with $\mathcal{P}(Y)$ the set of Borel probability measures on Y. More generally, we introduce the sets $\mathcal{M}_+(Y)$, $\mathcal{M}(Y) = \mathcal{M}(Y, \mathbb{R})$, and $\mathcal{M}(Y, \mathbb{R}^d)$, that are, respectively, the set of all positive finite measures, the set of all signed measures with finite total variation, and the set of all measures taking values in $\mathbb{R}^d$ with finite total variation. Recall that the total variation $|\nu| \in \mathcal{M}_+(Y)$ of a measure $\nu \in \mathcal{M}(Y; \mathbb{R}^d)$ is defined as

$$|\nu|(A) := \sup \left\{ \sum_{n=1}^{+\infty} |\nu(E_n)| \ : \ \bigcup E_n = A,\ E_i \cap E_j = \emptyset \text{ as } i \neq j \right\}.$$

Note that $\mathcal{P}(Y) \subset \mathcal{M}_+(Y) \subset \mathcal{M}(Y)$. The space $\mathcal{M}(Y; \mathbb{R}^d)$ is endowed with the *narrow topology* τ_N, i.e. the coarsest topology under which the functions let $\mathcal{M}(Y; \mathbb{R}^d) \ni \nu \mapsto \int \phi \cdot \mathrm{d}\nu$ are continuous for all $\phi \in C_b(Y; \mathbb{R}^d)$. $\mathcal{P}(Y)$ and $\mathcal{M}_+(Y)$ are closed subsets of $\mathcal{M}(Y)$.

Recall that, given a measurable function $f : Z_1 \to Z_2$, where $(Z_i, \mathcal{S}_i)$ are general measurable spaces, and a measure $\mu \in \mathcal{M}_+(Z_1)$, we denote with $f_\sharp \mu \in \mathcal{M}_+(Z_2)$ the push-forward measure, defined as

$$f_\sharp \mu(S) := \mu(f^{-1}(S)) \quad \forall S \in \mathcal{S}_2.$$

Under the Polish assumption on the ambient space Y, the narrow topology is completely characterized by the narrow convergence [Bog07, Theorem 8.9.4(ii)]: given $\nu_n, \nu \in \mathcal{M}(Y; \mathbb{R}^d)$, we say that ν_n narrowly converges to ν (we write $\nu_n \to \nu$) if

$$\int_Y \phi \cdot \mathrm{d}\nu_n \to \int_Y \phi \cdot \mathrm{d}\nu \quad \forall \phi \in C_b(Y; \mathbb{R}^d).$$

A nice characterization of compactness in the narrow topology has been given by Prohorov (see e.g. [Bog07, Theorem 8.6.2]). The theorem is stated for measures in $\mathcal{M}(Y; \mathbb{R}^d)$, and it is also true for $\mathcal{P}(Y)$ and $\mathcal{M}_+(Y)$, since they are closed subsets of $\mathcal{M}(Y)$.

Theorem 2.4. *Let $\mathcal{F} \subset \mathcal{M}(Y; \mathbb{R}^d)$. Then the following are equivalent:*

(1) $\mathcal{F}$ is relatively compact in the narrow topology;
(2) $\mathcal{F}$ is equi-bounded in total variation and equi-tight, i.e.

$$\sup_{\nu\in\mathcal{F}}|\nu|(Y)<+\infty \quad \textit{and} \quad \forall\varepsilon>0\ \exists K_\varepsilon\subset Y \textit{ compact s.t. } \sup_{\nu\in\mathcal{F}}|\nu|(Y\setminus K_\varepsilon)<\varepsilon.$$

Hereafter, unless otherwise stated, these spaces will always be equipped with the narrow topology; this, in turn, generates the corresponding Borel σ-algebra.

2.2.2. *Wasserstein metric.* Assume now the reference space is endowed with a (complete and separable) metric, and we refer to it with (X,d_X). Given $p\geq 1$, we introduce the space

$$\mathcal{P}_p(X):=\left\{\sigma\in\mathcal{P}(X)\ :\ \int_X d_X^p(x,\bar{x})\,\mathrm{d}\sigma(x) \text{ for some } \bar{x}\in Y\right\}.$$

The set $\mathcal{P}_p(X)$ is endowed with the p-Wasserstein metric defined as

$$W_{p,d_X}^p(\sigma_1,\sigma_2):=\min\left\{\int_{X\times X} d_X^p(x_1,x_2)\,\mathrm{d}\pi(x_1,x_2)\ :\ \pi\in\Gamma(\sigma_1,\sigma_2)\right\},$$

where $\Gamma(\sigma_1,\sigma_2)$ is the collection of all the transport plans (or couplings) π with marginals σ_1 and σ_2, i.e. all the probability measures $\pi\in\mathcal{P}(X\times X)$ satisfying $\pi(A\times X)=\sigma_1(A)$ and $\pi(X\times B)=\sigma_2(B)$ for all $A,B\in\mathcal{B}(X)$. When the distance d_X is clear from the context, we will simply write W_p. $(\mathcal{P}_p(X),W_p)$ is a complete and separable metric space.

By Kantorovich duality we have:

(2.7)
$$W_p^p(\sigma_1,\sigma_2)=\sup\left\{\int_X\phi\,\mathrm{d}\sigma_1+\int_X\psi\,\mathrm{d}\sigma_2\ :\ \phi,\psi\in C_b(X),\quad \phi(x_1)+\psi(x_2)\leq d^p(x_1,x_2)\right\}.$$

In the particular case $p=1$, the duality formula can be rewritten as

$$W_1(\sigma_1,\sigma_2)=\sup\left\{\int_X\phi\,\mathrm{d}\sigma_1-\int_X\phi\,\mathrm{d}\sigma_2\ :\ \phi\in\mathrm{Lip}_1(X)\right\},\tag{2.8}$$

where $\mathrm{Lip}_1(X)$ is the family of real Lipschitz functions with Lipschitz constant less or equal than 1.

When d_X is bounded, then $\mathcal{P}_p(X)=\mathcal{P}(X)$ and every metric W_{p,d_X} metrizes the narrow topology in $\mathcal{P}(X)$.

Remark 2.5. *So far, given (X,d_X) a complete and separable metric space, we introduced the Polish space $\mathcal{P}(X)$ (for which we only care about the narrow topology τ_N) and the Wasserstein spaces $\mathcal{P}_p(X)$ for any $p\geq 1$, inducing the Wasserstein topology $\tau_p=\tau_{W_p}$. When d_X is unbounded, it is well-known that τ_p is strictly finer than the restriction of τ_N to $\mathcal{P}_p(X)$. However, since $\mathcal{P}_p(X)$ is a Borel subset of $\mathcal{P}(X)$ and thus a Lusin space with respect to the narrow topology, and the p-Wasserstein topology is finer, then thanks to [Sch73, Corollary 2, pp. 101] the induced Borel σ-algebras $\mathcal{B}(\mathcal{P}_p(X))$ coincide.*

2.2.3. *Laws of random probability measures.* Given a Polish space (Y,τ_Y), the main objects of our study will be the so-called *laws of random probability measures*, or just *random measures*, $M\in\mathcal{P}(\mathcal{P}(Y))$. Since $\mathcal{P}(Y)$ is a Polish space, we observe that

- over $\mathcal{P}(\mathcal{P}(Y))$, we consider the narrow topology, induced by the underlying (Polish) narrow topology over $\mathcal{P}(Y)$. If δ_Y is any *bounded* metric inducing τ_Y, the narrow topology on $\mathcal{P}(\mathcal{P}(Y))$ is induced by the bounded Wasserstein metric $\widehat{\mathcal{W}}_1:=W_{1,W_{1,\delta}}$.

- When a (complete, separable) metric d_Y is assigned on the underlying space Y, then we can endow $\mathcal{P}_p(\mathcal{P}_q(Y))$, with $p, q \geq 1$, with the *Wasserstein on Wasserstein* metric $\mathcal{W}_{p,q} := W_{p,W_{q,d_Y}}$. We will only deal with the case $p = q$, and we use the notation $\mathcal{W}_p = \mathcal{W}_{p,p} = W_{p,W_{p,d_Y}}$.

Thanks to Remark 2.5, it is equivalent to consider a random measure $M \in \mathcal{P}(\mathcal{P}(Y))$ concentrated over $\mathcal{P}_p(Y)$ or a random measure $M \in \mathcal{P}(\mathcal{P}_p(Y))$, since the Borel σ-algebras induced on $\mathcal{P}_p(Y)$ by the narrow topology concides with the Borel σ-algebra induced by the L^p-Wasserstein metric. In particular, we can always work with random measures $M \in \mathcal{P}(\mathcal{P}(Y))$, possibly specifying later that it is concentrated over measures with finite p-moments.

It is worth highlighting a structure that will appear often in the following: given two Polish spaces Y, Z and a Borel map $f : Y \to Z$, we can define

$$\begin{aligned} F &:= f_\sharp : \mathcal{P}(Y) \to \mathcal{P}(Z) \\ F_\sharp &= f_{\sharp\sharp} : \mathcal{P}(\mathcal{P}(Y)) \to \mathcal{P}(\mathcal{P}(Z)), \end{aligned} \tag{2.9}$$

that are Borel maps with the topologies we considered (see Proposition D.8). A property of this nested push-forward is the following: for all $M \in \mathcal{P}(\mathcal{P}(Y))$ and $g : Z \to [0, +\infty]$

$$\int_{\mathcal{P}(Z)} \int_Z g(z) \,\mathrm{d}\nu(z) \,\mathrm{d}F_\sharp M(\nu) = \int_{\mathcal{P}(Y)} \int_Y g \circ f(y) \,\mathrm{d}\mu(y) \,\mathrm{d}M(\mu). \tag{2.10}$$

Remark 2.6. *Any random measure $M \in \mathcal{P}(\mathcal{P}(Y))$ can also be identified through the measure $\widetilde{M} := \int \mu \otimes \delta_\mu \,\mathrm{d}M(\mu) \in \mathcal{P}(Y \times \mathcal{P}(Y))$, i.e. the only measure for which, for all Borel $g : Y \times \mathcal{P}(Y) \to [0, +\infty)$ bounded it holds*

$$\int_{Y \times \mathcal{P}(Y)} g(y, \mu) \,\mathrm{d}\widetilde{M}(y, \mu) = \int_{\mathcal{P}(Y)} \int_Y g(y, \mu) \,\mathrm{d}\mu(y) \,\mathrm{d}M(\mu). \tag{2.11}$$

Remark 2.7. *The narrow topology is Polish when restricted either to $\mathcal{P}(Y)$ or $\mathcal{M}_+(Y)$ (see Section 2.2.2 and Lemma D.2). On the other hand, the narrow topology over $\mathcal{M}(Y)$ and $\mathcal{M}(Y; \mathbb{R}^d)$ cannot be metrized. However, the narrow topology is metrizable when restricted to sets of measures with bounded total variation, in analogy with the weak (and weak*) topology on Banach spaces (see [Bré11, Chapter 3]). Thanks to this fact, we obtain that the space $\mathcal{M}(Y; \mathbb{R}^d)$ is a Lusin space (see Appendix A), which will allow us to recover some useful properties about Borel measurability in these spaces.*

2.3. **Curves of measures.**

2.3.1. *Continuous curves.* Here we specialize the notation of 2.1.1 to the Polish spaces $\mathcal{P}(Y)$ or $\mathcal{P}(\mathcal{P}(Y))$, where (Y, τ) is a Polish space.

In both the spaces $C_T(\mathcal{P}(Y))$ and $C_T(\mathcal{P}(\mathcal{P}(Y)))$, the compact-open topology is Polish. Indeed, using the distances described above, the compact-open topology over $C_T(\mathcal{P}(Y))$ is induced by the sup distance $D_{\hat{W}_1}$, where $\hat{W}_1 := W_{1,\hat{d}}$ and $\hat{d}$ is a bounded distance inducing τ (notice that $\hat{W}_1 \neq W_1 \wedge 1$). Similarly, for $C_T(\mathcal{P}(\mathcal{P}(Y)))$, its compact-open topology is induced by the sup distance $D_{\hat{\mathcal{W}}_1}$, where $\hat{\mathcal{W}}_1 := W_{1,\hat{W}_1}$.

These distances are just a possible choice for inducing such topologies, but these specific choices will be useful for our purposes.

Following Remark 2.6, it is important to notice the identification between any curve of random measures $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(X)))$ with the measure $\widetilde{M_t} \otimes \mathrm{d}t \in \mathcal{M}_+([0,T] \times X \times \mathcal{P}(X))$, defined through the integration formula

$$\int f(t,x,\mu)\,\mathrm{d}\big(\widetilde{M}_t\otimes\,\mathrm{d}t\big)(t,x,\mu)=\int_0^T\int_{\mathcal{P}(X)}\int_X f(t,x,\mu)\,\mathrm{d}\mu(x)\,\mathrm{d}M_t(\mu)\,\mathrm{d}t, \tag{2.12}$$

for all $f:[0,T]\times X\times\mathcal{P}(X)\to[0,1]$ Borel measurable. Similarly, we introduce the measure $M_t\otimes\,\mathrm{d}t\in\mathcal{M}_+([0,T]\times\mathcal{P}(X))$ as

$$\int g(t,\mu)\,\mathrm{d}\big(M_t\otimes\,\mathrm{d}t\big)(t,\mu)=\int_0^T\int_{\mathcal{P}(X)} g(t,\mu)\,\mathrm{d}M_t(\mu)\,\mathrm{d}t, \tag{2.13}$$

for all $g:[0,T]\times\mathcal{P}(X)\to[0,1]$ Borel measurable.

2.3.2. *Topology over $\mathcal{P}(C_T(Y))$.* Given (Y,τ) a Polish space, the natural topology over $\mathcal{P}(C_T(Y))$ is the narrow topology with the compact-open as ground topology. With this topology, the space $\mathcal{P}(C_T(Y))$ is Polish as well, and a convenient metric that induces it is given by $W_{1,D_{\hat d}}$, where $\hat d$ is any bounded distance inducing τ (in particular, again we can take $\hat d=d\wedge 1$ where d is any distance on Y inducing τ).

2.3.3. *Absolutely continuous curves.* Let (X,d) be a complete and separable metric space. Consider the metric spaces $(\mathcal{P}_q(X),W_q)$ and $(\mathcal{P}_p(\mathcal{P}_q(X)),\mathcal{W}_{p,q})$, where $p,q\in[1,+\infty)$. To avoid confusion in the following, we restate the definition of the action of a curve in these cases.

Definition 2.8. *Let $p,q\in[1,+\infty)$ and $\boldsymbol{\mu}=(\mu_t)_{t\in[0,T]}\in C_T(\mathcal{P}_q(X))$. The p-action of $\boldsymbol{\mu}$ is defined as*

$$\mathcal{A}_{p,q}(\boldsymbol{\mu}):=\begin{cases}\int_0^T|\dot{\boldsymbol{\mu}}|^p_{W_q}(t)\,\mathrm{d}t & \text{if } \boldsymbol{\mu}\in AC_T(\mathcal{P}_q(X))\\ +\infty & \text{otherwise}\end{cases} \tag{2.14}$$

In the case $p=q$, we simply denote $\mathcal{A}_p=\mathcal{A}_{p,p}$.

Definition 2.9 (Absolutely continuous curves of random measures)**.** *Let $p,q,r\in[1,+\infty)$ and $\boldsymbol{M}=(M_t)_{t\in[0,T]}\in C_T(\mathcal{P}_p(\mathcal{P}_q(X)))$. Its r-action is then defined as*

$$\boldsymbol{A}_{r,p,q}(\boldsymbol{M}):=\begin{cases}\int_0^T|\dot{\boldsymbol{M}}|^r_{\mathcal{W}_{p,q}}(t)\,\mathrm{d}t & \text{if } \boldsymbol{M}\in AC_T(\mathcal{P}_p(\mathcal{P}_q(X)))\\ +\infty & \text{otherwise}\end{cases} \tag{2.15}$$

In the case $p=q=r$, we simply denote $\boldsymbol{A}_p=\boldsymbol{A}_{r,p,q}$.

In this paper, we will always deal with the case $p=q=r$. A fundamental theorem for our analysis is a lifting result due to [Lis07].

Theorem 2.10. *Let (X,d) be a complete and separable metric space and $p\in(1,+\infty)$. Let $\boldsymbol{\mu}=(\mu_t)_{t\in[0,T]}\in AC^p_T(\mathcal{P}_p(X))$. Then, there exists a lifting $\lambda\in\mathcal{P}(C_T(X))$ such that*

(1) $(e_t)_\sharp\lambda=\mu_t$ for any $t\in[0,T]$, where $e_t(\boldsymbol{x})=\boldsymbol{x}(t)$ for any $\boldsymbol{x}\in C_T(X)$;
(2) λ is concentrated over $AC^p_T(X)$ and

$$\int a_p(\boldsymbol{x})\,\mathrm{d}\lambda(\boldsymbol{x})=\mathcal{A}_p(\boldsymbol{\mu})<+\infty. \tag{2.16}$$

On the other hand, for any $p\in[1,+\infty)$, given $\lambda\in\mathcal{P}(C_T(X))$ concentrated over absolutely continuous curves, with $(e_0)_\sharp\lambda\in\mathcal{P}_p(X)$, the curve $\boldsymbol{\mu}:=((e_t)_\sharp\lambda)_{t\in[0,T]}$ belongs to $AC^p_T(\mathcal{P}_p(X))$ and it satisfies

$$|\dot{\boldsymbol{\mu}}|^p_{W_p}(t)\le\int|\dot{\boldsymbol{x}}|^p(t)\,\mathrm{d}\lambda(\boldsymbol{x}) \tag{2.17}$$

for $\mathcal{L}^1$-a.e. $t \in (0,T)$.

Remark 2.11. *Putting together the formulas* (2.16) *and* (2.17)*, we can see that a lifting λ as in the first part of the Theorem, is of minimal energy among all the possible lifting satisfying only the first condition.*

2.3.4. *Continuity equation over $\mathbb{R}^d$.* Given a Borel time-dependent vector field $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, we say that a curve of measure $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d))$ solves $\partial_t \mu_t + \operatorname{div}(v_t\mu_t) = 0$ if $\int_0^T \int |v(t,x)| \,\mathrm{d}\mu_t(x)\,\mathrm{d}t < +\infty$ and for all $\psi \in C_c^1((0,T)\times\mathbb{R}^d)$ it holds

$$\int_0^T \int_{\mathbb{R}^d} \partial_t\psi(t,x) + \nabla\psi(t,x)\cdot v(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t = 0. \tag{2.18}$$

Thanks to [AGS08, Lemma 8.1.2], it is not restrictive to assume that $t \mapsto \mu_t$ is continuous.

If the vector field is smooth enough to have that the ordinary differential equation given by

$$\begin{cases} \dot{\boldsymbol{x}}(t) = v(t, \boldsymbol{x}(t)) \\ \boldsymbol{x}(0) = \bar{x} \end{cases} \tag{2.19}$$

admits a unique solution $[0,T] \ni t \mapsto \boldsymbol{X}_t(\bar{x})$ for any $\bar{x} \in \mathbb{R}^d$, it holds that the unique solution of the continuity equation starting from $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ is

$$\mu_t := (\boldsymbol{X}_t)_\sharp \mu_0. \tag{2.20}$$

A similar scheme is still valid in a non-smooth setting. The so-called *finite dimensional superposition principle*, highlights it in a completely non-smooth setting (see e.g. [AC08]).

Theorem 2.12 (Finite dimensional superposition principle)**.** *Let $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ be a Borel vector field and $\boldsymbol{\mu} \in C_T(\mathcal{P}(\mathbb{R}^d))$ be satisfying* (2.18) *and $\int_0^T \int |v| \,\mathrm{d}\mu_t\,\mathrm{d}t < +\infty$. Then there exists a superposition solution $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ satisfying*

(1) $(e_t)_\sharp\lambda = \mu_t$ for all $t \in [0,T]$;
(2) $\lambda(AC_T(\mathbb{R}^d)) = 1$ and λ-a.e. $\boldsymbol{x} \in AC_T(\mathbb{R}^d)$ solves the integral formulation of (2.19)

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \int_0^t v(s, \boldsymbol{x}(s))ds \quad \forall t \in [0,T].$$

Conversely, given $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ satisfying (2) and

$$\int\int_0^T |v(t,\boldsymbol{x}(t))|\,\mathrm{d}t\,\mathrm{d}\lambda(\boldsymbol{x}) < +\infty,$$

then the curve of measures $\mu_t^\lambda := (e_t)_\sharp\lambda$ solves the continuity equation.

The structure of this result is very similar to the one of the lifting described in Theorem 2.10. The main difference is that the first result cares in selecting a minimal energy lifting, while the second takes in account the leading velocity vector field of the evolution. This finite dimensional superposition principle will play a fundamental role in the proof of Theorem 1.2.

2.4. **Recap.** Given a Polish space (Y,τ), we fixed a Polish topology on the following spaces:

(i) $C_T(Y)$, see 2.1.1;
(ii) $\mathcal{P}(Y)$, $\mathcal{M}_+(Y)$ and $\mathcal{M}(Y,\mathbb{R}^d)$, see 2.2.1;
(iii) $C_T(\mathcal{P}(Y))$ and $C_T(\mathcal{P}(\mathcal{P}(Y)))$, see 2.3.1;
(iv) $\mathcal{P}(C_T(Y))$, see 2.3.2.

Dealing with a specific metric structure (X, d), so that we care about the distance itself and not just its induced topology, we defined some subspaces of the spaces mentioned above. In some cases, they can be seen using a more general notation.

Definition 2.13. *Let Y be a Polish space and $F : Y \to [0, +\infty]$ a Borel function. Define the set*

$$\mathcal{P}_F(Y) := \left\{ \mu \in \mathcal{P}(Y) \ : \ \int_Y F(y)\,\mathrm{d}\mu(y) < +\infty \right\}. \tag{2.21}$$

Notice that, thanks to Lemma D.1, $\mathcal{P}_F(Y)$ is a Borel subset of $\mathcal{P}(Y)$. When F is lower semicontinuous, we can also say that $\mathcal{P}_F(Y)$ is an F_σ set, i.e. it is union of closed sets.

Then, given a metric space (X, d), we have:

(v) for any $p \geq 1$, $\mathcal{P}_p(X) = \mathcal{P}_F(X)$, with $F(x) := d^p(x, \bar{x})$ for some $\bar{x} \in X$. Its natural metric is W_p, see 2.2.2;

(vi) for any $p, q \geq 1$, $\mathcal{P}_p(\mathcal{P}_q(X)) = \mathcal{P}_F(\mathcal{P}(X)) \subset \mathcal{P}(\mathcal{P}(X))$, with $F(\mu) = W_q^p(\mu, \delta_{\bar{x}})$, for some $\bar{x} \in X$. Its natural metric is $\mathcal{W}_{p,q}$, see 2.2.2;

(vii) as in Remark 2.11, an object $\lambda \in \mathcal{P}(C_T(X))$ satisfying

$$\int \int_0^T |\dot{\boldsymbol{x}}|^p(t)\,\mathrm{d}t\,\mathrm{d}\lambda(\boldsymbol{x}) < +\infty$$

can be simply identified by writing $\lambda \in \mathcal{P}_{a_p}(C_T(X))$, where a_p is the finite energy of a continuous curve, see Definition 2.3.

Regarding the sets of absolutely continuous curves, notice that thanks to Lemma C.1 and Remark 2.5, we have:

(viii) $AC_T^p(X) \subset C_T(X)$ is a Borel subset, for $p > 1$;

(ix) $AC_T^p(\mathcal{P}_q(X)) \subset C_T(\mathcal{P}(X))$ is a Borel subset, for $q \geq 1$ and $p > 1$;

(x) $AC_T^r(\mathcal{P}_p(\mathcal{P}_q(X))) \subset C_T(\mathcal{P}(\mathcal{P}(X)))$ is a Borel subset, for $p, q \geq 1$ and $r > 1$.

3. Nested lifting for an absoltuely continuous curve of random measures

Let (X, d) be a complete and separable metric space, $\overline{x} \in X$ and $p \geq 1$. The goal of this section is to study the structure of absolutely continuous curves of random measures valued in $\mathcal{P}_p(\mathcal{P}_p(X))$, i.e. $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in AC_T^p(\mathcal{P}_p(\mathcal{P}_p(X)))$. Referring to the previous section for all the topological and metric notions, let us fix the notation that will be used for the rest of this section:

- a generic element of $\mathcal{P}_p(X)$ will be indicated as μ;
- a generic element of $\mathcal{P}_p(\mathcal{P}_p(X))$ will be indicated as M;
- a generic element of $C_T(X)$ will be indicated as $\boldsymbol{x} = (x_t)_{t\in[0,T]}$;
- a generic element of $AC_T^p(\mathcal{P}_p(X))$ will be indicated as $\boldsymbol{\mu} := (\mu_t)_{t\in[0,T]}$;
- a generic element of $AC_T^p(\mathcal{P}_p(\mathcal{P}_p(X)))$ will be indicated as $\boldsymbol{M} := (M_t)_{t\in[0,T]}$;
- following the notation of Definition 2.13, a generic element of $\mathcal{P}_{\overline{a}_p}\big(C_T(X)\big)$ will be indicated as λ, where

$$\overline{a}_p(\boldsymbol{x}) := d^p(\overline{x}, x_0) + a_p(\boldsymbol{x}) \tag{3.1}$$

for a fixed $\overline{x} \in X$. Note that $\mathcal{P}_{\overline{a}_p}\big(C_T(X)\big) \subset \mathcal{P}_{a_p}(C_T(X))$, because in addition we are asking that the marginal at time $t = 0$ (and then every marginal) is in $\mathcal{P}_p(X)$. We also introduce the notation

$$\lambda \in \mathcal{P}^{\min}_{\overline{a}_p}(C_T(X)) \subset \mathcal{P}_{\overline{a}_p}(C_T(X)) \tag{3.2}$$

for the liftings that, in addition, satisfy the minimality condition (2.16) too, i.e.

$$\mathcal{P}^{\min}_{\overline{a}_p}(C_T(X)) := \left\{ \lambda \in \mathcal{P}_{\overline{a}_p}(C_T(X)) \ : \ \int a_p(\boldsymbol{x}) d\lambda(\boldsymbol{x}) = \int_0^T |\dot{\boldsymbol{\lambda}}|^p(t) dt \right\},$$

where $\boldsymbol{\lambda} = \big((e_t)_\sharp \lambda\big)_{t\in[0,T]}$.

Using again Definition 2.13, we also introduce two sets of probability measures, that are, respectively, Borel subsets of $\mathcal{P}(C_T(\mathcal{P}(X)))$ and $\mathcal{P}(\mathcal{P}(C_T(X)))$:

- $\Lambda \in \mathcal{P}_{\bar{\mathcal{A}}_p}\big(C_T(\mathcal{P}(X))\big)$, where

$$\bar{\mathcal{A}}_p(\boldsymbol{\mu}) := W_p^p(\mu_0, \delta_{\overline{x}}) + \mathcal{A}_p(\boldsymbol{\mu}). \tag{3.3}$$

  In particular, each $\Lambda \in \mathcal{P}_{\bar{\mathcal{A}}_p}\big(C_T(\mathcal{P}(X))\big)$ is concentrated on $AC^p_T(\mathcal{P}_p(X))$ and $(\mathfrak{e}_t)_\sharp \Lambda \in \mathcal{P}_p(\mathcal{P}_p(X))$ for all $t \in [0,T]$;
- $\mathfrak{L} \in \mathcal{P}_{\overline{\mathfrak{A}}_p}\big(\mathcal{P}(C_T(X))\big)$ where

$$\overline{\mathfrak{A}}_p(\lambda) := \int \overline{a}_p d\lambda = \int d^p(\overline{x}, x_0) d\lambda(\boldsymbol{x}) + \int a_p(\boldsymbol{x}) d\lambda(\boldsymbol{x}). \tag{3.4}$$

  Notice that each $\mathfrak{L} \in \mathcal{P}_{\overline{\mathfrak{A}}_p}\big(\mathcal{P}(C_T(X))\big)$ is concentrated on the set $\mathcal{P}_{\overline{a}_p}\big(C_T(X)\big)$.

In this section, we prove Theorem 1.1, that links $\boldsymbol{M} \in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(X)))$ with $\Lambda \in \mathcal{P}_{\overline{\mathcal{A}}_p}(C_T(\mathcal{P}(X)))$ and $\mathfrak{L} \in \mathcal{P}_{\overline{\mathfrak{A}}_p}\big(\mathcal{P}(C_T(X))\big)$. To better understand the strategy of the proof, in the next subsection we expose it in an easier scenario.

3.1. **Couplings.** We discuss briefly how, given $M_0, M_1 \in \mathcal{P}_p(\mathcal{P}_p(X))$, we can associate to them:

- a coupling $\Pi \in \mathcal{P}_p\big(\mathcal{P}_p(X) \times \mathcal{P}_p(X)\big) = \mathcal{P}_F(\mathcal{P}(X) \times \mathcal{P}(X))$, with

$$F(\mu,\nu) = W_p^p(\mu, \delta_{\overline{x}}) + W_p^p(\nu, \delta_{\overline{x}}),$$

  i.e. it is the p-Wasserstein space built over the product metric space $\mathcal{P}_p(X) \times \mathcal{P}_p(X)$. We say that $\Pi \in \Gamma(M_0, M_1)$ if its marginals are M_0 and M_1;
- a random coupling, i.e. a probability measure $\mathfrak{P} \in \mathcal{P}_p\big(\mathcal{P}_p(X \times X)\big) = \mathcal{P}_F\big(\mathcal{P}(X \times X)\big)$, with

$$F(\pi) := \int \Big(d^p(x, \overline{x}) + d^p(y, \overline{x}) \Big) d\pi(x,y).$$

  We say that $\mathfrak{P} \in \mathrm{R}\Gamma(M_0, M_1)$ if $P^1_\sharp \mathfrak{P} = M_0$ and $P^2_\sharp \mathfrak{P} = M_1$, where

$$P^i : \mathcal{P}(X \times X) \to \mathcal{P}(X), \quad P^i(\pi) = p^i_\sharp \pi. \tag{3.5}$$

The plan $\Pi \in \mathcal{P}_p\big(\mathcal{P}_p(X) \times \mathcal{P}_p(X)\big)$ is simply selected as a $\mathcal{W}_p$-optimal plan between M_0 and M_1 , i.e. such that its marginals are M_0 and M_1, and it realizes the distance $\mathcal{W}_p$.

Given such $\Pi \in \mathcal{P}_p\big(\mathcal{P}_p(X) \times \mathcal{P}_p(X)\big)$, we build $\mathfrak{P} \in \mathcal{P}_p\big(\mathcal{P}_p(X \times X)\big)$ by defining a map $Q : \mathcal{P}_p(X) \times \mathcal{P}_p(X) \to \mathcal{P}_p(X \times X)$ that for any pair $\mu, \nu \in \mathcal{P}_p(X)$ gives (in a measurable way) an optimal plan $Q(\mu,\nu) \in \Gamma_0(\mu,\nu)$. To define such Q, let

$$\begin{aligned} P : \mathcal{P}(X \times X) &\to \mathcal{P}(X) \times \mathcal{P}(X) \\ \pi \quad &\mapsto (p^1, p^2)_\sharp \pi. \end{aligned} \tag{3.6}$$

This map is continuous, so it is Borel. Consider the space of optimal couplings

$$\mathcal{P}^{\rm opt}_p(X \times X) := \left\{ \pi \in \mathcal{P}_p(X \times X) \ : \ \int d^p(x,y) d\pi(x,y) = W_p^p(p^1_\sharp \pi, p^2_\sharp \pi) \right\}. \tag{3.7}$$

Notice that the map $P : \mathcal{P}_p^{\rm opt}(X\times X) \to \mathcal{P}_p(X)\times\mathcal{P}_p(X)$ is Borel and surjective, since for any couple $(\mu,\nu)\in\mathcal{P}_p(X)\times\mathcal{P}_p(X)$ there exists an optimal coupling. Then thanks to the measurable selection theorem A.10, there exists a (Souslin-Borel measurable, see Appendix A) right inverse P^{-1}. Then, let $Q := P^{-1}$ and define

$$
\begin{aligned}
Q_\sharp : \mathcal{P}_p\big(\mathcal{P}_p(X)\times\mathcal{P}_p(X)\big) &\to \mathcal{P}_p\big(\mathcal{P}_p(X\times X)\big)\\
\Pi \quad &\mapsto \quad \mathfrak{P} := Q_\sharp\Pi
\end{aligned}
\tag{3.8}
$$

By construction, each $\mathfrak{P}$ obtained in this way is concentrated on the set of optimal couplings $\mathcal{P}_p^{\rm opt}(X\times X)$. Finally, we have that $\mathfrak{P}\in \mathrm{R}\Gamma(M_0,M_1)$. Summing up, we have this result.

Proposition 3.1. *Let $M_0, M_1 \in \mathcal{P}_p(\mathcal{P}_p(X))$. Then, there exist $\Pi\in\Gamma(M_0,M_1)$ and $\mathfrak{P}\in \mathrm{R}\Gamma(M_0,M_1)$ such that*

$$
\mathcal{W}_p^p(M_0,M_1) = \int_{\mathcal{P}(X)\times\mathcal{P}(X)} W_p^p(\mu,\nu)d\Pi(\mu,\nu) = \int_{\mathcal{P}(X\times X)}\int_{X\times X} d^p(x,y)\pi(x,y)d\mathfrak{P}(\pi).
\tag{3.9}
$$

Whenever (3.9) *is satisfied, we say that $\Pi\in\Gamma_0(M_0,M_1)$ and $\mathfrak{P}\in\mathrm{R}\Gamma_0(M_0,M_1)$.*

This result shows the strategy we will follow to prove our nested superposition principle (both the metric and the differential one) and will play an important role for characterizing the geodesics of $(\mathcal{P}_p(\mathcal{P}_p(X)),\mathcal{W}_p)$ for $p>1$ (see §3.6).

3.2. **From** AC($\mathcal{P}$($\mathcal{P}$)) **to** $\mathcal{P}$(AC($\mathcal{P}$))**.** A similar strategy can be applied to the case of an absolutely continuous curve $\boldsymbol{M}\in AC_T^p(\mathcal{P}_p(\mathcal{P}_p(X)))$.

Notice that we can always associate to $\boldsymbol{M}$ a measure $\Lambda\in\mathcal{P}_{\bar{\mathcal{A}}_p}(C_T(\mathcal{P}(X)))$, using Theorem 2.10 with $(Y,d)=(\mathcal{P}_p(X),W_p)$, here is the specific statement.

Proposition 3.2. *For any curve $\boldsymbol{M}=(M_t)_{t\in[0,T]}\in AC^p\big(\mathcal{P}_p(\mathcal{P}_p(X))\big)$ there exists a lifting $\Lambda\in\mathcal{P}_{\bar{\mathcal{A}}_p}\big(C_T(\mathcal{P}_p(X))\big)$ such that*

$$
(\mathfrak{e}_t)_\sharp\Lambda = M_t \quad \textit{and} \quad \int \mathcal{A}_p(\boldsymbol{\mu})d\Lambda(\boldsymbol{\mu}) = \int_0^T |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t)dt < +\infty,
\tag{3.10}
$$

where $\mathfrak{e}_t(\boldsymbol{\mu})=\mu_t$.

As in the general case of Theorem 2.10, the measure Λ is possibly non-unique. Any possible selection Λ will be indicated as Lift($\boldsymbol{M}$), i.e. $\mathrm{Lift}(\boldsymbol{M}) := \{\Lambda\in\mathcal{P}_{\bar{\mathcal{A}}_p}\big(C_T(\mathcal{P}_p(X))\big) : (3.10) \text{ holds}\}$. The Proposition 3.2 can be restated as: if $\boldsymbol{M}\in AC_T^p(\mathcal{P}_p(\mathcal{P}_p(X)))$, then $\mathrm{Lift}(\boldsymbol{M})\neq\emptyset$.

3.3. **From** $\mathcal{P}$(AC($\mathcal{P}$)) **to** $\mathcal{P}$($\mathcal{P}$(AC))**.** In this subsection, we want to define a map that associates an element $\mathfrak{L}\in\mathcal{P}_{\bar{\mathfrak{A}}_p}(\mathcal{P}(C_T(X)))$ to $\Lambda\in\mathcal{P}_{\bar{\mathcal{A}}_p}(C_T(\mathcal{P}(X)))$. First, we need to define the map

$$
\begin{aligned}
E : \mathcal{P}\big(C_T(X)\big) &\to C_T(\mathcal{P}(X))\\
\lambda \quad &\mapsto \big((e_t)_\sharp\lambda\big)_{t\in[0,T]}
\end{aligned}
\tag{3.11}
$$

Lemma 3.3. *For all $\lambda\in\mathcal{P}(C_T(X))$, it holds $E[\lambda]\in C_T(\mathcal{P}(X))$ and the map E is continuous. Moreover, E is surjective from $\mathcal{P}_{\bar{a}_p}^{\min}(C_T(X))$ (see* (3.2)*) to $AC_T^p(\mathcal{P}_p(X))$, i.e.*

$$
E\big(\mathcal{P}_{\bar{a}_p}(C_T(X))\big) = E\big(\mathcal{P}_{\bar{a}_p}^{\min}(C_T(X))\big) = AC_T^p(\mathcal{P}_p(X)).
\tag{3.12}
$$

Proof. Consider the distances $D_{\hat W_1}$ and $W_{1,D_{\hat d}}$, respectively, on $C_T(\mathcal{P}(X))$ and $\mathcal{P}(C_T(X))$ to induce their topologies (see §2.4), where $\hat d := d \wedge 1$ and $\hat W_1 := W_{1,\hat d}$. We prove that the map E is 1-Lipschitz with these choices.
For all $\lambda \in \mathcal{P}\big(C_T(X)\big)$ and for all sequence $t_n \in [0,T]$ converging to t, it holds

$$W_{1,\hat d}\big((e_t)_\sharp\lambda,(e_{t_n})_\sharp\lambda\big) \le \int \hat d(\gamma_t,\gamma_{t_n})d\lambda(\gamma) \to 0$$

as $n \to +\infty$ by the dominated convergence theorem. Considering $\lambda,\rho \in \mathcal{P}\big(C_T(X)\big)$, we have

$$D_{W_{1,\hat d}}\big(E[\lambda],E[\rho]\big) = \sup_{t\in[0,T]} W_{1,\hat d}\big((e_t)_\sharp\lambda,(e_t)_\sharp\rho\big) \le W_{1,D_{\hat d}}(\lambda,\rho),$$

because $(e_t)_\sharp$ is a contraction, indeed saying that Π is a $W_{1,D_{\hat d}}$-optimal coupling between λ and ρ we have that

$$W_{1,\hat d}\big((e_t)_\sharp\lambda,(e_t)_\sharp\rho\big) \le \int \hat d(x^1_t,x^2_t)d\Pi(\boldsymbol{x}^1,\boldsymbol{x}^2) \le \int D_{\hat d}(\boldsymbol{x}^1,\boldsymbol{x}^2)d\Pi(\boldsymbol{x}^2,\boldsymbol{x}^2) = W_{1,D_{\hat d}}(\lambda,\rho).$$

Regarding the second part of the statement, first of all we need to prove that E is well defined: consider $\lambda \in \mathcal{P}_{\overline{a}_p}(C_T(X))$, then

$$\begin{aligned} W^p_{p,d}((e_t)_\sharp\lambda,(e_s)_\sharp\lambda) &\le \int d^p(x_t,x_s)d\lambda(\boldsymbol{x}) \le \int \left(\int_s^t |\dot{\boldsymbol{x}}|(r)dr\right)^p d\lambda(\boldsymbol{x}) \\ &\le |t-s|^{p-1}\int_s^t\int |\dot{\boldsymbol{x}}|^p(r)d\lambda(\boldsymbol{x})dr. \end{aligned}$$

Moreover, $(e_0)_\sharp\lambda \in \mathcal{P}_p(X)$. Putting everything together, the curve $\big((e_t)_\sharp\lambda\big)_{t\in[0,T]}$ is in $AC^p_T(\mathcal{P}_p(X))$, and by the Lebesgue theorem, for a.e. $t \in [0,T]$ it holds

$$\lim_{h\to 0}\frac{W^p_{p,d}((e_{t+h})_\sharp\lambda,(e_t)_\sharp\lambda)}{|h|^p} \le \int |\dot{\boldsymbol{x}}|^p(t)d\lambda(\boldsymbol{x}) < +\infty,$$

and in particular $\mathcal{A}_p(E[\lambda]) \le \int_0^T a_p(\boldsymbol{x})d\lambda(\boldsymbol{x})$. The surjectivity from $\mathcal{P}^{\min}_{\overline{a}_p}(C_T(X))$ to $AC^p_T(\mathcal{P}_p(X))$ is implied by Theorem 2.10. □

The goal is to find a measurable right inverse of the map E, in particular we would like to apply Theorem A.10. Recall that we know that E defined in (3.11) is not surjective: for example, consider two distinct points $x_0, x_1 \in X$ and define $\mu_t := (T-t)\delta_{x_0} + t\delta_{x_1}$ for $t \in [0,T]$. It is clear that $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(X))$, but it cannot be lifted to any measure $\mathcal{P}(C_T(X))$.

The situation is nicer if we restrict E as a map $E : \mathcal{P}^{\min}_{\overline{a}_p}\big(C_T(X)\big) \to AC^p_T\big(\mathcal{P}_p(X)\big)$, which is surjective thanks to Lemma 3.3.

Theorem 3.4. *There exists a Souslin-Borel measurable (see Appendix A) map* $G : AC^p_T(\mathcal{P}_p(X)) \to \mathcal{P}^{\min}_{\overline{a}_p}(C_T(X))$ *such that* $E \circ G[\boldsymbol{\mu}] = \boldsymbol{\mu}$ *for any* $\boldsymbol{\mu} \in AC^p_T(\mathcal{P}_p(X))$*, i.e.* $G(\boldsymbol{\mu})$ *is a lifting of* $\boldsymbol{\mu}$*. Moreover, for any* $\boldsymbol{\mu} \in AC^p_T(\mathcal{P}_p(X))$ *it holds*

$$\int a_p(\boldsymbol{x})d\big(G[\boldsymbol{\mu}]\big)(\boldsymbol{x}) = \mathcal{A}_p(\boldsymbol{\mu}). \tag{3.13}$$

In particular, the following map is well defined:

$$\begin{aligned} G_\sharp : \mathcal{P}_{\bar{\mathcal{A}}_p}\big(C_T(\mathcal{P}(X))\big) &\to \mathcal{P}_{\bar{\mathfrak{A}}_p}\big(\mathcal{P}(C_T(X))\big) \\ \Lambda \quad &\mapsto \quad \mathfrak{L} := G_\sharp\Lambda. \end{aligned} \tag{3.14}$$

Proof. The subset $\mathcal{P}^{\min}_{\bar{a}_p}(C_T(X))$ is a Borel subset of $\mathcal{P}(C_T(X))$, thanks to Lemma D.1. The same holds for the subset $AC^p_T(\mathcal{P}_p(X)) \subset C_T(\mathcal{P}(X))$. Then, because E is a surjection, we can apply Theorem A.10 to obtain a Souslin-Borel measurable map $G : AC^p_T(\mathcal{P}_p(X)) \to \mathcal{P}^{\min}_{\bar{a}_p}(C_T(X))$ such that $E \circ G[\boldsymbol{\mu}] = \boldsymbol{\mu}$ for all $\boldsymbol{\mu} \in AC^p_T(\mathcal{P}_p(X))$.
Given $\Lambda \in \mathcal{P}_{\bar{\mathcal{A}}_p}(C_T(\mathcal{P}(X)))$, thus concentrated on $AC^p_T(\mathcal{P}_p(X))$, thanks to Proposition A.8 and Corollary A.9, since G is Souslin-Borel measurable, we have that $\mathfrak{L} := G_\sharp\Lambda$ is a probability measure over Borel sets of $\mathcal{P}(C_T(X))$. It remains to show that $\mathfrak{L} \in \mathcal{P}_{\mathfrak{A}_p}(\mathcal{P}(C_T(X)))$. First of all, given $\overline{x} \in X$, it holds

$$\int\int d^p(\overline{x}, x_0)d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) = \int\int_X d^p(x, \overline{x})d\mu_0(x)d\Lambda(\boldsymbol{\mu}) = \int W_p^p(\mu_0, \delta_{\overline{x}})d\Lambda(\boldsymbol{\mu}) < +\infty.$$

Then

$$\begin{aligned}\int\int a_p(\boldsymbol{x})d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) &= \int_0^T \int\int |\dot{\boldsymbol{x}}|^p d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda)dt \\ &= \int_0^T \int\int |\dot{\boldsymbol{x}}|^p_d d\big(G[\boldsymbol{\mu}]\big)(\boldsymbol{x})d\Lambda(\boldsymbol{\mu})dt = \int_0^T \int |\dot{\boldsymbol{\mu}}|^p_{W_p} d\Lambda(\boldsymbol{\mu})dt < +\infty,\end{aligned} \tag{3.15}$$

where the last equality follows from $G[\boldsymbol{\mu}] \in \mathcal{P}^{\min}_{\bar{a}_p}(C_T(X))$ for all $\boldsymbol{\mu} \in AC^p_T(\mathcal{P}_p(X))$. □

Remark 3.5. *Looking closely at the previous proof, the last step highlights why it is important to invert the map E from the domain $\mathcal{P}^{\text{opt}}_{\bar{a}_p}(C_T(X))$, instead of $\mathcal{P}_{\bar{a}_p}(C_T(X))$, otherwise, we could not conclude that the measure $\mathfrak{L}$ belongs to $\mathcal{P}_{\bar{\mathfrak{A}}_p}(\mathcal{P}(C_T(X)))$.*

3.4. **From $\mathcal{P}(\mathcal{P}(\mathrm{AC}))$ to $\mathrm{AC}(\mathcal{P}(\mathcal{P}))$.** We conclude our construction by discussing the natural projection from $\mathfrak{L} \in \mathcal{P}_{\overline{\mathfrak{A}}_p}(\mathcal{P}(C_T(X)))$ to $AC^p_T(\mathcal{P}_p(\mathcal{P}_p(X)))$. We use the nested push-forward described in (2.9) using the evaluation map. For any $t \in [0, T]$, we define the maps

$$\begin{aligned} E_t : \mathcal{P}\big(C_T(X)\big) &\to \mathcal{P}(X) & \mathfrak{E} : \mathcal{P}(\mathcal{P}(C_T(X))) &\to C_T(\mathcal{P}(\mathcal{P}(X))) \\ \lambda &\mapsto (e_t)_\sharp\lambda, & \mathfrak{L} &\mapsto \big((E_t)_\sharp\mathfrak{L}\big)_{t\in[0,T]}. \end{aligned} \tag{3.16}$$

Proposition 3.6. *Let $\mathfrak{L} \in \mathcal{P}_{\mathcal{A}_p}\big(\mathcal{P}(C_T(X))\big)$ and define $M_t := (E_t)_\sharp\mathfrak{L}$ for any $t \in [0, T]$. Then, $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in AC^p_T\big(\mathcal{P}_p(\mathcal{P}_p(X))\big)$ and it holds*

$$|\dot{M}|^p_{\mathcal{W}_p}(t)dt \le \int\int |\dot{\boldsymbol{x}}|^p(t)d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) < +\infty \tag{3.17}$$

for a.e. $t \in (0, T)$. In particular

$$\int_0^T |\dot{M}|^p_{\mathcal{W}_p}(t)dt \le \int\int a_p(\boldsymbol{x})d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) < +\infty. \tag{3.18}$$

Proof. First, notice that $\boldsymbol{M} \in C_T(\mathcal{P}(\mathcal{P}(X)))$, since $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(X)))$. Now, we have to prove that $M_0 \in \mathcal{P}_p(\mathcal{P}_p(X))$ and $\int_0^T |\dot{M}|^p_{\mathcal{W}_p}(t)dt < +\infty$. First of all,

$$\int W_p^p(\mu, \delta_{\overline{x}})dM_0(\mu) = \int W_p^p((e_0)_\sharp\lambda, \delta_{\overline{x}})d\mathfrak{L}(\lambda) \le \int\int d^p(x_0, \overline{x})d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) < +\infty,$$

which implies that $M_0 \in \mathcal{P}_p(\mathcal{P}_p(X))$. Then

$$\mathcal{W}_p^p(M_t, M_s) \le \int_{\mathcal{P}\times\mathcal{P}} W_p^p(\mu, \nu)d\big((E_t, E_s)_\sharp\mathfrak{L}\big)(\mu, \nu) = \int W_p^p\big((e_t)_\sharp\lambda, (e_s)_\sharp\lambda\big)d\mathfrak{L}(\lambda)$$

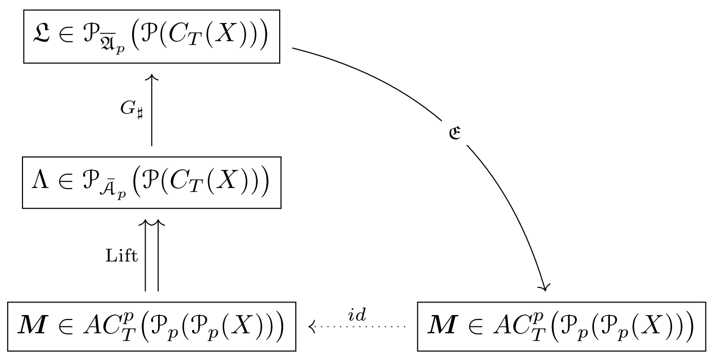
$$\leq \int\int d(x_t,x_s)^p d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) \leq |t-s|^{p-1}\int_s^t \int\int |\dot{\boldsymbol{x}}|_d^p(r)d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda)dr,$$

where we used $d(x_t,x_s) \leq \int_s^t |\dot{\boldsymbol{x}}|_d(r)dr$, Holder inequality and Fubini's theorem. This implies that $(M_t)_{t\in[0,T]}$ is absolutely continuous, and by Lebesgue theorem it holds

$$|\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t) \leq \int\int |\dot{\boldsymbol{x}}|_d^p(t)d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) \quad \text{for a.e. } t\in(0,T). \qquad \square$$

3.5. **Composition of maps.** In this section, we analyze the relations between the (possibly multivalued) map Lift, and the maps $G_\sharp$ and $\mathfrak{E}$.

Proposition 3.7. *The composition $\mathfrak{E}\circ(G_\sharp)\circ\mathrm{Lift}$, represented by the following diagram*



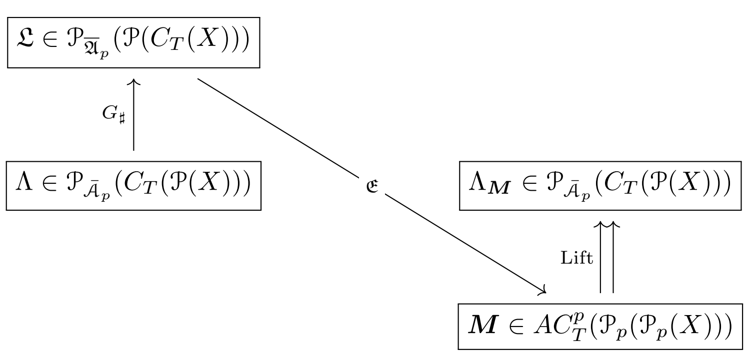
is the identity, i.e. $\mathfrak{E}\circ G_\sharp\circ\mathrm{Lift}(\boldsymbol{M}) = \boldsymbol{M}$ for any $\boldsymbol{M}\in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(X)))$. In other words, given $\boldsymbol{M}\in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(X)))$, for any $\Lambda\in\mathrm{Lift}(\boldsymbol{M})$ and defining $\mathfrak{L}:=G_\sharp\Lambda$, it holds $\mathfrak{E}(\mathfrak{L})=\boldsymbol{M}$.

Proof. Let $\boldsymbol{M}\in AC^p_T\big(\mathcal{P}_p(\mathcal{P}_p(X))\big)$ and apply to it the maps following the diagram above. Then for any $F:\mathcal{P}(X)\to[0,+\infty]$ Borel and for any $t\in[0,T]$, it holds

$$\int_{\mathcal{P}(X)} F(\mu)d\big((E_t)_\sharp\mathfrak{L}\big)(\mu) = \int F((e_t)_\sharp\lambda)d\mathfrak{L}(\lambda) = \int F\big((e_t)_\sharp(G(\boldsymbol{\mu})\big)d\Lambda(\boldsymbol{\mu})$$
$$= \int F(\mu_t)d\Lambda(\boldsymbol{\mu}) = \int_{\mathcal{P}(X)} F(\mu)dM_t(\mu),$$

where the second equality in the second line comes from $\mathfrak{L}=G_\sharp\Lambda$, the third one follows from $(e_t)_\sharp\circ G = E_t\circ G = \mathfrak{e}_t$, since $E\circ G=\mathrm{id}$, and the last one is a consequence of $(\mathfrak{e}_t)_\sharp\Lambda = M_t$. $\square$

Proposition 3.8. *Let $\Lambda\in\mathcal{P}_{\mathcal{A}_p}\big(C_T(\mathcal{P}_p(X)\big)$ and define $\mathfrak{L}=G_\sharp\Lambda$, $\boldsymbol{M}=\mathfrak{E}(\mathfrak{L})$ and take any $\Lambda_{\boldsymbol{M}}\in\mathrm{Lift}(\boldsymbol{M})$, according to the following diagram*



Then, both Λ *and* $\Lambda_{\boldsymbol{M}}$ *are lifting of* $\boldsymbol{M}$*, i.e.* $(\mathfrak{e}_t)_\sharp\Lambda = (\mathfrak{e}_t)_\sharp\Lambda_{\boldsymbol{M}} = M_t$ *for all* $t \in [0,T]$*, and for a.e.* $t \in [0,T]$ *it holds*

$$|\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t) = \int |\dot{\boldsymbol{\mu}}|^p_{W_{p,d}}(t)d\Lambda_{\boldsymbol{M}}(\boldsymbol{\mu}) \le \int |\dot{\boldsymbol{\mu}}|^p_{W_{p,d}}(t)d\Lambda(\boldsymbol{\mu}). \tag{3.19}$$

Proof. For all $t \in [0,T]$ and for all $F : \mathcal{P}_p(X) \to [0,+\infty)$ bounded Borel, we have

$$\begin{aligned}\int F(\mu)dM_t(\mu) &= \int F(\mu)d\big((E_t)_\sharp\mathfrak{L}\big)(\mu) = \int F\big((e_t)_\sharp\lambda\big)d\mathfrak{L}(\lambda) = \int F\big((e_t)_\sharp\lambda\big)d\big(G_\sharp\Lambda\big)(\lambda)\\ &= \int F\big((e_t)_\sharp G[\boldsymbol{\mu}]\big)d\Lambda(\boldsymbol{\mu}) = \int F(\mu_t)d\Lambda(\boldsymbol{\mu}) = \int F(\mu)d\big((\mathfrak{e}_t)_\sharp\Lambda\big)(\mu).\end{aligned}$$

For the second part, notice that

$$W_p^p(M_t,M_s) \le \int \mathcal{W}_p^p(\mu_t,\mu_s)d\Lambda'(\boldsymbol{\mu}) \le |t-s|^{p-1}\int\int_s^t |\dot{\boldsymbol{\mu}}|^p(r)drd\Lambda'(\boldsymbol{\mu})$$

for any $\Lambda' \in \mathcal{P}_{\mathcal{A}_p}\big(C_T(\mathcal{P}(X))\big)$ with marginal M_t at any time $t \in [0,T]$. This implies that

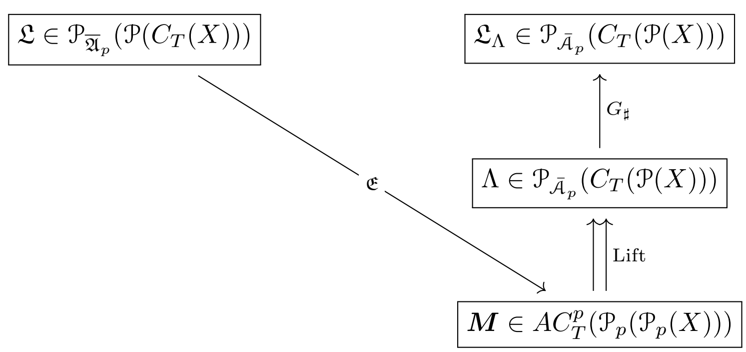
$$|\dot{\boldsymbol{M}}|^p(t) \le \int |\dot{\boldsymbol{\mu}}|^p(t)d\Lambda(\boldsymbol{\mu}) \quad \text{and} \quad |\dot{\boldsymbol{M}}|^p(t) \le \int |\dot{\boldsymbol{\mu}}|^p(t)d\Lambda_{\boldsymbol{M}}(\boldsymbol{\mu})$$

for a.e. $t \in [0,T]$. Moreover, thanks to Proposition 3.2, we have that

$$\int\int_0^T |\dot{\boldsymbol{\mu}}|^p(t)dtd\Lambda_{\boldsymbol{M}}(\boldsymbol{\mu}) = \int_0^T |\dot{\boldsymbol{M}}|^p(t)dt.$$

□

Proposition 3.9. *Let* $\mathfrak{L} \in \mathcal{P}_{\overline{\mathfrak{A}}_p}\big(\mathcal{P}_p(C_T(X))\big)$*. Define* $\boldsymbol{M} = \mathfrak{E}(\mathfrak{L})$*, take any* $\Lambda \in \mathrm{Lift}(\boldsymbol{M})$ *and define* $\mathfrak{L}_\Lambda = G_\sharp\Lambda$*, according to the following diagram*



Then $(E_t)_\sharp\mathfrak{L} = (E_t)_\sharp\mathfrak{L}_\Lambda = M_t$ *for all* $t \in [0,T]$*, and moreover*

$$|\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t) = \int\int |\dot{\boldsymbol{x}}|^p(t)d\lambda(\boldsymbol{x})d\mathfrak{L}_\Lambda(\lambda) \le \int\int |\dot{\boldsymbol{x}}|^p(t)d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda), \tag{3.20}$$

for a.e. $t \in [0,T]$.

Proof. The first part follows the same strategy above. Regarding the second part, from (3.15) and Proposition 3.8, we know that for any $s < t$

$$\int\int\int_s^t |\dot{\boldsymbol{x}}|^p_d(r)drd\lambda(\boldsymbol{x})d\mathfrak{L}_\Lambda(\lambda) = \int\int_s^t |\dot{\boldsymbol{\mu}}|^p_{W_p}(r)drd\Lambda(\boldsymbol{\mu}) = \int_s^t |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(r)dr.$$

Using Lebesgue theorem, we easily obtain that for a.e. $t \in [0, T]$ it holds

$$\int \int |\dot{\boldsymbol{x}}|_d^p(t) d\lambda(\boldsymbol{x}) d\mathfrak{L}_\Lambda(\lambda) = \int |\dot{\boldsymbol{\mu}}|_{W_p}^p(t) d\Lambda(\boldsymbol{\mu}) = |\dot{\boldsymbol{M}}|_{\mathcal{W}_p}^p(t) dt.$$

Moreover, from Proposition 3.6, for a.e. $t \in [0, T]$ it holds $|\dot{\boldsymbol{M}}|_{\mathcal{W}_p}^p(t) dt \le \int\int |\dot{\boldsymbol{x}}|_d^p(t) d\lambda(\boldsymbol{x}) d\mathfrak{L}(\lambda)$. □

3.6. **Geodesics of random measures.** In this subsection, we want to give a characterization for the *geodesics* in the space of random measures $(\mathcal{P}_p(\mathcal{P}_p(X)), \mathcal{W}_p)$. Assume that (X, d) is a complete, separable and geodesic metric space. It is well known that, under these assumptions on X, the space $(\mathcal{P}_p(X), W_p)$ is geodesic as well, for $p > 1$ (see e.g. [ABS24, Theorem 10.6]). Reiterating it, we already know that $(\mathcal{P}_p(\mathcal{P}_p(X)), \mathcal{W}_p)$ is geodesic.

With the notation introduced in §2.1.3, thanks to Corollary A.9, we define

$$\begin{gathered} \mathrm{GEO} : \mathcal{P}(X \times X) \to \mathcal{P}(C([0,1], X)), \quad \mathrm{GEO}(\pi) := \mathrm{geo}_\sharp \pi, \\ \mathrm{GEO}_t : \mathcal{P}(X \times X) \to \mathcal{P}(X), \quad \mathrm{GEO}_t(\pi) := (\mathrm{geo}_t)_\sharp \pi. \end{gathered} \tag{3.21}$$

Notice that, since geo is not Borel measurable in general, we cannot use Proposition D.8 to conclude that GEO (and GEO_t) is measurable. One way to deal with this would be to obtain GEO again using a measurable selection argument. For the sake of the presentation, we do not enter into details and we stick with the previous definitions, that work well whenever the geodesics are uniquely determined so that geo and GEO are Borel measurable.

On the other hand, we denote by $\mathrm{geo}_{\mathcal{P}_p} : \mathcal{P}_p(X) \times \mathcal{P}_p(X) \to \mathrm{Geo}(\mathcal{P}_p(X))$ the map defined in (2.6) in the geodesic space $\mathcal{P}_p(X)$. Similarly, $\mathrm{geo}_{t,\mathcal{P}_p}$ is its evaluation at time t, for all $t \in [0, 1]$.

First, we show how a geodesic connecting two given random measures $M_0, M_1 \in \mathcal{P}_p(\mathcal{P}_p(X))$ can be obtained either from $\Pi \in \Gamma_0(M_0, M_1)$ or $\mathfrak{P} \in \mathrm{R}\Gamma_0(M_0, M_1)$ (see §3.1).

Lemma 3.10. *Let $M_0, M_1 \in \mathcal{P}_p(\mathcal{P}_p(X))$. If $\mathfrak{P} \in \mathrm{R}\Gamma_0(M_0, M_1)$, then $M_t^{\mathfrak{P}} := (\mathrm{GEO}_t)_\sharp \mathfrak{P}$ is a geodesic in $(\mathcal{P}_p(\mathcal{P}_p(X)), \mathcal{W}_p)$ connecting M_0 and M_1.*
If $\Pi \in \Gamma_0(M_0, M_1)$, then $M_t^\Pi := (\mathrm{geo}_{t,\mathcal{P}_p})_\sharp \Pi$ is a geodesic in $(\mathcal{P}_p(\mathcal{P}_p(X)), \mathcal{W}_p)$ connecting M_0 and M_1.

Proof. Regarding the first claim, for all $0 \le s < t \le 1$, it holds

$$\begin{aligned} \mathcal{W}_p^p(M_s^{\mathfrak{P}}, M_t^{\mathfrak{P}}) &\le \int\int d^p(\mathrm{geo}_s(x,y), \mathrm{geo}_t(x,y)) d\pi(x,y) d\mathfrak{P}(\pi) \\ &= (t-s)^p \int\int d^p(x,y) d\pi(x,y) d\mathfrak{P}(\pi) = (t-s)^p \mathcal{W}_p^p(M_0, M_1). \end{aligned}$$

Similarly for the second claim:

$$\begin{aligned} \mathcal{W}_p^p(M_s^\Pi, M_t^\Pi) &\le \int W_p^p(\mathrm{geo}_{s,\mathcal{P}}(\mu,\nu), \mathrm{geo}_{t,\mathcal{P}}(\mu,\nu)) d\Pi(\mu,\nu) \\ &= (t-s)^p \int W_p^p(\mu,\nu) d\Pi(\mu,\nu) = (t-s)^p \mathcal{W}_p^p(M_0, M_1). \end{aligned}$$

□

The next proposition completely characterizes geodesics in terms of their liftings $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C([0,1], X)))$ and $\Lambda \in \mathcal{P}(C([0,1], \mathcal{P}(X)))$.

Proposition 3.11. *Let $\boldsymbol{M} = (M_t)_{t \in [0,1]} \in C([0,1], \mathcal{P}_p(\mathcal{P}_p(X)))$. The following are equivalent*

1. *$\boldsymbol{M}$ is a geodesic in $(\mathcal{P}_p(\mathcal{P}_p(X)), \mathcal{W}_p)$;*

(2) there exists $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C([0,1],X)))$ *lifting of* $\boldsymbol{M}$*, concentrated over* $\lambda \in \mathcal{P}(C([0,1],X))$ *that are, in turn, supported over* $\mathrm{Geo}(X)$*, and such that* $(E_{0,1})_\sharp \mathfrak{L} \in \mathrm{R}\Gamma_0(M_0, M_1)$*, where*

$$E_{0,1} : \mathcal{P}(C([0,1],X)) \to \mathcal{P}(X \times X), \quad E_{0,1}(\lambda) := (e_0, e_1)_\sharp \lambda; \tag{3.22}$$

(3) there exists $\Lambda \in \mathcal{P}(C([0,1],\mathcal{P}(X)))$ *lifting of* $\boldsymbol{M}$ *that is supported on* $\mathrm{Geo}(\mathcal{P}_p(X))$ *and such that* $(\mathfrak{e}_0, \mathfrak{e}_1)_\sharp \Lambda \in \Gamma_0(M_0, M_1)$.

Proof. (1) $\Longrightarrow$ (2): $\boldsymbol{M}$ being a geodesic implies that $\boldsymbol{M} \in AC_1^p(\mathcal{P}_p(\mathcal{P}_p(X)))$. Then, consider any $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C([0,1],X)))$ built as in Proposition 3.9, and we show it shares the wanted properties. By construction, it is a lifting of $\boldsymbol{M}$. Regarding the geodesic property, it holds

$$\begin{aligned} \mathcal{W}_p^p(M_0,M_1) &\le \int\int d^p(x_0,x_1) d\lambda(\boldsymbol{x}) d\mathfrak{L}(\lambda) \le \int\int \left(\int_0^1 |\dot{\boldsymbol{x}}|(r)dr\right)^p d\lambda(\boldsymbol{x}) d\mathfrak{L}(\lambda) \\ &\le \int\int\int_0^1 |\dot{\boldsymbol{x}}|^p(r) dr d\lambda(\boldsymbol{x}) d\mathfrak{L}(\lambda) = \int_0^1 |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(r)dr = \mathcal{W}_p^p(M_0,M_1), \end{aligned}$$

where the last equality follows from the fact that $\boldsymbol{M}$ is a geodesic and the second last one from Proposition 3.9. The previous computation forces all inequalities to be equalities. In particular, the first implies that $(E_{0,1})_\sharp \mathfrak{L} \in \mathrm{R}\Gamma_0(M_0,M_1)$, while the second and the third one imply, respectively, that for $\mathfrak{L}$-a.e. λ and for λ-a.e. $\boldsymbol{x}$, it holds

$$d(x_0,x_1) = \int_0^1 |\dot{\boldsymbol{x}}|(r)dr \quad \text{and} \quad |\dot{\boldsymbol{x}}| \text{ is constant},$$

that together imply that $\boldsymbol{x} \in \mathrm{Geo}(X)$.

(2) $\Longrightarrow$ (3): define $\Lambda := E_\sharp \mathfrak{L}$. It is immediate to verify that, if $\lambda \in \mathcal{P}(C([0,1],X))$ is supported over $\mathrm{Geo}(X)$ and $(e_0,e_1)_\sharp\lambda$ is optimal, then $\boldsymbol{\mu} := E(\lambda) \in \mathrm{Geo}(\mathcal{P}_p(X))$. We conclude observing $(\mathfrak{e}_0,\mathfrak{e}_1)_\sharp\Lambda = P_\sharp\big((E_{0,1})_\sharp\mathfrak{L}\big)$, where P is defined as in (3.6).

(3) $\Longrightarrow$ (1): thanks to Proposition 3.8 and the assumptions, it holds

$$\begin{aligned} \mathcal{W}_p^p(M_0,M_1) &\le \int_0^1 |\dot{\boldsymbol{M}}|^p(r)dr \le \int\int_0^1 |\dot{\boldsymbol{\mu}}|^p_{W_p}(r) dr d\Lambda(\boldsymbol{\mu}) \\ &= \int W_p^p(\mu_0,\mu_1) d\Lambda(\boldsymbol{\mu}) = \mathcal{W}_p^p(M_0,M_1). \end{aligned}$$

Reasoning as above, it implies that $\boldsymbol{M}$ is a constant speed geodesic in $(\mathcal{P}_p(\mathcal{P}_p(X)), \mathcal{W}_p)$. $\square$

The previous result can be improved when $X = \mathbb{R}^d$ (for simplicity, but the same holds whenever X is a Polish space for which there is uniqueness of geodesics), showing that all the objects involved are unique whenever we restrict ourselves to either $[0,t]$ or $[t,1]$, for $t \in (0,1)$, giving a non-branching property for geodesics of random measures.

First we need the operation of composition of random couplings, following [AGS08, Lemma 5.3.2].

Proposition 3.12. *Let* X *be a Polish space, and* $M_1, M_2, M_3 \in \mathcal{P}(X)$*. Let* $\mathfrak{P}_{1,2} \in \mathrm{R}\Gamma(M_1,M_2)$ *and* $\mathfrak{P}_{2,3} \in \mathrm{R}\Gamma(M_2,M_3)$*. Then, there exists* $\mathfrak{P}_{1,2,3} \in \mathcal{P}(\mathcal{P}(X \times X \times X))$ *such that*

$$P^{1,2}_\sharp \mathfrak{P}_{1,2,3} = \mathfrak{P}_{1,2}, \quad P^{2,3}_\sharp \mathfrak{P}_{1,2,3} = \mathfrak{P}_{2,3}, \tag{3.23}$$

where, for all $i,j = 1,2,3$

$$P^{i,j} : \mathcal{P}(X \times X \times X) \to \mathcal{P}(X \times X), \quad P^{i,j}(\theta) = p^{i,j}_\sharp \theta.$$

In particular, $P^{1,3}_\sharp \mathfrak{P}_{1,2,3} \in \mathrm{R}\Gamma(M_1,M_3)$.

Proof. The proof again uses a measurable selection argument. Let $\mathcal{A} \subset \mathcal{P}(X \times X) \times \mathcal{P}(X \times X)$ be the Borel subset defined as

$$\mathcal{A} = \left\{(\pi, \sigma) \in \mathcal{P}(X \times X) \times \mathcal{P}(X \times X) \,:\, p^2_\sharp \pi = p^1_\sharp \sigma\right\}.$$

Define the map

$$\mathrm{P} = (P^{1,2}, P^{2,3}) : \mathcal{P}(X \times X \times X) \to \mathcal{A}, \quad \mathrm{P}(\theta) := (p^{1,2}_\sharp \theta, p^{2,3}_\sharp \theta).$$

It is Borel thanks to Proposition D.8. Moreover, thanks to [AGS08, Lemma 5.3.2 & Remark 5.3.3], the map P is surjective. Thus, we can apply Theorem A.10 to obtain a Souslin-Borel measurable map $\mathrm{Q} : \mathcal{A} \to \mathcal{P}(X \times X \times X)$ that is a right-inverse of P.
Now, consider the disintegration of $\mathfrak{P}_{1,2}$ and $\mathfrak{P}_{2,3}$, respectively, with respect to the maps P^1 and P^2, to obtain that

$$\mathfrak{P}_{1,2} = \int_{\mathcal{P}} \mathfrak{P}_{1,2,\mu} dM_t(\mu), \quad \mathfrak{P}_{2,3} = \int_{\mathcal{P}} \mathfrak{P}_{2,3,\mu} dM_t(\mu),$$

where $\mathfrak{P}_{1,2,\mu} \in \mathcal{P}(\mathcal{P}(X \times X))$ is concentrated over couplings π for which $p^2_\sharp \pi = \mu$, for M_t-a.e. $\mu \in \mathcal{P}(X)$, and similarly for $\mathfrak{P}_{2,3,\mu}$. Notice that, for M_t-a.e. μ, the product measure $\mathfrak{P}_{1,2,\mu} \otimes \mathfrak{P}_{2,3,\mu}$ is concentrated over $\mathcal{A}$. Thus, we can define

$$\mathfrak{P}_{1,2,3} := \mathrm{Q}_\sharp \left(\int_{\mathcal{P}} \mathfrak{P}_{1,2,\mu} \otimes \mathfrak{P}_{2,3,\mu} dM_t(\mu)\right) = \int_{\mathcal{P}} \mathrm{Q}_\sharp \left(\mathfrak{P}_{1,2,\mu} \otimes \mathfrak{P}_{2,3,\mu}\right) dM_t(\mu), \tag{3.24}$$

It is not hard to show that $\mathfrak{P}_{1,2,3}$ shares the wanted properties. □

Theorem 3.13. *Let $\boldsymbol{M} = (M_t)_{t \in [0,1]} \in C([0,1], \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$ be a geodesic in $(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)), \mathcal{W}_p)$. Then, for every $t \in (0,1)$, $\mathrm{R}\Gamma_0(M_0, M_t)$ (resp. $\mathrm{R}\Gamma_0(M_t, M_1)$) contains a unique optimal random coupling $\mathfrak{P}_{0,t}$ (resp. $\mathfrak{P}_{t,1}$).*
Moreover, there exists a unique $\mathfrak{L}_{0,t} \in \mathcal{P}(\mathcal{P}(C([0,t], \mathbb{R}^d)))$ (resp. $\mathfrak{L}_{t,1} \in \mathcal{P}(\mathcal{P}(C([t,1], \mathbb{R}^d)))$) lifting of $(M_s)_{s \in [0,t]}$ (resp. $(M_s)_{s \in [t,1]}$) satisfying property (2) in Proposition 3.11.
Similarly, there exists a unique $\Lambda_{0,t} \in \mathcal{P}(C([0,t], \mathcal{P}(\mathbb{R}^d)))$ (resp. $\Lambda_{t,1} \in \mathcal{P}(C([t,1], \mathcal{P}(\mathbb{R}^d))))$) lifting of $(M_s)_{s \in [0,t]}$ (resp. $(M_s)_{s \in [t,1]}$) satisfying property (3) in Proposition 3.11.

Proof. Let $\mathfrak{P}_{0,t} \in \mathrm{R}\Gamma_0(M_0, M_t)$ and $\mathfrak{P}_{t,1} \in \mathrm{R}\Gamma_0(M_t, M_1)$ be any optimal random coupling. Applying Proposition 3.12, we obtain $\mathfrak{P}_{0,t,1} \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d))$ satisfying

$$P^{1,2}_\sharp \mathfrak{P}_{0,t,1} = \mathfrak{P}_{0,t}, \quad P^{2,3}_\sharp \mathfrak{P}_{0,t,1} = \mathfrak{P}_{t,1}, \quad \mathfrak{P}_{0,1} := P^{1,3}_\sharp \mathfrak{P}_{0,t,1} \in \mathrm{R}\Gamma(M_0, M_1). \tag{3.25}$$

Exploiting the geodesic property, we can also show that $\mathfrak{P}_{0,1} \in \mathrm{R}\Gamma_0(M_0, M_1)$:

$$\begin{aligned}
\mathcal{W}_p(M_0, M_1) &\leq \left(\iint |x-y|^p d\pi(x,y) d\mathfrak{P}_{0,1}(\pi)\right)^{\frac{1}{p}} = \left(\iint |x_1 - x_3|^p d\theta(x_1,x_2,x_3) d\mathfrak{P}_{0,t,1}(\theta)\right)^{\frac{1}{p}} \\
&\leq \left(\iint |x_1 - x_2|^p d\theta(x_1,x_2,x_3) d\mathfrak{P}_{0,t,1}(\theta)\right)^{\frac{1}{p}} + \left(\iint |x_2 - x_3|^p d\theta(x_1,x_2,x_3) d\mathfrak{P}_{0,t,1}(\theta)\right)^{\frac{1}{p}} \\
&= \mathcal{W}_p(M_0, M_t) + \mathcal{W}_p(M_t, M_1) = \mathcal{W}_p(M_0, M_1).
\end{aligned}$$

An important property is also hidden in the previous computation: since the last inequality is actually an equality, we know that for $\mathfrak{P}_{0,t,1}$-a.e. $\theta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ and for θ-a.e. (x_1, x_2, x_3), the three points are aligned, in the sense that there exists $\alpha \in (0,1)$ such that $x_2 - x_1 = \alpha(x_3 - x_1)$. Moreover, using again the geodesic property, it is not hard to show that

$\alpha = t$. In particular, for $\mathfrak{P}_{0,t,1}$-a.e. $\theta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ and for θ-a.e. (x_1, x_2, x_3), it holds $x_1 = \frac{x_2 - tx_3}{1-t}$ and $x_3 = \frac{x_2-(1-t)x_1}{t}$. Thus, define the functions

$$\ell_1(x_2, x_3) := \left(\frac{x_2 - tx_3}{1-t}, x_2, x_3\right), \quad \mathrm{L}_1 := (\ell_1)_\sharp : \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d),$$

$$\ell_3(x_1, x_2) := \left(x_1, x_2, \frac{x_2 - (1-t)x_1}{t}\right), \quad \mathrm{L}_3 := (\ell_3)_\sharp : \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d).$$

Thanks to the previous observations, we can conclude that $\mathfrak{P}_{0,t} = P^{1,2}_\sharp \mathfrak{P}_{0,t,1} = P^{1,2}_\sharp\big((\mathrm{L}_1)_\sharp \mathfrak{P}_{t,1}\big)$, that makes us conclude that $\mathfrak{P}_{0,t}$ is unique, since it has been chosen independently of $\mathfrak{P}_{t,1}$. Similarly, the uniqueness for $\mathfrak{P}_{t,1}$ is implied by $\mathfrak{P}_{t,1} = P^{2,3}_\sharp\big((\mathrm{L}_3)_\sharp \mathfrak{P}_{0,t}\big)$.

The uniqueness of $\mathfrak{L}_{0,t}$ follows from Proposition 3.11 and observing that the required properties force it to be equal to $P^{0,t}_\sharp \mathfrak{P}_{0,t}$, where

$$P^{0,t} := p^{0,t}_\sharp, \quad p^{0,t} : \mathbb{R}^d \times \mathbb{R}^d \to C([0,t], \mathbb{R}^d), \; p^{0,t}(x_1, x_2) = [s \mapsto (1-s)x_1 + sx_2].$$

The uniqueness of $\Lambda_{0,t}$ now follows, and similarly, the same holds for $\mathfrak{L}_{t,1}$ and $\Lambda_{t,1}$. □

3.7. **Barycenters.** In this subsection, we apply our result to show that, if $\boldsymbol{M} \in AC^p_T(\mathcal{P}(\mathcal{P}(X)))$, then the curves of its barycenters is in $AC^p_T(\mathcal{P}(X))$.

Definition 3.14. *Let $M \in \mathcal{P}(\mathcal{P}(X))$. The barycenter of M is the measure $\mathrm{bar}[M] \in \mathcal{P}(X)$ defined such that it satisfies*

$$\int_X f(x) d\big(\mathrm{bar}[M]\big)(x) = \int \int_X f(x) d\mu(x) dM(\mu) \tag{3.26}$$

for all Borel functions $f : X \to [0, +\infty]$.

Recalling the definition of $\widetilde{M} \in \mathcal{P}(X \times \mathcal{P}(X))$ given in Remark 2.6, we notice that the first marginal of $\widetilde{M}$ is indeed $\mathrm{bar}[M]$. Now, we show a nice property of the barycenter with respect to the nested push-forward described in (2.9).

Lemma 3.15. *Let X, Y be two Polish spaces, $M \in \mathcal{P}(\mathcal{P}(X))$ and $f : X \to Y$ a Borel map. Define the map $F : \mathcal{P}(X) \to \mathcal{P}(Y)$ as the push-forward of f, i.e. $F = f_\sharp$. Then it holds*

$$\mathrm{bar}[F_\sharp M] = f_\sharp(\mathrm{bar}[M]). \tag{3.27}$$

Proof. Let $N = F_\sharp M$. For all $g : Y \to [0, +\infty]$ Borel measurable, thanks to (2.10), it holds

$$\begin{aligned}\int_Y g(y) d\,\mathrm{bar}[N](y) &= \int_{\mathcal{P}(Y)} \int_Y g(y) d\nu(y) dN(\nu) = \int_{\mathcal{P}(X)} \int_X g(f(x)) d\mu(x) dM(\mu) \\ &= \int_X g(f(x)) d\big(\mathrm{bar}[M]\big)(x) = \int_Y g(y) d f_\sharp(\mathrm{bar}[N])(y).\end{aligned}$$

□

Given a curve $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(X)))$, we indicate with $\mathbf{bar}[\boldsymbol{M}] = (\mathrm{bar}[M_t])_{t\in[0,T]} \in C_T(\mathcal{P}(X))$ the curve of the barycenters.

Proposition 3.16. *If $\boldsymbol{M} \in AC^p_T\big(\mathcal{P}_p(\mathcal{P}_p(X))\big)$, then $\boldsymbol{m} := \mathbf{bar}[\boldsymbol{M}] \in AC^p_T(\mathcal{P}_p(X))$, and, for a.e. $t \in (0, T)$, it holds*

$$|\dot{\boldsymbol{m}}|^p_{W_p}(t) \le \int \int |\dot{\boldsymbol{x}}|^p(t) d\lambda(\boldsymbol{x}) d\mathfrak{L}(\lambda) = |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t). \tag{3.28}$$

Proof. For all $t \in [0,T]$, $\mathrm{bar}[M_t] \in \mathcal{P}_p(X)$, indeed

$$\int_X d^p(x,\bar{x})d\,\mathrm{bar}[M_t](x) = \int\int_X d^p(x,\bar{x})d\mu(x)dM_t(\mu) < +\infty.$$

Consider $\mathfrak{L} \in \mathcal{P}_{\mathcal{A}_p}\big(\mathcal{P}(C_T(X))\big)$ defined as $\mathfrak{L} := (G_\sharp) \circ \mathrm{Lift}(\boldsymbol{M})$. Define the map

$$(E_t, E_s) : \mathcal{P}(C_T(X)) \to \mathcal{P}(X \times X)$$
$$\lambda \mapsto (e_t, e_s)_\sharp \lambda,$$

and $\mathfrak{P}_{t,s} := (E_t, E_s)_\sharp \mathfrak{L} \in \mathcal{P}\big(\mathcal{P}(X \times X)\big)$. Notice that the barycenter of $\mathfrak{P}_{t,s}$, indicated as $\pi_{t,s} \in \mathcal{P}(X \times X)$, is a coupling between $\mathrm{bar}[M_t]$ and $\mathrm{bar}[M_s]$: using Lemma 3.15, indeed $\mathrm{bar}[(E_t)_\sharp \mathfrak{L}] = (e_t)_\sharp(\mathrm{bar}[\mathfrak{L}])$, and same for s. Thus, we have

$$\begin{aligned} W_p^p(\mathrm{bar}[M_t], \mathrm{bar}[M_s]) &\le \int\int d^p(x,y)d\pi(x,y)d\mathfrak{P}_{t,s}(\pi) \\ &= \int\int d^p(\gamma_t,\gamma_s)d\lambda(\gamma)d\mathfrak{L}(\lambda) \le |t-s|^{p-1}\int_s^t\int\int|\dot\gamma|^p(r)d\lambda(\gamma)d\mathfrak{L}(\lambda)dr, \end{aligned}$$

which implies that $\boldsymbol{m} := \mathbf{bar}[\boldsymbol{M}] \in AC_T(\mathcal{P}_p(X))$, and using Lebesgue theorem, it holds

$$|\dot{\boldsymbol{m}}|^p_{W_p}(t) \le \int\int |\dot{\boldsymbol{x}}|^p(t)d\lambda(\boldsymbol{x})d\mathfrak{L}(\lambda) = |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t) \quad \text{for a.e. } t \in [0,T]. \tag{3.29}$$

□

4. Continuity equation on random measures

4.1. **Derivations, continuity equation and a first superposition result.** Derivations are the natural objects that can be used to define an abstract continuity equation over a metric space X, see [ST17]. Here, we adapt this definition in the case $X = \mathcal{P}(\mathbb{R}^d)$, endowed with the narrow topology.

Definition 4.1 (Cylinder functions and Wasserstein gradient)**.** *A functional $F : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is called a C^1_c-cylinder function if there exists $k \in \mathbb{N}$, $\Phi = (\phi_1,\dots,\phi_k) \in C^1_c(\mathbb{R}^d;\mathbb{R}^k)$ and $\Psi \in C^1_b(\mathbb{R}^k)$ such that*

$$F(\mu) = \Psi\left(L_\Phi(\mu)\right), \quad L_\Phi(\mu) = \big(L_{\phi_1}(\mu),\dots,L_{\phi_k}(\mu)\big), \quad L_{\phi_i}(\mu) := \int_{\mathbb{R}^d}\phi_i(x)d\mu(x). \tag{4.1}$$

Its Wasserstein gradient is then defined as

$$\nabla_W F(x,\mu) := \sum_{i=1}^k \partial_i\Psi\left(L_\Phi(\mu)\right)\nabla\phi_i(x) \quad \forall x \in \mathbb{R}^d,\ \forall \mu \in \mathcal{P}(\mathbb{R}^d). \tag{4.2}$$

The collection of all the cylinder functions is called $\mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$. If $F = \Psi \circ L_\Phi$ with $\Phi \in C^1_b(\mathbb{R}^d)$ and $\Psi \in C^1_b(\mathbb{R}^k)$, we say that $F \in \mathrm{Cyl}^1_b(\mathcal{P}(\mathbb{R}^d))$.

Notice that, when $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$, we can consider the outer function $\Psi \in C^1_c(\mathbb{R}^k)$. Moreover, we have the inclusion $\mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)) \subset \mathrm{Cyl}^1_b(\mathcal{P}(\mathbb{R}^d))$.

Remark 4.2. *Since $\nabla_W F(x,\mu) = \nabla_x\left(\frac{d^+}{d\varepsilon}|_{\varepsilon=0}\ F((1-\varepsilon)\mu + \varepsilon\delta_x)\right)$, the Wasserstein gradient does not depend on the representation chosen for F.*

Definition 4.3 (L^p derivations)**.** *Let $M \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ and $p \geq 1$. An $L^p(M)$-derivation is a linear operator $B : \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)) \to L^p(M)$ such that*

$$B[FG] = GB[F] + FB[G] \quad M\text{-a.e.} \tag{4.3}$$

and there exists a non-negative function $c \in L^p(M)$ such that

$$|B[F](\mu)| \leq c(\mu)\|\nabla_W F(\cdot,\mu)\|_{L^{p'}(\mu)} \quad \forall F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)), \quad \text{for } M\text{-a.e. } \mu \in \mathcal{P}(\mathbb{R}^d), \tag{4.4}$$

where p' is the conjugate exponent of p.

We will extensively work with families of L^p-derivations: to be more specific, let $(M_t)_{t\in[0,T]} \subset \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ be a Borel family of random measures and $(B_t)_{t\in[0,T]}$ a family of $L^p(M_t)$-derivations such that $(t,\mu) \mapsto B_t[F](\mu)$ is Borel measurable for all $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ and there exists a non-negative function $c \in L^p(M_t \otimes dt)$ (recall the notation introduced in §2.3.1) such that

$$|B_t[F](\mu)| \leq c_t(\mu)\|\nabla_W F\|_{L^{p'}(\mu)} \; M_t \otimes dt\text{-a.e.} \quad \text{and} \quad \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} c_t^p(\mu) dM_t(\mu) dt < +\infty. \tag{4.5}$$

We will refer to such a kind of family of derivations as $L^p(M_t \otimes dt)$-derivations. Notice that, since all the measures involved are probability measures, an $L^p(M)$-derivation is also an $L^q(M)$-derivation for each $q \in [1,p]$.

Remark 4.4. *Thanks to* (4.4)*, given a representative of c, there exists $\mathcal{N} \subset \mathcal{P}(\mathbb{R}^d)$ such that $M(\mathcal{N}) = 0$ and $|B[F](\mu)| \leq c(\mu)\|\nabla_W F(\cdot,\mu)\|_{L^{p'}(\mu)} < +\infty$ for all $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ and $\mu \in \mathcal{N}^c$. Similarly for $L^p(M_t \otimes dt)$-derivations, exploiting* (4.5)*, there exists $\widetilde{\mathcal{N}} \subset \mathcal{P}(\mathbb{R}^d) \times [0,T]$ such that $M_t \otimes dt(\widetilde{\mathcal{N}}) = 0$ and $|B_t[F](\mu)| \leq c_t(\mu)\|\nabla_W F(\cdot,\mu)\|_{L^{p'}(\mu)} < +\infty$ for all $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ and $(\mu,t) \in \widetilde{\mathcal{N}}^c$.*

Definition 4.5 (Continuity equation on random measures)**.** *Let $(M_t)_{t\in[0,T]}$ be a continuous curve of Borel probability measures over $\mathcal{P}(\mathbb{R}^d)$, i.e. $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$. Let $(B_t)_{t\in[0,T]}$ be a family of $L^1(M_t)$-derivations, according to* (4.5)*. We say that the continuity equation*

$$\frac{d}{dt} M_t + \mathrm{div}_{\mathcal{P}} \big(B_t M_t\big) = 0 \tag{4.6}$$

is satisfied if

$$\forall F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)) \quad \frac{d}{dt}\int_{\mathcal{P}} F(\mu)\, dM_t(\mu) = \int_{\mathcal{P}} B_t[F](\mu) dM_t(\mu) \tag{4.7}$$

in the sense of distributions in $(0,T)$, where $\mathcal{P}$ is short for $\mathcal{P}(\mathbb{R}^d)$.

Remark 4.6. *The assumption that the curve $t \mapsto M_t$ is continuous, is not restrictive. Indeed, if such curve is just Borel measurable, thanks to Lemma C.7, we can always find a (unique) continuous representative of it.*

Before proceeding, we state and prove two technical, but useful, lemmas.

Lemma 4.7 (Chain rule)**.** *Let $M \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ and B be an $L^1(M)$-derivation. Then, for all $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ of the form $F = \Psi(L_\Phi(\mu))$ as in* (4.1)*, for M-a.e. $\mu \in \mathcal{P}(\mathbb{R}^d)$ it holds*

$$B[F](\mu) = \sum_{i=1}^k \partial_i \Psi(L_\Phi(\mu)) B[L_{\phi_i}](\mu). \tag{4.8}$$

Proof. Thanks to the Leibniz rule, by linearity and induction (4.8) holds when Ψ is a polynomial. When Ψ is not a polynomial, consider (p_n) a sequence of polynomial approximating Ψ uniformly on compact sets, together with its first derivatives. By the boundedness of Φ and its first derivatives, we conclude. $\square$

Lemma 4.8. *Let $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ and $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ such that $(\mathfrak{e}_t)_\sharp \Lambda = M_t$ for all $t \in [0,T]$. Let $p \geq 1$. Let $(B_t)_{t\in[0,T]}$ be a family of $L^p(M_t \otimes dt)$-derivations and $c \in L^p(M_t \otimes dt)$ as in (4.5). Then the functions*

$$(t, \boldsymbol{\mu}) \mapsto B_t[F](e_t(\boldsymbol{\mu})), \quad (t, \boldsymbol{\mu}) \mapsto c_t^p(e_t(\boldsymbol{\mu})) \tag{4.9}$$

are $\mathcal{L}^1_T \otimes \Lambda$-measurable and well-defined. Moreover, for $\mathcal{L}^1_T \otimes \Lambda$-a.e. $(t, \boldsymbol{\mu})$ and for all $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$, it holds

$$|B_t[F](e_t(\boldsymbol{\mu}))| \leq c_t(e_t(\boldsymbol{\mu}))|\nabla_W F(\cdot, e_t(\boldsymbol{\mu})\|_{L^{p'}(e_t(\boldsymbol{\mu}))} < +\infty, \quad \int_0^T c_t(e_t(\boldsymbol{\mu}))dt < +\infty. \tag{4.10}$$

Proof. The functions in (4.9) are composition of measurable maps, so they are measurable. Now, consider $\widetilde{\mathcal{N}}$ as in Remark 4.4 and define the function

$$\begin{aligned} \widetilde{E} : [0,T] \times C_T(\mathcal{P}(\mathbb{R}^d)) &\to [0,T] \times \mathcal{P}(\mathbb{R}^d) \\ (t, \boldsymbol{\mu}) &\mapsto (t, \mu_t), \end{aligned}$$

where we mean that $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]}$, so that $\mu_t = e_t(\boldsymbol{\mu})$. Notice that $\widetilde{E}_\sharp(\mathcal{L}^1_T \otimes \Lambda) = dt \otimes M_t$, so $\widetilde{E}^{-1}(\widetilde{\mathcal{N}})$ is a negligible set w.r.t. $\mathcal{L}^1_T \otimes \Lambda$, which implies (4.10). To conclude, notice that

$$\int \int_0^T c_t^p(\mu_t) dt d\Lambda(\boldsymbol{\mu}) = \int_0^T \int c_t^p(\mu_t) d\Lambda(\boldsymbol{\mu}) dt = \int_0^T \int_{\mathcal{P}} c_t^p(\mu) dM_t(\mu) dt < +\infty.$$

$\square$

Now, we prove a first superposition result in terms of derivations. Our proof strongly relies on the techniques developed in [AT14]: indeed, we embed the space $\mathcal{P}(\mathbb{R}^d)$ in $\mathbb{R}^\infty$, where an infinite-dimensional version of Theorem 2.12 holds (see Appendix B).

We are going to use a similar notation as for the purely metric setting:

- $\boldsymbol{\gamma} := (\gamma_t)_{t\in[0,T]} \subset C_T(\mathbb{R}^d)$;
- $\lambda \in \mathcal{P}\big(C_T(\mathbb{R}^d)\big)$;
- $\boldsymbol{\mu} := (\mu_t)_{t\in[0,T]} \in C_T\big(\mathcal{P}(\mathbb{R}^d)\big)$;
- $M \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$;
- $\boldsymbol{M} := (M_t)_{t\in[0,T]} \in C_T\big(\mathcal{P}(\mathcal{P}(\mathbb{R}^d))\big)$;
- $\Lambda \in \mathcal{P}\big(C_T(\mathcal{P}(\mathbb{R}^d))\big)$;
- $\mathfrak{L} \in \mathcal{P}\big(\mathcal{P}(C_T(\mathbb{R}^d))\big)$.

Theorem 4.9. *Let $\boldsymbol{M} = (M_t)_{t\in[0,T]}$ be a continuous curve of probability measures over $\mathcal{P}(\mathbb{R}^d)$. Let $(B_t)_{t\in[0,T]}$ be a Borel family of derivations and $c \in L^1(M_t \otimes dt)$ as in (4.5) with $p = 1$. Assume that $(M_t)_{t\in[0,T]}$ satisfies the continuity equation $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(B_t M_t) = 0$ in the sense of Definition 4.5. Then there exists a probability measure $\Lambda \in \mathcal{P}\big(C_T(\mathcal{P}(\mathbb{R}^d)\big)$, such that*

- *$(\mathfrak{e}_t)_\sharp \Lambda = M_t$ for all $t \in [0,T]$;*
- *for Λ-a.e. curve $(\mu_t)_{t\in[0,T]}$ and a.e. $t \in [0,T]$, $B_t[F](\mu_t)$ and $c_t(\mu_t)$ are well defined and it holds*

$$\int_0^T c_t(\mu_t) dt < +\infty \quad \textit{and} \quad \partial_t \mu_t + \mathrm{div}\left(B_t \mu_t\right) = 0 \quad \textit{for } \Lambda\textit{-a.e. } (\mu_t)_{t\in[0,T]} \tag{4.11}$$

in the sense of distributions in duality with C^1_c functions, i.e. for all $\phi \in C^1_c(\mathbb{R}^d)$ and $\psi \in C^1_c(0,T)$

$$\int_0^T \int_{\mathbb{R}^d} \psi'(t)\phi(x) d\mu_t(x) dt = -\int_0^T \psi(t) B_t[L_\phi](\mu_t) dt. \tag{4.12}$$

Proof. **Step 1**: here we use the result presented in Appendix B. Let $\mathcal{A} = \{\phi_1, \phi_2, \dots\} \subset C^1_c(\mathbb{R}^d)$ and ι as in Appendix B, i.e.

- ϕ_k is Lipschitz with respect to $|\cdot| \wedge 1$ for all $k \in \mathbb{N}$. In particular, $\|\nabla \phi_k\|_\infty \leq 1$ for all $k \in \mathbb{N}$;
- $\mathrm{Span}(\mathcal{A})$ dense in $C^1_0(\mathbb{R}^d)$, i.e. functions that, together with their first derivatives, converge to 0 at infinity;
- $\hat{W}_1(\mu,\nu) := W_{1,|\cdot|\wedge 1}(\mu,\nu) = \sup_k \int_{\mathbb{R}^d} \phi_k d(\mu - \nu)$;

and

$$\iota : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^\infty, \quad \iota(\mu) = \big(L_{\phi_1}(\mu), L_{\phi_2}(\mu), \dots\big). \tag{4.13}$$

Step 2: define $\mathtt{m}_t := \iota_\sharp M_t$. We prove that $(\mathtt{m}_t) \in AC_T(\mathcal{P}(\mathbb{R}^\infty), W_{1,D_\infty})$ (see (B.1)). For any $s,t \in [0,T]$, consider $\Pi_{s,t}$ an optimal transport plan between M_t and M_s realizing $\hat{\mathcal{W}}_1(M_t,M_s) := W_{1,\hat{W}_1}(M_t,M_s)$. Then $(\iota,\iota)_\sharp \Pi_{t,s}$ is a transport plan between $\mathtt{m}_t$ and $\mathtt{m}_s$, so that

$$\begin{aligned} W_{1,D_\infty}(\mathtt{m}_t,\mathtt{m}_s) &\leq \int D_\infty(\mathrm{x},\mathrm{y})\ d(\iota,\iota)_\sharp \Pi_{t,s}(x,y) = \int \sup_k |L_{\phi_k}(\mu) - L_{\phi_k}(\nu)| \wedge 1\ d\Pi_{t,s}(\mu,\nu) \\ &= \int \sup_k |L_{\phi_k}(\mu) - L_{\phi_k}(\nu)| d\Pi_{t,s}(\mu,\nu) = \int \hat{W}_1(\mu,\nu) d\Pi_{t,s}(\mu,\nu) = \hat{\mathcal{W}}_1(M_t,M_s), \end{aligned}$$

and we are done thanks to Lemma C.7.

Step 3: define, component-wisely, the vector field $v : [0,T] \times \mathbb{R}^\infty \to \mathbb{R}^\infty$ as

$$v^{(k)}_t(\mathrm{x}) := \begin{cases} B_t[L_{\phi_k}](\iota^{-1}(\mathrm{x})) & \text{if } \mathrm{x} \in \iota(\mathcal{P}(\mathbb{R}^d)) \\ 0 & \text{if } \mathrm{x} \notin \iota(\mathcal{P}(\mathbb{R}^d)). \end{cases} \tag{4.14}$$

Notice that

$$\int_0^T \int_{\mathbb{R}^\infty} |v^{(k)}_t(\mathrm{x})| d\mathtt{m}_t(\mathrm{x}) dt = \int_0^T \int_{\mathcal{P}} |B_t[L_{\phi_k}](\mu_t)| dM_t(\mu) dt \leq \int_0^T \int_{\mathcal{P}} c_t(\mu) dM_t(\mu) dt < +\infty,$$

and for any cylinder function $F : \mathbb{R}^\infty \to \mathbb{R}$, i.e. such that there exists $n \in \mathbb{N}$ for which $F(\mathrm{x}) = F(x_1, \dots, x_n)$, it holds (thanks to Lemma 4.8)

$$\begin{aligned} &\frac{d}{dt} \int_{\mathbb{R}^\infty} F(\mathrm{x}) d\mathtt{m}_t(\mathrm{x}) = \frac{d}{dt} \int_{\mathcal{P}} F(\iota(\mu)) dM_t(\mu) = \frac{d}{dt} \int_{\mathcal{P}} F(L_{\phi_1}(\mu), \dots, L_{\phi_n}(\mu)) dM_t(\mu) \\ &= \sum_{i=1}^n \int_{\mathcal{P}} \partial_i F(L_{\phi_1}(\mu), \dots, L_{\phi_n}(\mu))\ B_t[L_{\phi_k}](\mu_t) dM_t(\mu) = \int_{\mathbb{R}^\infty} \nabla F(\mathrm{x}) \cdot \big(v^{(1)}_t, \dots, v^{(n)}_t\big) d\mathtt{m}_t(\mathrm{x}). \end{aligned}$$

Then, we can apply Theorem B.5, to obtain the existence of a measure $\mathtt{L} \in \mathcal{P}\big(C_T(\mathbb{R}^\infty, d_\infty)\big)$ satisfying

$$(e_t)_\sharp \mathtt{L} = \mathtt{m}_t$$

and $\mathtt{L}$-a.e. $\tilde{\gamma} \in C_T(\mathbb{R}^\infty)$ is weakly absolutely continuous with

$$\frac{d}{dt} \tilde{\gamma}(t) = v_t(\tilde{\gamma}(t)),$$

in the sense that each component $\tilde{\gamma}^{(k)}$ of the curve is in $AC([0,T],\mathbb{R})$ and $\frac{d}{dt}\tilde{\gamma}^{(k)} = v^{(k)}_t(\tilde{\gamma}(t))$.

Step 4: we prove that for $\mathtt{L}$-a.e. $\tilde{\gamma}$, $\tilde{\gamma} \in AC_T(\iota(\mathcal{P}(\mathbb{R}^d)), D_\infty)$. In particular, for every $t \in [0,T]$, $\tilde{\gamma}(t) \in \iota(\mathcal{P})$. First, observe that for $\mathtt{L}$-a.e. $\tilde{\gamma}$, $\tilde{\gamma}(t) \in \iota(\mathcal{P}(\mathbb{R}^d))$ for any $t \in \mathbb{Q} \cap [0,T]$. Fix $t \in [0,T] \cap \mathbb{Q}$, then

$$\mathtt{L}\big(\{\tilde{\gamma} \ : \ \tilde{\gamma}(t) \in \iota(\mathcal{P}(\mathbb{R}^d))\}\big) = \mathtt{m}_t(\iota(\mathcal{P}(\mathbb{R}^d))) = 1,$$

so we conclude because $[0,T] \cap \mathbb{Q}$ is countable.

Now, for $\mathtt{L}$-a.e. $\big(\tilde{\gamma}(t)\big)_{t\in[0,T]}$, $\tilde{\gamma} \in AC_T(\mathbb{R}^\infty, D_\infty)$. In fact, $\mathtt{L}$-a.e. curve $\tilde{\gamma}$ and for any $s,t \in [0,T]$, it holds

$$D_\infty(\tilde{\gamma}(t), \tilde{\gamma}(s)) = \sup_n |\tilde{\gamma}_n(t) - \tilde{\gamma}_n(s)| \wedge 1 \leq \sup_n \int_s^t |v_r^{(n)}(\tilde{\gamma}(r))| dr \leq \int_s^t \sup_n |v_r^{(n)}(\tilde{\gamma}_r)| dr.$$

Moreover, notice that

$$\begin{aligned} \int \int_0^T \sup_n |v_r^{(n)}(\tilde{\gamma}_r)| dr d\mathtt{L}(\tilde{\gamma}) &= \int_0^T \int \sup_n |v_r^{(n)}(\tilde{\gamma}_r)| d\mathtt{L}(\tilde{\gamma}) dr = \int_0^T \int \sup_n |v_r^{(n)}(\mathrm{x})| d\mathtt{m}_r(\mathrm{x}) dr \\ &\leq \int_0^T \int \sup_n |v_r(n)(\iota(\mu))| dM_r(\mu) dr \leq \int_0^T \int c_r(\mu) dM_r(\mu) dr < +\infty, \end{aligned}$$

where we used that $\sup_n |v_r(n)(\iota(\mu))| = \sup_n |B_r[L_{\phi_n}](\mu)| \leq c_r(\mu)$. This implies that for $\mathtt{L}$-a.e. $\tilde{\gamma}$ it holds that

$$\int_0^T \sup_n |v_r^{(n)}(\tilde{\gamma}_r)| dr < +\infty$$

and thanks to the inequalities above, we have that for $\mathtt{L}$-a.e. $\tilde{\gamma}$, $\tilde{\gamma} \in AC_T(\mathbb{R}^\infty, D_\infty)$. The above properties show that for $\mathtt{L}$-a.e. $\tilde{\gamma}$, $\tilde{\gamma}(t) \in \iota(\mathcal{P}(\mathbb{R}^d))$ for any $t \in [0,T]$, thanks to Lemma B.1

Step 5: thanks to the previous step, $\tilde{\Lambda}$ is concentrated over $C_T(\iota(\mathcal{P}), D_\infty)$, thus, thanks to Lemma C.3, it can be seen as a probability measure over it, with its natural induced compact-open topology. Now, consider the function

$$\begin{aligned} \Theta : C_T(\iota(\mathcal{P}), D_\infty) &\to C_T(\mathcal{P}(\mathbb{R}^d)) \\ \tilde{\gamma} \quad &\mapsto \quad \mu_t := \iota^{-1}(\tilde{\gamma}(t)). \end{aligned} \tag{4.15}$$

It is well-defined, and thanks to the previous considerations we are allowed to define $\Lambda := \Theta_\sharp \mathtt{L}$, because $\mathtt{L}$ is concentrated over curves that are absolutely continuous with respect to D_∞, and in particular on curves that are continuous with respect to it.

By the properties of $\mathtt{L}$ and the fact that $e_t \circ \Theta(\tilde{\gamma}) = \iota^{-1}(\tilde{\gamma}(t))$, it is straightforward that $(e_t)_\sharp \Lambda = M_t$, so that we can apply Lemma 4.8 and then for Λ-a.e. $(\mu_t)_{t\in[0,T]}$ it holds

$$\frac{d}{dt} \int_X \phi_k(x) d\mu_t(x) = \frac{d}{dt} L_{\phi_k}(\mu_t) = B_t[L_{\phi_k}](\mu_t) \tag{4.16}$$

in the sense of distribution in $(0,T)$, for all $k \in \mathbb{N}$.

By the density of $\mathrm{Span}(\mathcal{A})$ in $C_c^1(\mathbb{R}^d)$, it is immediate to prove that, if a curve $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]}$ satisfies (4.16), then (4.11) holds. □

4.2. **Correspondence between AC curves and solution to the continuity equation.** In this subsection we find a natural correspondence between solutions to the continuity equation associated with a family of L^p-derivations, and curves in $AC_T^p(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$.

Proposition 4.10 (From CE to AC). *In the setting of Theorem 4.9, assume $p \geq 1$, $M_0 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ and*

$$\int_0^T \int_{\mathcal{P}} c_t^p(\mu) dM_t(\mu) dt < +\infty, \tag{4.17}$$

i.e. $(B_t)_{t\in[0,T]}$ is a family of $L^p(M_t)$ derivations. Then, the probability measure Λ given by Theorem 4.9 is concentrated over $\boldsymbol{\mu} \in AC_T^p(\mathcal{P}_p(\mathbb{R}^d))$ and

$$\int_0^T |\dot{\boldsymbol{\mu}}|_{W_p}^p(t) dt \leq \int_0^T c_t^p(\mu_t) dt < +\infty \quad \text{for } \Lambda\text{-a.e. } \boldsymbol{\mu}. \tag{4.18}$$

In particular, $\Lambda \in \mathcal{P}_{\bar{\mathcal{A}}_p}(C_T(\mathcal{P}(\mathbb{R}^d)))$ and $\boldsymbol{M} \in AC_T^p(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$.

Proof. Thanks to the hypothesis $M_0 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$, we already know that

$$\int W_p^p(\mu_0, \delta_{\bar{x}}) d\Lambda(\boldsymbol{\mu}) < +\infty.$$

Moreover, thanks to Lemma 4.8 and Fubini's theorem, it holds that

$$\int_0^T \int c_t^p(\mu_t) d\Lambda(\boldsymbol{\mu}) dt = \int_0^T \int_{\mathcal{P}} c_t^p(\mu) dM_t(\mu) dt < +\infty,$$

so it is left to prove that Λ-a.e. $\boldsymbol{\mu}$ is in $AC_T(\mathcal{P}_p(\mathbb{R}^d))$ and (4.18) holds.

Case $p = 1$: let $\phi \in C_b^1(\mathbb{R}^d)$ with $\|\nabla\phi\|_\infty \leq 1$. Then for any $s < t$ it holds

$$\int_{\mathbb{R}^d} \phi d(\mu_t - \mu_s) \leq \int_s^t B_r[L_\phi](\mu_r) dr \leq \int_s^t c_r(\mu_r) dr,$$

so that passing to the supremum w.r.t. ϕ on the left hand side, we have $W_1(\mu_s, \mu_t) \leq \int_s^t c_r(\mu_r) dr$, which implies (4.18). Then, $\Lambda \in \mathcal{P}_{\overline{\mathcal{A}}_1}(C_T(\mathcal{P}(\mathbb{R}^d)))$ by definition and $\boldsymbol{M} \in AC_T^1(\mathcal{P}_1(\mathcal{P}_1(\mathbb{R}^d)))$ easily follows.

Case $p > 1$: for Λ-a.e. $\boldsymbol{\mu}$, we find a vector field $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ for which $\partial_t\mu_t + \mathrm{div}(v_t\mu_t) = 0$ is satisfied. We know that Λ-a.e. $\boldsymbol{\mu}$ solves the continuity equation $\partial_t\mu_t + \mathrm{div}(B_t\mu_t) = 0$, in the sense that for all $\psi \in C_c^1(0,T)$ and $\phi \in C_c^1(\mathbb{R}^d)$ it holds

$$\int_0^T c_t^p(\mu_t) dt < +\infty \quad \text{and} \quad \int_0^T \psi'(t) \int_{\mathbb{R}^d} \phi(x) d\mu_t(x) dt = -\int_0^T \psi(t) B_t[L_\phi](\mu_t) dt. \tag{4.19}$$

Fix a curve $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d))$ with such properties and let $\mathcal{S}$ be the collection of real valued functions from $[0,T] \times \mathbb{R}^d$ defined as

$$\mathcal{S} := \mathrm{Span}\left(\left\{(t,x) \mapsto \psi(t)\nabla_W F(x,\mu_t) \ : \ \psi \in C^1([0,T]),\ F \in \mathrm{Cyl}_c^1(\mathcal{P}(\mathbb{R}^d))\right\}\right).$$

Define the functional

$$\mathcal{B} : \mathcal{S} \to \mathbb{R}, \quad \mathcal{B}(H) := \sum_{k=1}^n \int_0^T \psi_k(t) B_t[F_k](\mu_t) dt, \tag{4.20}$$

where the general form for H is $H(t,x) = \sum_{k=1}^n \psi_k(t)\nabla_W F_k(x,\mu_t)$. We have:

- $\mathcal{B}$ is linear (due to the linearity of all the B_t's) and well defined, i.e. it does not depend on the representation chosen for $H \in \mathcal{S}$ indeed if

$$H = \sum_{i=1}^m \zeta_i(t)\nabla_W G_i(x,\mu_t) = \sum_{k=1}^n \psi_k(t)\nabla_W F_k(x,\mu_t),$$

this implies that for each $t \in [0,T]$ the Wasserstein gradient of the cylinder function $\mu \mapsto \sum_i \zeta_i(t)G_i(\mu) - \sum_k \psi_k(t)F_k(\mu)$ is null in μ_t. Then by linearity of the integral and of each B_t, it holds

$$\begin{aligned}\left|\sum_{k=1}^n \int_0^T \psi_k(t)B_t[F_k](\mu_t)dt - \sum_{i=1}^m \int_0^T \zeta_i(t)B_t[G_i](\mu_t)dt\right|& \\ = \left|\int_0^T B_t\left[\sum_{i=1}^m \zeta_i(t)G_i - \sum_{k=1}^n \psi_k(t)F_k\right](\mu_t)dt\right|& \\ \le \int_0^T c_t(\mu_t)\left\|\nabla_W\left(\sum_{i=1}^m \zeta_i(t)G_i - \sum_{k=1}^n \psi_k(t)F_k\right)(\cdot,\mu_t)\right\|_{L^{p'}(\mu_t)} dt = 0;&\end{aligned}$$

- for any $H \in \mathcal{S}$, it holds

$$|\mathcal{B}(H)| \le \left(\int_0^T c_t^p(\mu_t)dt\right)^{\frac{1}{p}} \|H\|_{L^{p'}(\mu_t\otimes dt;\mathbb{R}^d)}.$$

This implies that, by Hahn-Banach theorem, $\mathcal{B}$ can be extended to a continuous and linear functional on the set $L^{p'}(\mu_t \otimes dt;\mathbb{R}^d)$ and it can be represented by a Borel measurable vector field $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$. In particular, v satisfies

$$\|v\|_{L^p(\mu_t\otimes dt;\mathbb{R}^d)} \le \left(\int_0^T c_t^p(\mu_t)dt\right)^{\frac{1}{p}} \quad \text{and} \quad \mathcal{B}(H) = \int_0^T \int_{\mathbb{R}^d} v(t,x)\cdot H(t,x)d\mu_t(x)dt, \tag{4.21}$$

for all $H \in \mathcal{S}$. This implies that $\partial_t\mu_t+\mathrm{div}(v_t\mu_t) = 0$, indeed for all $\psi \in C^1_c(0,T)$ and $\phi \in C^1_c(\mathbb{R}^d)$, considering function $H(t,x) := \psi(t)\nabla\phi(x)$ and substituting in (4.21), we have

$$\begin{aligned}\int_0^T \psi'(t)\int_{\mathbb{R}^d} \phi(x)d\mu_t(x)dt &= -\int_0^T \psi(t)B_t[L_\phi](\mu_t)dt \\ &= -\,\mathcal{B}(H) = -\int_0^T \psi(t)\int_{\mathbb{R}^d} v(t,x)\cdot\nabla\phi(x)d\mu_t(x)dt.\end{aligned}$$

Thanks to [AGS08, Theorem 8.3.1], we conclude that $\boldsymbol{\mu} \in AC_T(\mathcal{P}_p(\mathbb{R}^d))$

$$\int_0^T |\dot\mu|^p_{W_p}(t)dt \le \int_0^T\int_{\mathbb{R}^d} |v(t,x)|^p d\mu_t dt \le \int_0^T c_t^p(\mu_t)dt < +\infty.$$

Then, $\Lambda \in \mathcal{P}_{\bar{\mathcal{A}}_p}(C_T(\mathcal{P}(\mathbb{R}^d)))$ and $\boldsymbol{M} \in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$ follow, respectively, from Lemma 4.8 and Proposition 3.8. □

4.2.1. *Non-local vector fields.* Before proceeding, we introduce the notion of *non-local vector field*, and we see how it is connected to the one of derivation. As for derivations, we introduce 'L^p-non-local vector fields'. We will use the notation $\widetilde{M}$ and $\widetilde{M_t} \otimes dt$ introduced in Remark 2.6 and §2.3.1.

Definition 4.11 (L^p-non-local vector fields)**.** *Let $M \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ and $p \ge 1$. We say that $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ is an $L^p(\widetilde{M})$-non-local vector field if*

$$\int_{\mathcal{P}}\int_{\mathbb{R}^d} |b(x,\mu)|^p d\mu(x)dM(\mu) < +\infty. \tag{4.22}$$

As for derivations, we will often work with a family of non-local vector field, indexed by time $t \in [0,T]$. In particular, given a Borel family of random measures $(M_t)_{t\in[0,T]} \subset \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, an $L^p(\widetilde{M}_t \otimes dt)$-non-local vector field is a Borel measurable function $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that

$$\int_0^T \int_{\mathcal{P}} \int_{\mathbb{R}^d} |b(t,x,\mu)|^p d\mu(x) dM_t(\mu) dt < +\infty. \tag{4.23}$$

Definition 4.12. *Let $(M_t)_{t\in[0,T]} \subset \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ be a curve of random measures and $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ an $L^p(\widetilde{M}_t \otimes dt)$ non-local vector field. We say that $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ holds, if for all $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ it holds*

$$\frac{d}{dt} \int_{\mathcal{P}(\mathbb{R}^d)} F(\mu) dM_t(\mu) = \int_{\mathcal{P}} \int_{\mathbb{R}^d} \nabla_W F(x,\mu) \cdot b_t(x,\mu) d\mu(x) dM_t(\mu), \tag{4.24}$$

in the sense of distribution in $(0,T)$.

Remark 4.13. *Notice that a non-local vector field always induces a derivation. To be more specific, let $M \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ and $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be an $L^p(\widetilde{M})$ non-local vector field, then for M-a.e. $\mu \in \mathcal{P}(\mathbb{R}^d)$, the following quantity defines an $L^p(M)$-derivation*

$$B^{(b)}[F](\mu) := \int_{\mathbb{R}^d} b(x,\mu) \cdot \nabla_W F(x,\mu) d\mu(x) \tag{4.25}$$

where a feasible c is given by

$$c^{(b)}(\mu) = \|b(\cdot,\mu)\|_{L^p(\mu)}.$$

The same relation holds between $L^p(\widetilde{M}_t \otimes dt)$-non-local vector field and $L^p(M_t \otimes dt)$-derivations. In particular, in the context of Definition 4.12, if $p \geq 1$ and $M_0 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$, then thanks to Proposition 4.10, it always holds that

$$|\dot{M}|^p_{\mathcal{W}_p}(t) \leq \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |b_t(x,\mu)|^p d\mu(x) dM_t(\mu) < +\infty \quad \text{for a.e. } t \in [0,T]. \tag{4.26}$$

Lemma 4.14. *Let $(M_t)_{t\in[0,T]} \subset \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ be a curve of random measures and $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ an $L^p(\widetilde{M}_t \otimes dt)$ non-local vector fields, with $p \geq 1$, satisfying $\partial M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$. Then, (4.24) is satisfied also for $F \in \mathrm{Cyl}^1_b(\mathcal{P}(\mathbb{R}^d))$.*

Proof. Let $\Psi \in C^1_b(\mathbb{R}^k)$ and $\phi_i \in C^1_b(\mathbb{R}^d)$ for $i \leq k$. Consider a cut-off function $\rho \in C^1_c(\mathbb{R}^d)$ such that $0 \leq \rho \leq 1$, $\rho(x) = 1$ for all $|x| \leq 1$ and $\rho = 0$ for all $|x| \geq 2$. Then, for all $R > 1$ define $\rho_R(x) := \rho(x/R)$, $\phi_{i,R}(x) = \phi_i(\rho_R(x))$ and $F_R = \Psi(L_{\phi_{1,R}}, \dots, L_{\phi_{k,R}}) \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$. We know that for all $\xi \in C^1_c(0,T)$ it holds

$$\int_0^T \xi'(t) \int_{\mathcal{P}(\mathbb{R}^d)} F_R(\mu) dM_t(\mu) dt = \int_0^T \xi(t) \int_{\mathcal{P}} \int_{\mathbb{R}^d} b_t(x,\mu) \cdot \nabla_W F_R(x,\mu) d\mu(x) dM_t(\mu) dt,$$

so we want to pass to the limit on both sides as $R \to +\infty$. Regarding the LHS:

$$\left| \int_0^T \xi'(t) \int_{\mathcal{P}(\mathbb{R}^d)} F_R(\mu) dM_t(\mu) dt - \int_0^T \xi'(t) \int_{\mathcal{P}(\mathbb{R}^d)} F(\mu) dM_t(\mu) dt \right|$$
$$\leq \|\xi'\|_\infty \|\nabla\Psi\|_\infty \sum_{i=1}^k \int_0^T \int_{\mathcal{P}} \int_{\mathbb{R}^d} |\phi_{i,R}(x) - \phi_i(x)| d\mu(x) dM_t(\mu) dt \to 0,$$

thanks to dominated convergence theorem. Regarding the RHS:

$$\left|\int_0^T \xi(t)\int_{\mathcal{P}}\int_{\mathbb{R}^d} b_t(x,\mu)\cdot\big(\nabla_W F_R(x,\mu)-\nabla_W F(x,\mu)\big)d\mu(x)dM_t(\mu)dt\right|\to 0$$

again by dominated convergence theorem, since $\nabla_W F_R(x,\mu)\to\nabla_W F(x,\mu)$ pointwise in (x,μ), and the domination is given by $2\|\xi\|_\infty\sum_{i=1}^k\|\partial_i\Psi\|_\infty\|\nabla\phi_i\|_\infty\|b_t(\cdot,\mu)\|_{L^p(\mu)}$. □

Example 4.15. *The curve $(M_t)_{t\in[0,T]}$ introduced in (1.7) solves the continuity equation $\partial M_t+\mathrm{div}_{\mathcal{P}}(b_tM_t)=0$, with b as in (1.8). Indeed, using the notation $\underline{x}=(x_1,\dots,x_N)\in(\mathbb{R}^d)^N$ for all $F=\Psi\circ L_\Phi\in\mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ we have*

$$\begin{aligned}\frac{d}{dt}\int_{\mathcal{P}(\mathbb{R}^d)}F(\mu)dM_t(\mu)=&\frac{d}{dt}\int_{\mathbb{R}^{dN}}\Psi\left(\frac1N\sum_{i=1}^N\phi_1(x_i),\dots,\frac1N\sum_{i=1}^N\phi_k(x_i)\right)dm_t(\underline{x})\\ =&\frac1N\sum_{j=1}^k\int_{\mathbb{R}^{dN}}\sum_{i=1}^N\partial_j\Psi\left(L_\Phi(\iota(\underline{x}))\right)\nabla\phi_j(x_i)\cdot b_t(x_i,\underline{x})dm_t(\underline{x})\\ =&\sum_{j=1}^k\int_{\mathbb{R}^{dN}}\int_{\mathbb{R}^d}\partial_j\Psi(L_\Phi(\iota(\underline{x})))\nabla\phi_j(x)\cdot b_t(x,\iota(\underline{x}))d[\iota(\underline{x})](x)dm_t(\underline{x})\\ =&\int_{\mathcal{P}(\mathbb{R}^d)}\int_{\mathbb{R}^d}\nabla_W F(x,\mu)\cdot b_t(x,\mu)d\mu(x)dM_t(\mu).\end{aligned}$$

Consider now an absolutely continuous curve of random measures $\boldsymbol{M}\in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$, our goal is to build a non-local vector field such that the curve solves the continuity equation associated to it. Before proceeding, we need to define some useful objects in the following:

- given a curve $\boldsymbol{M}\in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$, define $\Xi^{\boldsymbol{M}}\in\mathcal{M}_+([0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d))$ such that for all $F:[0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to[0,1]$ Borel measurable, it holds

$$\int F(t,x,\mu)d\Xi^{\boldsymbol{M}}(t,x,\mu)=\int_0^T\int_{\mathcal{P}}\int_{\mathbb{R}^d}F(t,x,\mu)d\mu(x)dM_t(\mu)dt. \tag{4.27}$$

  It coincides with the measure already indicated as $\widetilde{M_t}\otimes dt$, we just use this in some contexts for the sake of notation;
- given a measure $\mathfrak{L}\in\mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$, define the measure $\Xi^{\mathfrak{L}}\in\mathcal{M}_+([0,T]\times C_T(\mathbb{R}^d)\times\mathcal{P}(C_T(\mathbb{R}^d)))$ such that for all $H:[0,T]\times C_T(\mathbb{R}^d)\times\mathcal{P}(C_T(\mathbb{R}^d))\to[0,1]$ Borel measurable, it holds

$$\int H(t,\gamma,\lambda)d\Xi^{\mathfrak{L}}(t,\gamma,\lambda)=\int_0^T\int\int H(t,\gamma,\lambda)d\lambda(\gamma)d\mathfrak{L}(\lambda)dt. \tag{4.28}$$

Proposition 4.16 (From AC to CE)**.** *Let $\boldsymbol{M}=(M_t)_{t\in[0,T]}\in AC^p_T(\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$ for some $p>1$. Then, there exists an $L^p(\widetilde{M_t}\otimes dt)$ non-local vector field $b:[0,T],\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}^d$ such that*

$$\int_{\mathcal{P}}\int_{\mathbb{R}^d}|b(t,x,\mu)|^pd\mu(x)dM_t(\mu)=|\dot{M}|^p_{\mathcal{W}_p}(t)\quad\text{for a.e. } t\in[0,T], \tag{4.29}$$

and satisfying the continuity equation $\partial M_t+\mathrm{div}_{\mathcal{P}}(b_tM_t)=0$, in the sense of (4.24).

Proof. Using the results of Section 3, let $\Lambda_{\boldsymbol{M}} \in \mathrm{Lift}(\boldsymbol{M})$ and $\mathfrak{L} := G_\sharp(\Lambda_{\boldsymbol{M}}) \in \mathcal{P}_{\bar{\mathfrak{A}}_p}(\mathcal{P}(C_T(\mathbb{R}^d)))$, and consider the measures $\Xi^{\boldsymbol{M}}$ and $\Xi^{\mathfrak{L}}$ defined above. Thanks to propositions 3.7 and 3.9, we have

$$\int_0^T \int \int |\dot{\gamma}|^p(t) d\lambda(\gamma) d\mathfrak{L}(\lambda) dt = \int_0^T |\dot{M}|^p_{\mathcal{W}_p}(t) dt < +\infty,$$

which implies that the map $(t, \gamma, \lambda) \mapsto D(t, \gamma) := \dot{\gamma}(t)$ is in $L^p(\Xi^{\mathfrak{L}}; \mathbb{R}^d)$ (see Lemma C.2). Consider the map

$$\begin{aligned} \mathcal{E} : [0,T] \times C_T(\mathbb{R}^d) \times \mathcal{P}(C_T(\mathbb{R}^d)) &\to [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \\ (t, \gamma, \lambda) &\mapsto (t, \gamma(t), (e_t)_\sharp \lambda), \end{aligned} \tag{4.30}$$

and notice that $\Xi^{\boldsymbol{M}} = \mathcal{E}_\sharp \Xi^{\mathfrak{L}}$. Then, thanks to [ABS24, Lemma 17.3] (see also Remark 5.11), there exists a function $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that

$$\mathcal{E}_\sharp(D\ \Xi^{\mathfrak{L}}) = b\ \Xi^{\boldsymbol{M}}. \tag{4.31}$$

We show that b is the non-local vector field we are looking for: indeed

$$\int_0^T \int_{\mathcal{P}} \int_{\mathbb{R}^d} |b(t,x,\mu)|^p d\mu(x) dM_t(\mu) dt = \int |b(t,x,\mu)|^p d\Xi^{\boldsymbol{M}}(t,x,\mu) dt \le \int |\dot{\gamma}(t)|^p d\Xi^{\mathfrak{L}}(t,\gamma,\lambda) < +\infty, \tag{4.32}$$

again thanks to [ABS24, Lemma 17.3]. Moreover, for any $\psi \in C^1_c(0,T)$ and $F = \Psi \circ L_\Phi \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$, it holds

$$\begin{aligned} \int_0^T \psi'(t) \int_{\mathcal{P}} F(\mu) dM_t(\mu) dt &= \int \psi'(t) F(\mu) d\Xi^{\boldsymbol{M}}(t,x,\mu) = \int \psi'(t) F((e_t)_\sharp \lambda) d\Xi^{\mathfrak{L}}(t,\gamma,\lambda) \\ &= \int \int \int_0^T \psi'(t) \Psi \left(\int \phi_1(\gamma(t)) d\lambda(\gamma), \dots, \int \phi_k(\gamma(t)) d\lambda(\gamma) \right) dt d\lambda(\gamma) d\mathfrak{L}(\lambda) \\ &= - \int \int \int_0^T \psi(t) \sum_{i=1}^k \partial_i \Psi(L_\Phi((e_t)_\sharp \lambda)) \nabla \phi_i(\gamma(t)) \cdot \dot{\gamma}(t)\ dt d\lambda(\gamma) d\mathfrak{L}(\lambda) \\ &= - \int \psi(t) \nabla_W F\big(\gamma(t), (e_t)_\sharp \lambda\big) \cdot d\big(D\ \Xi^{\mathfrak{L}}\big)(t,\gamma,\lambda) \\ &= - \int \psi(t) \nabla_W F(x,\mu) \cdot d\big(b\ \Xi^{\boldsymbol{M}}\big)(t,x,\mu) \\ &= - \int_0^T \psi(t) \int_{\mathcal{P}} \int_{\mathbb{R}^d} \nabla_W F(x,\mu) \cdot b(t,x,\mu) d\mu(x) dM_t(\mu) dt, \end{aligned}$$

where in the second last equality we exploited the definition of $\mathcal{E}$ and the characterization of b given by (4.31). Putting together (4.26), (4.32) and Proposition (3.6), then (4.29) follows. □

Remark 4.17. *Notice that, thanks to* (4.29), *the vector field we built is minimal in an L^p-sense, and because of the strict convexity of $|\cdot|^p$ for $p > 1$, such vector field is unique, in the sense that any other $L^p(M_t \otimes dt)$-non-local vector field $\tilde{b}$ satisfying $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(\tilde{b}_t M_t) = 0$ and* (4.29) *coincides with b for $\Xi^{\boldsymbol{M}}$-a.e. $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$.*

A consequence of the characterization of absolutely continuous curves with curves that solve a continuity equation, is a Benamou-Brenier-type formula for random measures.

Theorem 4.18 (Benamou-Brenier formula)**.** *Let $p > 1$. For all $M_0, M_1 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ it holds*

$$\mathcal{W}_p^p(M_0, M_1) = \min\left\{ \int_0^1 \int_{\mathcal{P}} \int_{\mathbb{R}^d} |b_t(x,\mu)|^p d\mu(x) dM_t(\mu) dt \,:\, \partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0 \right\}. \tag{4.33}$$

Proof. Thanks to (4.26), all the competitors for the right-hand side satisfy the inequality

$$\mathcal{W}_p^p(M_0, M_1) \le \int_0^1 \int_{\mathcal{P}} \int_{\mathbb{R}^d} |b_t(x,\mu)|^p d\mu(x) dM_t(\mu) dt.$$

On the other hand, Lemma 3.10 gives the existence of a constant speed geodesic $(M_t)_{t\in[0,1]} \in C([0,1], \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$. Then, Proposition 4.16 gives the existence of a non-local vector field $b : [0,1] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ satisfying $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ and (4.29), from which it follows that the curve (M_t) and the non-local vector field b are optimal for (4.33). □

4.3. **The tangent and cotangent bundle to $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$.** In this subsection, we define the tangent and cotangent bundle as closure in a suitable Lebesgue space of the Wasserstein gradient of cylinder functions. Then, following the same argument of [AGS08, §8.4] we characterize the non-local vector fields of minimal energy (see Remark 4.17) as elements of the tangent bundle.

Before proceeding, let us recall the duality pairing map between Lebesgue spaces: given any measurable space $(X, \mathcal{F})$ endowed with a finite positive measure σ, then for any $p \in (1, +\infty)$ the duality pairing is defined as

$$j_p : L^p(\sigma; \mathbb{R}^d) \to L^{p'}(\sigma; \mathbb{R}^d), \quad j_p(V)(x) := \begin{cases} |V(x)|^{p-2} V(x) & \text{if } V(x) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.34}$$

Definition 4.19. *Let $p \in (1, +\infty)$ and $M \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ and recall the definition of $\widetilde{M} \in \mathcal{P}(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d))$ from Remark 2.6. Then we define, respectively, the cotangent and the tangent space of $\mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ at M as*

$$\mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)) := \mathrm{Clos}_{L^{p'}(\widetilde{M};\mathbb{R}^d)} \left\{ \nabla_W F \;:\; F \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d)) \right\} \subseteq L^{p'}(\widetilde{M}; \mathbb{R}^d); \tag{4.35}$$

$$\mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)) := \mathrm{Clos}_{L^{p}(\widetilde{M};\mathbb{R}^d)} \left\{ j_{p'}(\nabla_W F) \;:\; F \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d)) \right\} \subseteq L^{p}(\widetilde{M}; \mathbb{R}^d). \tag{4.36}$$

Notice that the tangent and the cotangent space are in duality by the maps j_p and $j_{p'}$, i.e. $\mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)) = j_{p'}(\mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$ and $\mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)) = j_p(\mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)))$.

Remark 4.20. *The tangent space could be defined only considering infinitely-smooth cylinder functions, that is*

$$\mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d)) = \mathrm{Clos}_{L^{p}(\widetilde{M};\mathbb{R}^d)} \left\{ j_{p'}(\nabla_W(\Psi \circ L_\Phi)) : k \in \mathbb{N},\, \Psi \in C_c^\infty(\mathbb{R}^k),\, \Phi \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^k) \right\},$$

since any C_c^1 function can be uniformly approximated by functions in C_c^∞.

Lemma 4.21. *Let $p > 1$, $M \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ and $b \in L^p(\widetilde{M}; \mathbb{R}^d)$. Then $b \in \mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ if and only if $\|b + b'\|_{L^p(\widetilde{M};\mathbb{R}^d)} \ge \|b\|_{L^p(\widetilde{M};\mathbb{R}^d)}$ for all $b' \in L^p(\widetilde{M}; \mathbb{R}^d)$ such that $\langle b', \omega \rangle = 0$ for all $\omega \in \mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$.*
In particular, for every $b \in L^p(\widetilde{M}; \mathbb{R}^d)$ there exists a unique element $\Pi(b) \in \mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ in the set of vector fields $b' \in L^p(\widetilde{M}; \mathbb{R}^d)$ satisfying $\langle b, \omega \rangle = \langle b', \omega \rangle$ for all $\omega \in \mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ and $\Pi(b)$ is the element of minimal norm in this class.

Notice that the condition that $\langle b', \omega\rangle = 0$ for all $\omega \in \mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ is equivalent to ask that $\langle b', \nabla_W F\rangle = 0$ for all $F \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d))$, which, consistently with our notation, can be written in the compact form $\mathrm{div}_{\mathcal{P}}(b'M) = 0$.

Proof. As in [AGS08, Lemma 8.4.2], by convexity of the L^p-norm to the power p and the fact that $pj_p(b)$ belongs to its subdifferential at the function b, we have that $\|b + b'\|^p_{L^p} \geq \|b\|_{L^p}$ for all b' satisfying $\mathrm{div}_{\mathcal{P}}(b'M) = 0$ if and only if $\langle j_p(b), b'\rangle = 0$ for all b' as before, and by the Hahn-Banach theorem this happens if and only if $j_p(b) \in \mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$. This is equivalent to say $b = j_{p'}(j_p(b)) \in \mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$. The last part follows from the fact that the class of vector fields b' satisfying $\langle b, \omega\rangle = \langle b', \omega\rangle$ for all $\omega \in \mathrm{CoTan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ is closed and convex, so that by strict convexity of the L^p-norm there exists a unique element of minimum norm in it and by the previous characterization it belongs to $\mathrm{Tan}_M \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$. □

Proposition 4.22. *Let* $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ *such that* $M_0 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ *and* $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ *be an* $L^p(\widetilde{M}_t \otimes dt)$*-non-local vector field. Assume that the continuity equation* $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ *holds, in the sense of Definition 4.12. Then* $M_t \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ *for all* $t \in [0, T]$ *and the following are equivalent:*

(1) $b_t \in \mathrm{Tan}_{M_t} \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ *for a.e.* $t \in [0, T]$*;*
(2) $\int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |b_t(x, \mu)|^p d\mu(x)dM_t(\mu) \leq |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t)$ *for a.e.* $t \in [0, T]$*;*
(3) $\int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |b_t(x, \mu)|^p d\mu(x)dM_t(\mu)dt = \int_0^T |\dot{\boldsymbol{M}}|^p_{\mathcal{W}_p}(t)dt.$

Proof. The equivalence between (2) and (3) follows from (4.26). Regarding (1) $\Longrightarrow$ (2), consider $\tilde{b}$ be the $L^p(\widetilde{M}_t \otimes dt)$-non-local vector field given by Proposition 4.16. The goal is to show that $b_t = \tilde{b}_t$ for a.e. $t \in [0, T]$, as functions of $L^p(\widetilde{M}_t)$. By (4.29) and (4.26), it holds that

$$\int_{\mathcal{P}} \int_{\mathbb{R}^d} |\tilde{b}_t(x, \mu)|^p d\mu(x)dM_t(\mu) \leq \int_{\mathcal{P}} \int_{\mathbb{R}^d} |b_t(x, \mu)|^p d\mu(x)dM_t(\mu) \quad \text{for a.e. } t \in [0, T]. \tag{4.37}$$

Moreover, the curve $(M_t)_{t\in[0,T]}$, by assumption and by construction of $\tilde{b}$, satisfies the continuity equations $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ and $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(\tilde{b}_t M_t) = 0$, which implies that for all $\xi \in C^1_c(0, T)$ and $F \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d))$ it holds

$$\int_0^T \xi(t) \int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} b_t \cdot \nabla_W F d\widetilde{M}_t \, dt = \int_0^T \xi(t) \int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} \tilde{b}_t \cdot \nabla_W F d\widetilde{M}_t \, dt.$$

Localizing this equality in time, it holds

$$\int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} b_t \cdot \nabla_W F d\widetilde{M}_t = \int_{\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)} \tilde{b}_t \cdot \nabla_W F d\widetilde{M}_t \quad \text{for a.e. } t \in [0, T],$$

and together with (4.37), since $b_t \in \mathrm{Tan}_{M_t} \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ for a.e. $t \in [0, T]$, we can apply Lemma 4.21 to conclude that $b_t(x, \mu) = \tilde{b}_t(x, \mu)$ for $\widetilde{M}_t$-a.e. (x, μ) and for a.e. $t \in [0, T]$.
Regarding (2) $\Longrightarrow$ (1), let us introduce the auxiliar space

$$\mathcal{V} := \mathrm{Clos}_{L^p(\Xi^{\boldsymbol{M}};\mathbb{R}^d)} \mathrm{Span}\left\{ j_p\left(\xi(t)\nabla_W F(x, \mu)\right) \ : \ \xi \in C^1_c(0, T), \ F \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d)) \right\}.$$

Following the argument of Lemma 4.21, it is not hard to prove that a non-local vector field $v : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ belongs to $\mathcal{V}$ if and only if $\|v + v'\|_{L^p} \geq \|v\|_{L^p}$ for all $v' \in L^p(\Xi^{\boldsymbol{M}}; \mathbb{R}^d)$ satisfying

$$\int_0^T \xi(t) \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} \nabla_W F(x, \mu) \cdot v'(t, x, \mu)d\mu(x)dM_t(\mu)dt = 0.$$

Thus, if we assume condition (2), because of its equivalence to (3) and the minimality given by Remark 4.17, it holds $b \in \mathcal{V}$. Thus we conclude proving that $v \in \mathcal{V} \implies v_t \in \mathrm{Tan}_{M_t} \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ for a.e. $t \in [0,T]$. This easily follows by a pointwise argument: fix two sequences of functions $\xi_n \in C^1_c(0,T)$ and $F_n \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d))$ such that $j_p(\xi_n \nabla_W F_n) \to v$ in $L^p(\Xi^{\boldsymbol{M}};\mathbb{R}^d)$. Up to consider a subsequence, it holds that

$$j_p(\xi_n \nabla_W F_n)(t,\cdot,\cdot) = |\xi_n(t)|^{p-2}\xi_n(t)|\nabla_W F_n(\cdot,\cdot)|^{p-2}\nabla_W F_n(\cdot,\cdot) \to v(t,\cdot,\cdot) \quad \text{ in } L^p(\widetilde{M}_t),$$

for a.e. $t \in [0,T]$. In particular, fixing a time $t \in [0,T]$ for which the one above holds, the sequence of cylinder functions $F_{t,n} \in \mathrm{Cyl}_c(\mathcal{P}(\mathbb{R}^d))$ defined by $F_{t,n}(x,\mu) := \xi_n(t)F_n(x,\mu)$ is such that $j_p(\nabla_W F_{t,n}) \to v_t$ in $L^p(\widetilde{M}_t)$. By definition of tangent space, this proves that if $v \in \mathcal{V}$, then $v_t \in \mathrm{Tan}_{M_t} \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ for a.e. $t \in [0,T]$. □

4.4. **Derivations and vector fields.** In this subsection, we show that any family of derivations $(B_t)_{t\in[0,T]}$ is induced by a family of non-local vector fields, as in Remark 4.11, whenever the continuity equation $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(B_t M_t) = 0$ is satisfied.

Theorem 4.23. *Let $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ and $(B_t)_{t\in[0,T]}$ an $L^p(M_t \otimes dt)$-derivation, such that $\partial M_t + \mathrm{div}_{\mathcal{P}}(B_t M_t) = 0$. Then there exists an $L^p(\widetilde{M}_t \otimes dt)$-non-local vector field $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ such that for $M_t \otimes dt$-a.e. (μ,t) it holds*

$$B_t[F](\mu) = \int_{\mathbb{R}^d} b(t,x,\mu)\cdot \nabla_W F(x,\mu)d\mu(x), \quad \forall F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)). \tag{4.38}$$

The proof of this result is very similar to the one of Proposition 4.16. Indeed, putting together Proposition 4.10 and Proposition 4.16, we have a vector field $v : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that for any $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ and for a.a. $t \in (0,T)$, it holds

$$\int_{\mathcal{P}}\int_{\mathbb{R}^d} v(t,x,\mu)\cdot \nabla_W F(x,\mu)d\mu(x)dM_t(\mu) = \int_{\mathcal{P}} B_t[F](\mu)dM_t(\mu).$$

The non-trivial part is to localize this equality with respect to the variable μ, to prove (4.38). In the proof, we are going to see how this localization can be done in various steps, mainly in steps 2 and 3 below.

Proof of Theorem 4.23. **Step 1**: *superposition and nested metric lifting.* Using first Proposition 4.10 and then Theorem 3.4, we obtain a probability measure $\mathfrak{L} \in \mathcal{P}_{\mathfrak{A}_p}(\mathcal{P}(C_T(\mathbb{R}^d)))$ that satisfies:

(i) $(E_t)_\sharp \mathfrak{L} = M_t$;
(ii) $\mathfrak{L}$-a.e. $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ is such that $\mu_t := (e_t)_\sharp \lambda$ solves the continuity equation (4.11), since $E_\sharp \mathfrak{L} = \Lambda$;
(iii) $\int_0^T \int\int |\dot\gamma|^p(t) d\lambda(\gamma)d\mathfrak{L}(\lambda)dt < +\infty$.

Consider its associated measure $\Xi^{\mathfrak{L}} \in \mathcal{M}_+([0,T] \times C_T(\mathbb{R}^d) \times \mathcal{P}(C_T(\mathbb{R}^d)))$ as in (4.28).
Step 2: *localization step.* Consider the function $\mathcal{F} : (t,\gamma,\lambda) \mapsto (t,\gamma(t),\lambda) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}(C_T(\mathbb{R}^d))$, and notice that $\mathcal{F}_\sharp \Xi^{\mathfrak{L}} = \Xi^{\mathfrak{L},\boldsymbol{M}}$, where $\Xi^{\mathfrak{L},\boldsymbol{M}} \in \mathcal{M}_+\big([0,T] \times \mathbb{R}^d \times \mathcal{P}(C_T(\mathbb{R}^d))\big)$ is defined such that for each $G : [0,T] \times \mathbb{R}^d \times \mathcal{P}(C_T(\mathbb{R}^d)) \to [0,1]$ Borel measurable it holds

$$\int G(t,x,\lambda)d\Xi^{\mathfrak{L},\boldsymbol{M}}(t,x,\lambda) = \int_0^T \int\int G(t,\gamma(t),\lambda)d\lambda(\gamma)d\mathcal{L}(\lambda)dt.$$

Since the function $\dot\gamma \in L^p(\Xi^{\mathfrak{L}})$, then thanks to [ABS24, Lemma 17.3], it holds that there exists a function $\tilde b \in L^p(\Xi^{\mathfrak{L},\boldsymbol{M}};\mathbb{R}^d)$ such that

$$\mathcal{F}_\sharp(D\ \Xi^{\mathfrak{L}})(dt,dx,d\lambda) = \tilde b(t,x,\lambda)\ \Xi^{\mathfrak{L},\boldsymbol{M}}(dt,dx,d\lambda),$$

where $D(t,\gamma,\lambda)=\dot\gamma(t)$ as before.

Claim: for $\mathfrak{L}$-a.e. λ, $\mu_t := (e_t)_\sharp\lambda$ solves $\partial_t\mu_t + \mathrm{div}_x(\tilde b(t,x,\lambda)\mu_t)=0$.
Indeed, for any $F:\mathcal{P}(C_T(\mathbb{R}^d))\to[0,1]$ Borel measurable, $\psi\in C^1_c(0,T)$ and $\phi\in C^1_c(\mathbb{R}^d)$ it holds

$$\begin{aligned}\int F(\lambda)&\left[\int_0^T\psi'(t)\int_{\mathbb{R}^d}\phi(x)d\mu_t(x)dt\right]d\mathfrak{L}(\lambda)=\\
&=\int F(\lambda)\left[\int_0^T\psi'(t)\int_{C_T(\mathbb{R}^d)}\phi(\gamma(t))d\lambda(\gamma)dt\right]d\mathfrak{L}(\lambda)=\\
&=-\int F(\lambda)\left[\int_0^T\psi(t)\int_{C_T(\mathbb{R}^d)}\nabla\phi(\gamma(t))\cdot\dot\gamma(t)d\lambda(\gamma)dt\right]d\mathfrak{L}(\lambda)\\
&=-\int F(\lambda)\psi(t)\nabla\phi(\gamma(t))\cdot d\left(D\Xi^{\mathfrak{L}}\right)(t,\gamma,\lambda)\\
&=-\int F(\lambda)\psi(t)\nabla\phi(x)\cdot\tilde b(t,x,\lambda)d\Xi^{\mathfrak{L},\boldsymbol{M}}(t,x,\lambda)\\
&=\int F(\lambda)\left[-\int_0^T\psi(t)\int_{\mathbb{R}^d}\nabla\phi(x)\cdot\tilde b(t,x,\lambda)d\mu_t(x)dt\right]d\mathfrak{L}(\lambda).\end{aligned}$$

Step 3: *definition of the non-local vector field.* Define the continuous map $\mathcal{G}:[0,T]\times\mathbb{R}^d\times\mathcal{P}(C_T(\mathbb{R}^d))\to[0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)$ as

$$\mathcal{G}(t,x,\lambda)=(t,x,(e_t)_\sharp\lambda),$$

and notice that $\mathcal{G}_\sharp\Xi^{\mathfrak{L},M}:=\Xi^{\boldsymbol{M}}$. At this point, consider the disintegration of $\Xi^{\mathfrak{L},\boldsymbol{M}}$ w.r.t. $\mathcal{G}$, i.e. the Borel map $(t,x,\mu)\mapsto\Xi^{\mathfrak{L},\boldsymbol{M}}_{t,x,\mu}$, that is well-defined $\Xi^{\boldsymbol{M}}$-almost everywhere. Then define the non-local vector field as

$$b(t,x,\mu)=\int\tilde b(s,z,\lambda)d\Xi^{\mathfrak{L},\boldsymbol{M}}_{t,x,\mu}(s,z,\lambda),$$

which is measurable thanks to the measurability of $\tilde b$ and of the disintegration $(t,x,\mu)\mapsto\Xi^{\mathfrak{L},\boldsymbol{M}}_{t,x,\mu}$ (see Lemma D.1). By construction, it's easy to verify that $b\in L^p(\Xi^{\boldsymbol{M}};\mathbb{R}^d)$.

Step 4: *representation for the disintegration* $\Xi^{\mathfrak{L},\boldsymbol{M}}_{t,x,\mu}$. For any $t\in[0,T]$, disintegrate the measure $\mathfrak{L}$ with respect to the map $E_t=(e_t)_\sharp$, to obtain that there exists a family of probability measures $\{\mathfrak{L}_{t,\mu}\}_{\mu\in\mathcal{P}(\mathbb{R}^d)}\subset\mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ such that

$$\mathfrak{L}=\int_{\mathcal{P}(\mathbb{R}^d)}\mathfrak{L}_{t,\mu}dM_t(\mu).$$

Then $\Xi^{\mathfrak{L},\boldsymbol{M}}_{t,x,\mu}=\delta_t\otimes\delta_x\otimes\mathfrak{L}_{t,\mu}$ for $\Xi^{\boldsymbol{M}}$-a.e. (t,x,μ). Indeed for any $G:[0,T]\times\mathbb{R}^d\times\mathcal{P}(C_T(\mathbb{R}^d))\to[0,+\infty]$ Borel measurable map, we have

$$\begin{aligned}\int G d\Xi^{\mathfrak{L},\boldsymbol{M}}&=\int_0^T\int\left(\int G(t,x,\lambda)d\big((e_t)_\sharp\lambda\big)(x)\right)d\mathfrak{L}(\lambda)dt\\
&=\int_0^T\left[\int_{\mathcal{P}(\mathbb{R}^d)}\left(\int_{\mathbb{R}^d}\int G(t,x,\lambda)d\mathfrak{L}_{t,\mu}(\lambda)d\mu(x)\right)dM_t(\mu)\right]dt\\
&=\int\left[\int G(t,x,\lambda)d\mathfrak{L}_{t,\mu}(\lambda)\right]d\Xi^{\boldsymbol{M}}(t,x,\mu)\end{aligned}$$

$$= \int \left[\int \int \int G(s,z,\lambda)\ d\delta_t(s) d\delta_x(z) d\mathfrak{L}_{t,\mu}(\lambda) \right] d\Xi^{\boldsymbol{M}}(t,x,\mu),$$

so we conclude by mean of the uniqueness of the disintegration. In particular, for $\Xi^{\boldsymbol{M}}$-a.e. (t,x,μ), it holds

$$b(t,x,\mu) = \int \tilde{b}(t,x,\lambda) d\mathfrak{L}_{t,\mu}(\lambda). \tag{4.39}$$

Step 5: *conclusion.* We are left to prove that the non-local vector field b satisfies (4.38). From Step 2 and the properties of $\mathfrak{L}$, we know that for $\mathfrak{L}$-a.e. λ, for any $\psi \in C_c^1(0,T)$ and $\phi \in C_c^1(\mathbb{R}^d)$, it holds

$$\begin{aligned}\int_0^T \psi(t) B_t[L_\phi]\big((e_t)_\sharp\lambda\big) dt &= -\int_0^T \psi'(t) \int_{\mathbb{R}^d} \phi(x) d\big((e_t)_\sharp\lambda\big)(x) dt \\ &= \int_0^T \psi(t) \int_{\mathbb{R}^d} \tilde{b}(t,x,\lambda) \cdot \nabla\phi(x) d\big((e_t)_\sharp\lambda\big)(x) dt,\end{aligned}$$

which implies, thanks to Lemma 4.7, that for all $F \in \mathrm{Cyl}_c^1(\mathcal{P}(\mathbb{R}^d))$ it holds

$$B_t[F]\big((e_t)_\sharp\lambda\big) = \int_{\mathbb{R}^d} \tilde{b}(t,x,\lambda) \cdot \nabla_W F(x,(e_t)_\sharp\lambda) d\big((e_t)_\sharp\lambda\big)(x),$$

for a.e. $t \in [0,T]$ and $\mathfrak{L}$-a.e. λ. At this point, for a.e. $t \in [0,T]$ and M_t-a.e. $\mu \in \mathcal{P}(\mathbb{R}^d)$, we can integrate both sides w.r.t. $\mathfrak{L}_{t,\mu}$: since $(e_t)_\sharp\lambda = \mu$ for $\mathfrak{L}_{t,\mu}$-a.e. λ, the left hand side is constant, while on the right hand side, we can switch the order of integration to obtain, thanks to (4.39), that

$$B_t[F](\mu) = \int_{\mathbb{R}^d} \left(\int \tilde{b}(t,x,\lambda) d\mathfrak{L}_{t,\mu}(\lambda) \right) \cdot \nabla_W F(x,\mu) d\mu(x) = \int_{\mathbb{R}^d} b(t,x,\mu) \cdot \nabla_W F(x,\mu) d\mu(x),$$

for a.e. $M_t \otimes dt$-a.e. (μ,t) and for all $F \in \mathrm{Cyl}_c^1(\mathcal{P}(\mathbb{R}^d))$. □

5. Nested superposition principle

The main goal of this section is to prove Theorem 1.2. The strategy is similar to the one used in Section 3 to prove Theorem 1.1. In particular, given a Borel measurable non-local vector field $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, the main objects of our study are:

(i) a curve of random measures $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ for which b is an $L^1(\widetilde{M}_t \otimes dt)$-non-local vector field and such that, according to Definition 4.12, it holds

$$\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0; \tag{5.1}$$

(ii) a probability measure $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ satisfying $\int \int_0^T \int |b(t,x,\mu_t)| d\mu_t(x) dt d\Lambda(\boldsymbol{\mu}) < +\infty$ and concentrated over curves of measures $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d))$ that solves

$$\partial_t \mu_t + \mathrm{div}(b_t(\cdot,\mu_t)\mu_t) = 0; \tag{5.2}$$

(iii) a probability measure $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ satisfying $\int \int a_1(\gamma) d\lambda(\gamma) d\mathfrak{L}(\lambda) < +\infty$ (see Definition 2.3) and concentrated over $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ that, in turn, are concentrated over $\gamma \in AC_T(\mathbb{R}^d)$ that are solutions of

$$\dot{\gamma}(t) = b(t,\gamma_t,(e_t)_\sharp\lambda). \tag{5.3}$$

We introduce two sets associated with a generic non-local vector field, that will play a fundamental role in the proof of Theorem 1.2.

Definition 5.1. *Let $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be a Borel map. Then, the set* $\mathrm{CE}(b) \subset C_T(\mathcal{P}(\mathbb{R}^d))$ *of solutions to the continuity equation driven by the non-local vector field b is defined by*

$$\begin{aligned} \mathrm{CE}(b) := \Big\{ (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d)) \ : \ \int\int |b_t(x,\mu_t)| d\mu_t(x) dt < +\infty, \\ \partial_t \mu_t + \mathrm{div}(b_t(\cdot,\mu_t)\mu_t) = 0 \Big\}. \end{aligned} \tag{5.4}$$

The set $\mathrm{SPS}(b) \subset \mathcal{P}(C_T(\mathbb{R}^d))$ *of superposition solutions of the particle systems* (5.3) *is defined by*

$$\begin{aligned} \mathrm{SPS}(b) := \{\lambda \in & \mathcal{P}(C_T(\mathbb{R}^d)) \ : \ \int\int |b_t(\gamma(t),(e_t)_\sharp\lambda)| d\lambda(\gamma) dt < +\infty, \\ & \lambda\big(\mathrm{AC}_T(\mathbb{R}^d)\big) = 1, \ \dot\gamma(t) = b_t(\gamma_t,(e_t)_\sharp\lambda) \ \mathcal{L}^1_T \otimes \lambda\text{-}a.e.\}. \end{aligned} \tag{5.5}$$

Note that the properties of $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ and $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ listed in (ii) and (iii), can be summarized by saying that Λ is concentrated over $\mathrm{CE}(b)$ and $\mathfrak{L}$ is concentrated over $\mathrm{SPS}(b)$. A crucial point will be the Borel measurability of these sets, for which we refer to Proposition 5.7 and Proposition 5.8.

Let us start noticing that the hierarchy described in the introduction between general objects $\boldsymbol{M} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$, $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ and $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$, is preserved when we require them to satisfy the conditions listed above. Before proceeding recall that

$$E : \mathcal{P}(C_T(\mathbb{R}^d)) \to C_T(\mathcal{P}(\mathbb{R}^d)), \quad E(\lambda) := ((e_t)_\sharp\lambda)_{t\in[0,T]};$$

$$\mathfrak{e}_t : C_T(\mathcal{P}(\mathbb{R}^d)) \to \mathcal{P}(\mathbb{R}^d), \quad \mathfrak{e}_t(\boldsymbol{\mu}) := \mu_t; \quad E_t : \mathcal{P}(C_T(\mathbb{R}^d)) \to \mathcal{P}(\mathbb{R}^d), \quad E_t(\lambda) = (e_t)_\sharp\lambda.$$

Proposition 5.2. *If $\lambda \in \mathrm{SPS}(b)$, then $E(\lambda) \in \mathrm{CE}(b)$. In particular, if $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ is concentrated over $\mathrm{SPS}(b)$, then $\Lambda := E_\sharp\mathfrak{L} \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ is concentrated over $\mathrm{CE}(b)$.*

Proof. Let $\lambda \in \mathrm{SPS}(b)$ and $\mu_t := (e_t)_\sharp\lambda$. Then for any $\xi \in C^1_c(0,T)$ and $\phi \in C^1_c(\mathbb{R}^d)$, it holds

$$\begin{aligned} \int_0^T \xi'(t) \int_{\mathbb{R}^d} \phi(x) d\mu_t(x) dt &= \int_0^T \xi'(t) \int \phi(\gamma(t)) d\lambda(\gamma) dt = -\int\int_0^T \xi(t) \nabla\phi(\gamma_t) \cdot \dot\gamma(t) dt d\lambda(\gamma) \\ &= -\int_0^T \xi(t) \int \nabla\phi(\gamma_t) \cdot b(t,\gamma_t,(e_t)_\sharp\lambda) d\lambda(\gamma) dt = -\int_0^T \xi(t) \int_{\mathbb{R}^d} \nabla\phi(x) \cdot b(t,x,\mu_t) d\mu_t(x) dt. \end{aligned}$$

□

Proposition 5.3. *Let $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ be concentrated on $\mathrm{CE}(b)$ and such that*

$$\int\int_0^T \int |b_t(x,\mu_t)| d\mu_t(x) dt d\Lambda(\boldsymbol{\mu}) < +\infty.$$

Then the curve of random measures defined as $M_t := (\mathfrak{e}_t)_\sharp\Lambda$, solves the continuity equation $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$, in the sense of Definition 4.12.

Proof. Let $\xi \in C^1_c(0,T)$ and $F = \Psi(L_\Phi) \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$, then

$$\begin{aligned} \int_0^T \xi'(t) \int_{\mathcal{P}(\mathbb{R}^d)} F(\mu) dM_t(\mu) dt &= \int_0^T \xi'(t) \int F(\mu_t) d\Lambda(\boldsymbol{\mu}) dt \\ &= \int\int_0^T \xi'(t) \psi\left(\int \phi_1 d\mu_t, \dots, \int \phi_k d\mu_t\right) dt d\Lambda(\boldsymbol{\mu}) \end{aligned}$$

$$= -\int\int_0^T \xi(t)\sum_{i=1}^k \partial_i\psi(L_\Phi(\mu_t))\int_{\mathbb{R}^d}\nabla\phi_i(x)\cdot b_t(x,\mu_t)d\mu_t(x)dtd\Lambda(\boldsymbol{\mu})$$
$$= -\int_0^T \xi(t)\int\int_{\mathbb{R}^d}\nabla_W F(x,\mu_t)\cdot b_t(x,\mu_t)d\mu_t(x)d\Lambda(\boldsymbol{\mu})dt$$
$$= -\int_0^T \xi(t)\int_{\mathcal{P}(\mathbb{R}^d)}\int_{\mathbb{R}^d}\nabla_W F(x,\mu)\cdot b_t(x,\mu)d\mu(x)dM_t(\mu)dt.$$

□

Corollary 5.4. *Let $\mathfrak{L}\in\mathcal{P}(\mathcal{P}(AC_T(\mathbb{R}^d)))$ be concentrated on* SPS(b) *and such that*

$$\int\int_0^T\int |b_t(\gamma_t,(e_t)_\sharp\lambda)|d\lambda(\gamma)dtd\mathfrak{L}<+\infty.$$

Then the curve of random measures defined as $M_t := (E_t)_\sharp\mathfrak{L}$, solves $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t)=0$.

Proof. Simply notice that $E_t = \mathfrak{e}_t\circ E$, and we conclude thanks to Propositions 5.2 and 5.3. □

5.1. **Superposition: from $\boldsymbol{M}$ to Λ.** Here, similarly to Section 3, we would like to somehow invert the result of Proposition 5.3, i.e. to define a measure $\Lambda\in\mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ concentrated over CE(b) starting from a curve $\boldsymbol{M}\in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ that solves the continuity equation $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t)=0$. This is a simple consequence of the superposition theorem 4.9 for derivations.

Theorem 5.5. *Let $\boldsymbol{M}=(M_t)_{t\in[0,T]}\in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ and $b:[0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}^d$ be an $L^1(\widetilde{M_t}\otimes dt)$-non-local vector fields, according to Definition 4.11.*
Assume that the continuity equation $\partial_t M_t+\mathrm{div}_{\mathcal{P}}(b_t M_t)=0$ is satisfied. Then there exists a measure $\Lambda\in\mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ that is concentrated over CE(b) *and $(\mathfrak{e}_t)_\sharp\Lambda = M_t$ for all $t\in[0,T]$.*

Proof. Consider the family of $L^1(M_t\otimes dt)$-derivations induced by the family of non-local vector fields b, i.e.

$$B_t[F](\mu) := \int_{\mathbb{R}^d} b(t,x,\mu)\cdot\nabla_W F(x,\mu)d\mu(x)\quad \forall F\in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)), \tag{5.6}$$

that is well-defined M_t-a.e. and for almost all $t\in(0,T)$, since

$$\int_0^T\int_{\mathcal{P}(\mathbb{R}^d)}\int_{\mathbb{R}^d}|b(t,x,\mu)|d\mu(x)dM_t(\mu)dt<+\infty. \tag{5.7}$$

It is an $L^1(M_t\otimes dt)$ family of derivations, so we can apply Theorem 4.9 to obtain $\Lambda\in\mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ such that $(\mathfrak{e}_t)_\sharp\Lambda = M_t$ for all $t\in[0,T]$ and Λ-a.e. $\boldsymbol{\mu}=(\mu_t)_{t\in[0,T]}$ satisfies

$$\frac{d}{dt}\int_{\mathbb{R}^d}\phi(x)d\mu_t(x) = B_t[L_\phi](\mu_t) = \int_{\mathbb{R}^d} b(t,x,\mu_t)\cdot\nabla\phi(x)d\mu_t(x),$$

for all $\phi\in C^1_c(\mathbb{R}^d)$, in the sense of distributions in $(0,T)$. □

5.2. **Nested superposition: from Λ to $\mathfrak{L}$.** Here, we want to invert the result of Proposition 5.2, i.e. to build a measure $\mathfrak{L}\in\mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ concentrated over SPS(b) to a given measure $\Lambda\in\mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ concentrated over CE(b). Let us introduce some notations: we will denote

$$Y := [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d),\qquad Z:=[0,T]\times C_T(\mathbb{R}^d)\times\mathcal{P}(C_T(\mathbb{R}^d)). \tag{5.8}$$

Then, we define the maps

$$\begin{aligned} &\kappa : C_T(\mathcal{P}(\mathbb{R}^d)) \to \mathcal{M}_+(Y), \quad \kappa(\boldsymbol{\mu}) = dt \otimes \big(\mu_t \otimes \delta_{\mu_t}\big), \\ &\mathfrak{K} : \mathcal{P}(C_T(\mathbb{R}^d)) \to \mathcal{M}_+(Z), \quad \mathfrak{K}(\lambda) := \mathcal{L}^1_T \otimes \lambda \otimes \delta_\lambda. \end{aligned} \tag{5.9}$$

For a fixed Borel non-local vector field $b : [0,T] \times \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, we define the subset $\hat{\mathrm{CE}}(b) \subset \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d)$ as

$$\begin{aligned} \hat{\mathrm{CE}}(b) := \big\{(\hat\mu, \hat\nu) \in \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d) \ : \ &\hat\mu \in \mathrm{Im}(\kappa),\ b \in L^1(\hat\mu), \\ &\hat\nu = b\hat\mu,\ \partial_t\hat\mu + \mathrm{div}(\hat\nu) = 0\big\}, \end{aligned} \tag{5.10}$$

where the equation $\partial_t\hat\mu + \mathrm{div}(\hat\nu) = 0$ must be understood as

$$\forall \xi \in C_c((0,T) \times \mathbb{R}^d) \ \int \partial_t\xi(t,x) d\hat\mu(t,x,\mu) + \int \nabla_x\xi(t,x) \cdot d\hat\nu(t,x,\mu) = 0,$$

and, lastly, define the set $\hat{\mathrm{SPS}}(b) \subset \mathcal{M}_+(Z)$ as

$$\begin{aligned} \hat{\mathrm{SPS}}(b) := \Big\{ \hat\lambda \in \mathcal{M}_+(Z) \ : \ &\hat\lambda \in \mathrm{Im}(\mathfrak{K}),\ \hat\lambda\big(([0,T] \times AC_T(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d))^c\big) = 0, \\ &\hat D \in L^1(\hat\lambda),\ \int |\hat D - b \circ \hat E| d\hat\lambda = 0 \Big\}, \end{aligned} \tag{5.11}$$

where $\hat D(t,\gamma,\lambda) = D(t,\gamma)$, according to Lemma C.2, for all $(t,\gamma,\lambda) \in Z$, and

$$\hat E : Z \to [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d), \quad \hat E(t,\gamma,\lambda) = (t, \gamma(t), (e_t)_\sharp\lambda). \tag{5.12}$$

Notice that $\hat E$ coincide with the map $\mathcal{E}$ already defined in (4.30), we just call it $\hat E$ here for the sake of notation.

The proof of the nested superposition principle relies on the measurability of the sets $\mathrm{CE}(b)$, $\mathrm{SPS}(b)$, $\hat{\mathrm{CE}}(b)$ and $\hat{\mathrm{SPS}}(b)$, for which the following lemma will be fundamental.

Lemma 5.6. *The maps κ and $\mathfrak{K}$ are continuous and injective. In particular, $\mathrm{Im}(\kappa) \subset \mathcal{M}_+(Y)$ and $\mathrm{Im}(\mathfrak{K}) \subset \mathcal{M}_+(Z)$ are Borel subsets.*

Proof. The two functions are clearly continuous, and the injectivity follows by uniqueness of the disintegration with respect to the projection on the time variable. The last part of the statement follows from Lemma A.2 and the definition of Lusin set. □

Proposition 5.7. *Let $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be Borel. It holds*

$$\mathrm{CE}(b) = \kappa^{-1}(\pi^1(\hat{\mathrm{CE}}(b))), \tag{5.13}$$

where $\pi^1(\hat\mu,\hat\nu) = \hat\mu$ is the projection on the first coordinate. Moreover, the set $\hat{\mathrm{CE}}(b) \subset \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d)$ is Borel and, in particular, $\mathrm{CE}(b) \subset C_T(\mathcal{P}(\mathbb{R}^d))$ is Borel.

Proof. Thanks to Lemma D.1, Lemma D.13 and Lemma 5.6, the sets

$$B_1 := \{\hat\mu \in \mathcal{M}_+(Y) \ : \ \hat\mu \in \mathrm{Im}(\kappa),\ b \in L^1(\hat\mu)\}$$

$$B := \{(\hat\mu,\hat\nu) \in \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d) \ : \ \hat\mu \in \mathrm{Im}(\kappa),\ b \in L^1(\hat\mu),\ \hat\nu = b\hat\mu\}$$

are Borel, respectively, in $\mathcal{M}_+(Y)$ and $\mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d)$. Moreover, notice that $B_1 = \pi^1(B)$. Then, the set

$$\hat{\mathrm{CE}}(b) := \{(\hat\mu,\hat\nu) \in \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d) \ : \ b \in L^1(\hat\mu),\ \hat\nu = b\hat\mu,\ \partial_t\hat\mu + \mathrm{div}(\hat\nu) = 0\}$$

is Borel, since it is a relatively closed subset of B. Notice that the projection on the first coordinate restricted to B, i.e. $\pi^1|_B : B \to \mathcal{M}_+(Y)$, is continuous and injective, then it satisfies $\pi^1(B') \in \mathcal{B}\big(\mathcal{M}_+(Y)\big)$ for all $B' \subset B$ Borel set (see Lemma A.2 and Corollary A.3). Thus, we conclude proving (5.13): if $\boldsymbol{\mu} \in \mathrm{CE}(b)$, then $(\kappa(\boldsymbol{\mu}), b\kappa(\boldsymbol{\mu})) \in \hat{\mathrm{CE}}(b)$; on the other hand, if $\boldsymbol{\mu} \in \kappa^{-1}(\pi^1(\hat{\mathrm{CE}}(b)))$, then $(\hat{\mu}, \hat{\nu}) = \big(\kappa(\boldsymbol{\mu}), b\kappa(\boldsymbol{\mu})\big) \in \hat{\mathrm{CE}}(b)$, which implies that

$$\int_0^T \int |b_t(x, \mu_t)| d\mu_t(x) dt = \int_Y |b_t(x, \mu)| d\kappa(\boldsymbol{\mu})(t, x, \mu) < +\infty,$$

and

$$\begin{aligned} 0 &= \int_Y \frac{\partial}{\partial t}\xi(t, x) d\hat{\mu}(t, x, \mu) + \int_Y \nabla_x \xi(t, x) \cdot d\hat{\nu}(t, x, \mu) \\ &= \int_Y \frac{\partial}{\partial t}\xi(t, x) d\hat{\mu}(t, x, \mu) + \int_Y \nabla_x \xi(t, x) \cdot b(t, x, \mu) d\hat{\mu}(t, x, \mu) \\ &= \int_0^T \int_{\mathbb{R}^d} \frac{\partial}{\partial t}\xi(t, x) d\mu_t(x) dt + \nabla_x \xi(t, x) \cdot b(t, x, \mu_t) d\mu_t(x) dt, \end{aligned}$$

which means that $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]}$ satisfies the continuity equation, and so $\boldsymbol{\mu} \in \mathrm{CE}(b)$. □

Proposition 5.8. *Let $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be Borel. It holds*

$$\mathrm{SPS}(b) = \mathfrak{K}^{-1}(\hat{\mathrm{SPS}}(b)). \tag{5.14}$$

Moreover, the set $\hat{\mathrm{SPS}}(b) \subset \mathcal{M}_+(Z)$ is Borel and, in particular, $\mathrm{SPS}(b) \subset \mathcal{P}(C_T(\mathbb{R}^d))$ is Borel.

Proof. The set $\hat{\mathrm{SPS}}(b)$ is Borel thanks to Lemma 5.6, Lemma C.1, Lemma C.2 and Lemma D.1, and we conclude by proving (5.14). If $\lambda \in \mathrm{SPS}(b)$, then calling $\hat{\lambda} = \mathfrak{K}(\lambda)$ we have:

- since λ is concentrated over $AC_T(\mathbb{R}^d)$, it holds $\hat{\lambda}\big(([0, T] \times AC_T(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d))^c\big) = 0$;
- $\int |D - b \circ \hat{E}| d\hat{\lambda} = 0$ is equivalent to $\dot{\gamma}(t) = b(t, \gamma(t), (e_t)_\sharp \lambda)$ $\mathcal{L}^1_T \otimes \lambda$-a.e.;
- thanks to the previous point, $\hat{D} \in L^1(\hat{\lambda})$ is equivalent to $\int\int |b(t, \gamma(t), (e_t)_\sharp \lambda)| d\lambda dt < +\infty$.

So, $\mathfrak{K}(\lambda) \in \hat{\mathrm{SPS}}(b)$. On the other hand, if $\lambda \in \mathfrak{K}^{-1}(\hat{\mathrm{SPS}}(b))$, then $\hat{\lambda} := \mathfrak{K}(\lambda) \in \hat{\mathrm{SPS}}(b)$, which means that

- $\lambda\big((AC_T(\mathbb{R}^d))^c\big) = 0$;
- $\dot{\gamma}(t) = b(t, \gamma(t), (e_t)_\sharp \eta)$ for $\hat{\lambda}$-a.e. (t, γ, η). Since $\hat{\lambda} = \mathcal{L}^1_T \otimes \lambda \otimes \delta_\lambda$, we have that $\dot{\gamma}(t) = b(t, \gamma(t), (e_t)_\sharp \lambda)$ $\mathcal{L}^1_T \otimes \lambda$-a.e.;
- thanks to the previous point,
$$\int\int |b(t, \gamma(t), (e_t)_\sharp \lambda)| d\lambda dt = \int |\hat{D}| d\hat{\lambda} < +\infty.$$

Thus, $\lambda \in \mathrm{SPS}(b)$ and the proof is concluded. □

Now, everything is set to prove the following theorem.

Theorem 5.9. *Let $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be Borel measurable. Then, there exists a Souslin-Borel measurable map $G_b : \mathrm{CE}(b) \to \mathcal{P}(C_T(\mathbb{R}^d))$ satisfying $\mathrm{Im}\, G_b \subset \mathrm{SPS}(b)$ and $E \circ G_b(\boldsymbol{\mu}) = \boldsymbol{\mu}$ for all $\boldsymbol{\mu} \in \mathrm{CE}(b)$.*
In particular, if $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ is concentrated over $\mathrm{CE}(b)$, then the measure $\mathfrak{L} := (G_b)_\sharp \Lambda \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ is well defined and concentrated over $\mathrm{SPS}(b)$, satisfying $E_\sharp \mathfrak{L} = \Lambda$.

Proof. Notice that, thanks to Proposition 5.2, the map

$$E|_{\mathrm{SPS}(b)} : \mathrm{SPS}(b) \to \mathrm{CE}(b)$$

is well defined. Moreover, thanks to the finite dimensional superposition principle, Theorem 2.12, it is surjective. Indeed for all $\boldsymbol{\mu} \in \mathrm{CE}(b)$ there exists a lifting $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ such that $(e_t)_\sharp \lambda = \mu_t$, λ is supported over a.c. curves that solves $\dot\gamma(t) = b(t, \gamma_t, \mu_t)$ and

$$\int\int |b_t(\gamma(t),(e_t)_\sharp\lambda)| d\lambda(\gamma) dt = \int\int |b_t(x,\mu_t)| d\mu_t(x) dt < +\infty.$$

In other words, $\lambda \in \mathrm{SPS}(b)$. Then, thanks to Proposition 5.7 and Proposition 5.8, we can apply Theorem A.10 to obtain a Souslin-Borel measurable map $G_b : \mathrm{CE}(b) \to \mathrm{SPS}(b)$, thanks to which we can define $\mathfrak{L} := (G_b)_\sharp \Lambda$, because of Corollary A.9. Moreover, being G_b a right-inverse of E, we also have the equality $\Lambda = E_\sharp \mathfrak{L}$, so it preserves the hierarchy shown in Proposition 5.2. □

5.3. **Universality in the measurable selection.** In this subsection, we show a possible universal decomposition for a measurable selection map G_b, highlighting the dependence on the non-local vector field b in it.
We will use the same notations of the previous subsection introduced in (5.8)-(5.12), but here it will be fundamental to introduce also the following sets:

$$\widehat{\mathrm{CE}} := \big\{(\hat\mu,\hat\nu) \in \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d) \ : \ \hat\mu \in \mathrm{Im}(\kappa), \quad \partial_t\hat\mu + \mathrm{div}\,\hat\nu = 0, \quad \hat\nu \ll \hat\mu\big\},$$

$$\widehat{\mathrm{SPS}} := \Big\{\hat\lambda \in \mathcal{M}_+(Z) \ : \ \hat\lambda \in \mathrm{Im}(\mathfrak{K}), \quad \hat\lambda\big(([0,T] \times AC_T(\mathbb{R}^d) \times \mathcal{P}(C_T(\mathbb{R}^d)))^c\big) = 0, \\ \hat D \in L^1(\hat\lambda), \quad \exists v : Y \to \mathbb{R}^d \text{ s.t. } \hat D = v \circ \hat E \ \hat\lambda\text{-a.e.}\Big\}. \tag{5.15}$$

These sets are strictly related to the sets $\widehat{\mathrm{CE}}(b)$ and $\widehat{\mathrm{SPS}}(b)$ (see (5.10) and (5.11)), the difference is that they allow a general Borel field b instead of an a priori fixed one.
Indeed, for any element $(\hat\mu,\hat\nu) \in \widehat{\mathrm{CE}}$ there exists a Borel non-local vector field $v : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that $\hat\nu = v\hat\mu$. Moreover, since $\hat\mu \in \mathrm{Im}(\kappa)$, there exists a curve $\boldsymbol{\mu} \in C_T(\mathcal{P}(\mathbb{R}^d))$ such that $\hat\mu = \kappa(\boldsymbol{\mu})$, and by the first condition, it also holds that $\partial_t\mu_t + \mathrm{div}(v_t(\cdot,\mu_t)\mu_t) = 0$. We remark that the presence of $\hat\nu$ is also telling which vector field (defined $\hat\mu$-a.e.) drives the continuity equation. In particular, $\widehat{\mathrm{CE}}$ is formally the union of all $\widehat{\mathrm{CE}}(b)$ as b varies.
Regarding the other set, any $\hat\lambda \in \widehat{\mathrm{SPS}}$ can be represented as $\hat\lambda = \mathcal{L}^1 \otimes \lambda \otimes \delta_\lambda$ for some $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$, and by the other conditions it holds that λ is concentrated over absolutely continuous curves that solve the ordinary differential equation $\dot\gamma(t) = v_t(\gamma(t),(e_t)_\sharp\lambda)$. As before, $\widehat{\mathrm{SPS}}$ is formally the a union of all $\widehat{\mathrm{SPS}}(b)$ as b varies.

At this point, for all $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ Borel measurable we have the following commutative diagram:

$$\begin{array}{ccc} \mathrm{SPS}(b) & \xrightarrow{\ \mathfrak{K}\ } & \widehat{\mathrm{SPS}} \\ {\scriptstyle E}\big\downarrow & & \big\downarrow{\scriptstyle \hat{\mathcal{E}}} \\ \mathrm{CE}(b) & \xrightarrow{\ \kappa_b\ } & \widehat{\mathrm{CE}} \end{array} \tag{5.16}$$

where

$$\hat{\mathcal{E}}(\hat\lambda) := \big(\hat E_\sharp\hat\lambda, \hat E_\sharp(\hat D\hat\lambda)\big), \quad \kappa_b(\hat\mu) := \big(\kappa(\boldsymbol{\mu}), b\kappa(\boldsymbol{\mu})\big).$$

In the following, we show that we can obtain a right inverse $\hat{\mathcal{G}} : \widehat{\mathrm{CE}} \to \widehat{\mathrm{SPS}}$ of the map $\hat{\mathcal{E}}$, performing again a measurable selection. Then, we infer that $\hat{\mathcal{G}} \circ \kappa_b(\mathrm{CE}(b)) \subseteq \mathfrak{K}(\mathrm{SPS}(b))$ for any measurable b, giving the diagram

$$\begin{array}{ccc} \mathrm{SPS}(b) & \xleftarrow{\mathfrak{K}^{-1}} & \hat{\mathrm{SPS}} \\ & & \Big\uparrow \hat{\mathcal{G}} \\ \mathrm{CE}(b) & \xrightarrow{\kappa_b} & \hat{\mathrm{CE}} \end{array} \tag{5.17}$$

Then, composition $\mathfrak{K}^{-1} \circ \hat{\mathcal{G}} \circ \kappa_b$ will be a right inverse of $E|_{\mathrm{SPS}_b}$. This strategy is an alternative proof for Theorem 5.9, providing a more refined right inverse of the map $E|_{\mathrm{SPS}_b}$. Moreover, the measurable selection map $\hat{\mathcal{G}}$ is universal, in the sense that it does not depend on b.

In the following, we make rigorous this strategy step by step, starting with a more general result that will be useful but could be of independent interest.

Lemma 5.10. *Let R, S be Polish spaces, $e : S \to R$ Borel measurable and $f : S \to \mathbb{R}^d$ a Borel function. Let $\lambda \in \mathcal{M}_+(S)$ be such that $f \in L^1(\lambda)$. Assume that*

$$\mathcal{H}\left(e_\sharp(f\lambda)|e_\sharp\lambda\right) = \mathcal{H}(f\lambda|\lambda),$$

where

$$\mathcal{H}(\nu|\mu) := \int \sqrt{1 + \left|\frac{d\nu}{d\mu}\right|^2} d\mu + |\nu^\perp|, \quad \nu = \frac{d\nu}{d\mu}\mu + \nu^\perp \text{ with } \nu^\perp \perp \mu.$$

Then, calling $v : R \to \mathbb{R}^d$ a version for the density of $e_\sharp(f\lambda)$ w.r.t. $e_\sharp\lambda$, we have that

$$f = v \circ e \ \lambda\text{-a.e.}$$

Proof. First of all, notice that

$$\mathcal{H}\left(e_\sharp(f\lambda)|e_\sharp\lambda\right) = \int_R \sqrt{1 + |v|^2} d(e_\sharp\lambda), \quad \mathcal{H}\left(f\lambda|\lambda\right) = \int_S \sqrt{1 + |f|^2} d\lambda.$$

Consider the disintegration of λ with respect to the map e, i.e. the family of probability measures $\{\lambda_x\}_{x\in R} \subset \mathcal{P}(S)$ such that

$$\lambda = \int_R \lambda_x d(e_\sharp\lambda)(x), \quad \lambda_x(e^{-1}(\{x\}) = 1.$$

Define the map

$$\tilde{v}(x) := \int_S f(y) d\lambda_x(y),$$

which is well-define $e_\sharp\lambda$-a.e. because $f \in L^1(\lambda)$ and Borel measurable thanks to Corollary D.6

Claim: $v = \tilde{v}$ $e_\sharp\lambda$-a.e.

To prove the claim, it suffices to prove that $\tilde{v}$ is a density for $e_\sharp(f\lambda)$ w.r.t. $e_\sharp\lambda$. For any $g : R \to \mathbb{R}^d$ it holds

$$\begin{aligned}\int_R g(x) \cdot de_\sharp(f\lambda)(x) &= \int_S g(e(y)) \cdot f(y) d\lambda(y) = \int_R \left(\int_S g(e(y)) \cdot f(y) d\lambda_x(y)\right) d(e_\sharp\lambda)(x) \\ &= \int_R g(x) \cdot \left(\int_S f(y) d\lambda_x(y)\right) d(e_\sharp\lambda) = \int_R g(x)\tilde{v}(x) d(e_\sharp\lambda)(x).\end{aligned}$$

Now, for simplicity, we call $H(z) = \sqrt{1 + |z|^2}$. Then it always holds

$$\mathcal{H}\left(e_\sharp(f\lambda)|e_\sharp\lambda\right) = \int_R H(\tilde{v}(x)) d(e_\sharp\lambda)(x) = \int_X H\left(\int_S f(y) d\lambda_x(y)\right) d(e_\sharp\lambda)(x)$$

$$\leq \int_R \int_S H(f(y))d\lambda_x(y)d(e_\sharp\lambda)(x) = \int_S H(f(y))d\lambda(y) = \mathcal{H}\left(f\lambda|\lambda\right),$$

where we used Jensen inequality. By hypothesis, we have equality, and by strict convexity of G the equality can hold if and only if for $(e_\sharp\lambda)$-a.e. $x \in R$, f is constant λ_x-a.e. In particular, f is constant $\lambda_{e(y)}$-a.e., for λ-a.e. $y \in S$, and looking in the definition of $\tilde{v}$, we conclude that $\tilde{v}(e(y)) = f(y)$ λ-a.e. □

Remark 5.11. *It is known that the same conclusion of the previous lemma is obtained if*

$$\int_R |v|^p d(e_\sharp\lambda) = \int_S |f|^p d\lambda, \tag{5.18}$$

for $p \in (1, +\infty)$. Unfortunately, this argument fails when $p = 1$, as the following counterexample shows. Indeed, under the same assumptions of the lemma, but assuming only (5.18) with $p = 1$, it is still true that $e_\sharp\eta \ll e_\sharp\lambda$. Anyway, it is not said that $|f| = |v \circ e|$: take $S = \mathbb{R}^2$, $R = \mathbb{R}$, $\lambda = \frac{1}{2}\mathcal{H}^1_{[0,1]\times\{-1\}} + \frac{1}{2}\mathcal{H}^1_{[0,1]\times\{1\}}$, $e(x, y) = x$, $f(x, y) = 1/2$ is $x < 0$ and $f(x, y) = 3/2$ if $x \geq 0$. Then

$$\eta := f\lambda = \frac{1}{4}\mathcal{H}^1_{[0,1]\times\{-1\}} + \frac{3}{4}\mathcal{H}^1_{[0,1]\times\{1\}}, \quad e_\sharp\lambda = e_\sharp\eta = \mathcal{L}^1_{[0,1]},$$

which means that $v \equiv 1$. Moreover, it's easy to verify that

$$\int |f|d\lambda = \frac{1}{2}\int_0^1 \frac{1}{2}dx + \frac{1}{2}\int_0^1 \frac{3}{2}dx = 1 = |e_\sharp\eta|(\mathbb{R}),$$

but it does not hold that $|f| = |v \circ e|$ for λ-a.e. $y \in S$. So, when $p = 1$, we can only conclude that

$$f = \frac{|f|}{|v \circ e|} v \circ e \quad \lambda\text{-a.e.}$$

In view of this, our result is a version of the result for $p > 1$, when we only have $f \in L^1$. Of course, the function $H(z) = \sqrt{1 + z^2}$ can be substituted with any strictly convex function $\tilde{H}$ satisfying $\lim_{z\to+\infty} \tilde{H}(z)/x < +\infty$.

Thanks to this lemma, we can then rewrite the set $\hat{\text{SPS}}$ as

$$\begin{aligned} \hat{\text{SPS}} := \Big\{ \hat{\lambda} \in \mathcal{M}_+(Z) \ : \ \hat{\lambda} \in \text{Im}(\mathfrak{K}), \quad \hat{\lambda}\big(([0, T] \times AC_T(\mathbb{R}^d) \times \mathcal{P}(C_T(\mathbb{R}^d)))^c\big) = 0, \\ \mathcal{H}\left(\hat{E}_\sharp(\hat{D}\hat{\lambda})|\hat{E}_\sharp\hat{\lambda}\right) = \int \sqrt{1 + |\hat{D}|^2} d\hat{\lambda}, \quad \hat{D} \in L^1(\hat{\lambda}) \Big\}. \end{aligned} \tag{5.19}$$

Indeed, thanks to Lemma 5.10, the constraint that there exists $v : Y \to \mathbb{R}^d$ such that $\hat{D} = v \circ \hat{E}$ $\hat{\lambda}$-a.e. can be rewritten by

$$\mathcal{H}\left(\hat{E}_\sharp(\hat{D}\hat{\lambda})|\hat{E}_\sharp\hat{\lambda}\right) = \int \sqrt{1 + |\hat{D}|^2} d\hat{\lambda}. \tag{5.20}$$

We are now ready to prove the following.

Theorem 5.12. *The sets $\hat{\text{CE}} \subset \mathcal{M}_+(Y) \times \mathcal{M}(Y; \mathbb{R}^d)$ and $\hat{\text{SPS}} \subset \mathcal{M}_+(Z)$, and the map $\hat{\mathcal{E}}$ are Borel measurable. Moreover,*

$$\hat{\mathcal{E}}|_{\hat{\text{SPS}}} : \hat{\text{SPS}} \to \hat{\text{CE}} \tag{5.21}$$

is surjective. In particular, there exists a Souslin-Borel measurable map $\hat{\mathcal{G}} : \hat{\text{CE}} \to \hat{\text{SPS}}$ such that $\hat{\mathcal{E}} \circ \hat{\mathcal{G}} = \text{id}_{\hat{\text{CE}}}$. Moreover, for all $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ Borel, it holds $\hat{\mathcal{E}}(\hat{\text{SPS}}(b)) = \hat{\text{CE}}(b)$ and $\hat{\mathcal{G}}(\hat{\text{CE}}(b)) \subset \hat{\text{SPS}}(b)$.

Proof. **Step 1**: the set $\hat{\mathrm{CE}}$ is measurable, since $\partial_t\hat\mu + \operatorname{div}\hat\nu = 0$ is a closed condition, $\hat\nu \ll \hat\mu$ is a Borel condition (see Corollary D.15) and $\mathrm{Im}(\kappa)$ is Borel measurable (see Lemma 5.6).
The set $\hat{\mathrm{SPS}}$ is Borel since: the set $\mathrm{Im}(\mathfrak{K})$ is Borel thanks to Lemma 5.6; from Lemma D.1, we have that the evaluation on Borel sets is Borel and $\hat D \in L^1(\hat\lambda)$ is a Borel condition; the map $\mathcal{M}_+(Z) \ni \hat\lambda \mapsto \hat D\hat\lambda \in \mathcal{M}(Z;\mathbb{R}^d)$ is Borel measurable (see Corollary D.12), as well as the push-forward operation through $\hat E$ (see Proposition D.8). Then the equality (5.20) is a Borel condition, again by Lemma D.1 and the fact that $\mathcal{H}$ is jointly l.s.c. (see [AFP00, Chapter 2]).

Step 2: as before, $\hat{\mathcal{E}}$ is the composition of Borel measurable, thanks to Corollary D.12 and Proposition D.8, thus it is Borel measurable.
Moreover, it maps $\hat{\mathrm{SPS}}$ in $\hat{\mathrm{CE}}$: let $\hat\lambda \in \hat{\mathrm{SPS}}$ and $(\hat\mu,\hat\nu) := \hat{\mathcal{E}}(\hat\lambda)$. Since $\hat\lambda \in \mathrm{Im}(\mathfrak{K})$, there exists $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ such that $\hat\lambda = \mathfrak{L}^1_T \otimes \lambda \otimes \delta_\lambda$, which implies that $\hat{\mathcal{E}}(\hat\lambda) = \mathcal{L}^1_T \otimes (\mu_t \otimes \delta_{\mu_t})$, where $\mu_t := (e_t)_\sharp\lambda$ for all $t \in [0,T]$. Moreover, the equality (5.20) implies (see Lemma 5.10) the existence of a vector field $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that $\hat D = b\circ\hat E$ for $\hat\lambda$-a.e. (t,γ,η). By definition of $\hat{\mathcal{E}}$ and again by Lemma 5.10, $\hat\nu = b\hat\mu$. Regarding the continuity equation, for all $\xi \in C^1_c((0,T)\times\mathbb{R}^d)$, it holds

$$\begin{aligned}\int \frac{\partial}{\partial t}\xi(t,x)d\hat\mu + \int \nabla_x\xi(t,x)d\hat\nu &= \int \frac{\partial}{\partial t}\xi(t,\gamma(t))d\hat\lambda + \int \nabla_x\xi(t,\gamma(t))\cdot D(t,\gamma)d\hat\lambda \\ &= \int\int_0^T \frac{\partial}{\partial t}\xi(t,\gamma(t))dtd\lambda(\gamma) = 0.\end{aligned}$$

Then, $\hat{\mathcal{E}}$ maps $\hat{\mathrm{SPS}}$ in $\hat{\mathrm{CE}}$.

Step 3: given $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and $\hat\lambda \in \hat{\mathrm{SPS}}(b)$ of the form $\hat\lambda = \mathcal{L}^1_T \otimes \lambda \otimes \delta_\lambda$, for some $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$, using the same notation of the previous step we have

$$\begin{aligned}0 = \int \frac{\partial}{\partial t}\xi(t,x)d\hat\mu + \int \nabla_x\xi(t,x)d\hat\nu &= \int \frac{\partial}{\partial t}\xi(t,\gamma(t))d\hat\lambda + \int \nabla_x\xi(t,\gamma(t))\cdot D(t,\gamma)d\hat\lambda \\ &= \int \frac{\partial}{\partial t}\xi(t,\gamma(t))d\hat\lambda + \int \nabla_x\xi(t,\gamma(t))\cdot b(t,\gamma(t),(e_t)_\sharp\eta)d\hat\lambda \\ &= \int_0^T\int \frac{\partial}{\partial t}\xi(t,\gamma(t))d\lambda dt + \int_0^T\int \nabla_x\xi(t,\gamma(t))\cdot b(t,\gamma(t),(e_t)_\sharp\lambda)d\lambda dt \\ &= \int_0^T\int \frac{\partial}{\partial t}\xi(t,x)d\mu_t(x)dt + \int_0^T\int \nabla_x\xi(t,x)\cdot b(t,x,\mu_t)d\mu_t dt,\end{aligned}$$

i.e. $\hat{\mathcal{E}}(\hat{\mathrm{SPS}}(b)) \subseteq \hat{\mathrm{CE}}(b)$. The equality follows from the surjectivity, which we prove in the next step.

Step 4: we are left to prove the surjectivity of the map $\hat{\mathcal{E}}$. For any $(\hat\mu,\hat\nu) \in \hat{\mathrm{CE}}$, we have that $\hat\mu$ is of the form $\mathcal{L}^1_T \otimes (\mu_t \otimes \delta_{\mu_t})$ for some $(\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d))$, and μ_t solves the continuity equation $\partial_t\mu_t + \operatorname{div}(b_t(\cdot,\mu_t)\mu_t) = 0$, where b is a density for $\hat\nu$ w.r.t. $\hat\mu$. Then we can use the finite dimensional superposition principle (Theorem 2.12) to obtain $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ that is in $\mathrm{SPS}(b)$. It is not hard to verify that $\hat\lambda := \mathfrak{K}(\lambda) \in \hat{\mathrm{SPS}}$ and $\hat{\mathcal{E}}(\hat\lambda) = (\hat\mu,\hat\nu)$.

Step 5: we can finally apply Theorem A.10 to obtain the existence of $\hat{\mathcal{G}} : \hat{\mathrm{CE}} \to \hat{\mathrm{SPS}}$ satisfying the requirements. Then, we are only left to prove that $\hat{\mathcal{G}}(\hat{\mathrm{CE}}(b)) \subset \hat{\mathrm{SPS}}(b)$, for which it is sufficient to show that $\hat{\mathcal{E}}|^{-1}_{\hat{\mathrm{SPS}}}(\hat{\mathrm{CE}}(b)) = \hat{\mathrm{SPS}}(b))$. Let $(\hat\mu,\hat\nu) \in \hat{\mathrm{CE}}(b)$ with $\hat\mu = \mathcal{L}^1_T \otimes (\mu_t \otimes \delta_{\mu_t})$ for some $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d))$ and consider $\hat\lambda \in \hat{\mathrm{SPS}}$ such that $\hat{\mathcal{E}}(\hat\lambda) = (\hat\mu,\hat\nu)$. Since

$\hat{\text{SPS}} \subset \text{Im}(\mathfrak{K})$, there exists $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ such that $\hat{\lambda} = \mathcal{L}^1_T \otimes \lambda \otimes \delta_\lambda$. Moreover, since

$$\mathcal{H}\left(\hat{E}_\sharp(\hat{D}\hat{\lambda})|\hat{E}_\sharp\hat{\lambda}\right) = \mathcal{H}\left(\hat{D}\hat{\lambda}|\hat{\lambda}\right),$$

we have that there exists a Borel measurable function $v : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ such that $\hat{D} = v \circ \hat{E}$ for $\hat{\lambda}$-a.e. (t,γ,η). In particular, $\hat{\lambda} \in \hat{\text{SPS}}(v)$, and thanks to Lemma 5.12 we have that $\hat{\mathcal{E}}(\hat{\lambda}) = (\hat{\mu},\hat{\nu}) \in \hat{\text{CE}}(v)$. By assumption $(\hat{\mu},\hat{\nu}) \in \hat{\text{CE}}(b)$, thus it holds $b = v$ for $\hat{\mu}$-a.e. (t,x,μ). Since $\hat{\mu} = \hat{E}_\sharp\hat{\lambda}$, we have that $b \circ \hat{E} = v \circ \hat{E} = D$ $\hat{\lambda}$-almost everywhere. In particular $\hat{\lambda} \in \hat{\text{SPS}}(b)$. □

We are now ready to conclude the argument we presented above, stating the main result. Let $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be a Borel measurable map. Define $\mathcal{M}_{+,b}(Y) := \left\{\hat{\mu} \in \mathcal{M}_+(Y) : b \in L^1(\hat{\mu};\mathbb{R}^d)\right\}$ and the map

$$\begin{aligned} V_b : \mathcal{M}_{+,b}(Y) &\to \mathcal{M}_+(Y) \times \mathcal{M}(Y;\mathbb{R}^d) \\ \hat{\mu} &\mapsto (\hat{\mu}, b\hat{\mu}), \end{aligned} \tag{5.22}$$

which is a Borel map thanks to Corollary D.12. Then, the map κ_b introduced in (5.16), is given by $\kappa_b = V_b \circ \kappa$, resulting Borel measurable. Recall that the (left) inverse of $\mathfrak{K}$ is given by $\mathfrak{K}^{-1} = \frac{1}{T}\pi^2_\sharp$.

Corollary 5.13. *The function $\mathfrak{K}^{-1} \circ \hat{\mathcal{G}} \circ \kappa_b$ maps* CE(b) *to* SPS(b) *and is a right inverse of the map $E|_{\text{CE}(b)}$, i.e. $E \circ \mathfrak{K}^{-1} \circ \hat{\mathcal{G}} \circ \kappa_b = \text{id}_{\text{CE}(b)}$. In particular, for any $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ concentrated over* CE(b)*, the random measure $\mathfrak{L} := (\mathfrak{K}^{-1} \circ \hat{\mathcal{G}} \circ \kappa_b)_\sharp\Lambda$ is concentrated over* SPS(b) *and satisfies the properties of Theorem 1.2.*

Proof. For simplicity, let us call $G_b := \mathfrak{K}^{-1} \circ \hat{\mathcal{G}} \circ \kappa_b$. First of all, notice that $\kappa_b = V_b \circ k$ is well defined and maps $\boldsymbol{\mu} \in \text{CE}(b)$ in $(\hat{\mu},\hat{\nu}) \in \hat{\text{CE}}(b) \cap \text{Im}(\kappa)$. Then, $\hat{\mathcal{G}}$ selects a measure $\hat{\lambda} \in \hat{\text{SPS}}$ such that $\hat{\mathcal{E}}(\hat{\lambda}) = (\hat{\mu},\hat{\nu})$.

Since $\hat{\mathcal{G}}(\text{CE}(b)) \subset \hat{\text{SPS}}(b)$, then $\hat{\lambda} := \hat{\mathcal{G}}(\kappa_b(\boldsymbol{\mu})) \in \hat{\text{SPS}}(b)$. We need to show that $\lambda := \mathfrak{K}^{-1}(\hat{\lambda}) \in \text{SPS}(b)$. All the needed checks are pretty straightforward:

- since $\lambda = \pi^2_\sharp\hat{\lambda}$, we have that $\lambda(AC_T(\mathbb{R}^d)) = 0$;
- for $\hat{\lambda}$-a.e. (t,γ,η) it holds $D(t,\gamma) = b(t,\gamma(t),(e_t)_\sharp\eta)$, which immediately implies that $D(t,\gamma) = b(t,\gamma(t),(e_t)_\sharp\lambda)$ for $\mathcal{L}^1_T \otimes \lambda$-a.e. (t,γ);
- $\int\int |b(t,\gamma_t,(e_t)_\sharp\lambda)|d\lambda(\gamma)dt = \int |D|d\hat{\lambda} < +\infty$.

Thus, we conclude that G_b maps CE(b) in SPS(b). We are left to show that G_b is a right inverse of E. Let $\lambda = G_b(\boldsymbol{\mu})$, then for all $f : [0,T] \times \mathbb{R}^d \to [0,+\infty]$ Borel, it holds

$$\int f(t,\gamma(t))d\lambda(\gamma) = \int f(t,\gamma(t))d\hat{\lambda}(t,\gamma,\eta) = \int f(t,x)d\hat{\mu}(t,x,\mu) = \int f(t,x)d\mu_t(x)dt,$$

where $\hat{\lambda} = \mathfrak{K}(\lambda)$ and $\hat{\mu} = \hat{E}_\sharp\hat{\lambda} = \mathcal{L}^1_T \otimes (\mu_t \otimes \delta_{\mu_t})$. In particular, $(e_t)_\sharp\lambda = \mu_t$ for all $t \in [0,T]$. □

6. Existence and uniqueness: the Lipschitz case

In this section, we are going to prove a uniqueness result for the solution of the continuity equation $\partial_t M_t + \text{div}_{\mathcal{P}}(b_t M_t) = 0$, under a Lipschitz assumption of the vector field b with respect to the variables $(x,\mu) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$.

Here we state first a useful lemma, and then the main uniqueness theorem of this section, whose proof is divided into several parts and postponed to the next subsection.

Lemma 6.1. *Let $p \geq 1$. Let $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ and $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ an $L^p(\widetilde{M_t} \otimes dt)$-non-local vector field (see Definition 4.11) satisfying $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$. The following properties hold:*

(i) if M_0 is concentrated over $\mathcal{P}_p(\mathbb{R}^d)$, then the same holds for M_t for all $t \in [0,T]$;
(ii) if $M_0 \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$, then $M_t \in \mathcal{P}_p(\mathcal{P}_p(\mathbb{R}^d))$ for all $t \in [0,T]$.

Proof. (i) Consider $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ given by Theorem 4.9. Thanks to (4.23) and the properties of Λ, we have that for Λ-a.e. $\boldsymbol{\mu} = (\mu_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathbb{R}^d))$ it holds

$$\int_0^T \int_{\mathbb{R}^d} |b(t,x,\mu_t)|^p d\mu_t(x)dt < +\infty,$$

and $\partial_t\mu_t + \mathrm{div}(b_t(\cdot,\mu_t)\mu_t) = 0$. Moreover, since $(\mathfrak{e}_0)_\sharp\Lambda = M_0$, we have that $\mu_0 \in \mathcal{P}_p(\mathbb{R}^d)$ for M_0-a.e. μ_0. Then, for any $(\mu_t)_{t\in[0,T]}$ with these properties, consider $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ given by Theorem 2.12, so that

$$\begin{aligned} W_p^p(\mu_t,\mu_0) &\leq \int_0^t \int |\dot\gamma|^p(s)d\lambda(\gamma)ds = \int_0^t \int |b(s,\gamma_s,(e_s)_\sharp\lambda)|^p(s)d\lambda(\gamma)ds \\ &= \int_0^t \int_{\mathbb{R}^d} |b(s,x,\mu_s)|^p d\mu_s(x)ds < +\infty \end{aligned}$$

and given a transport plan $\pi_{0,t}$ realizing $W_p^p(\mu_0,\mu_t)$, we have

$$\int_{\mathbb{R}^d} |x|^p d\mu_t(x) \leq 2^p \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi_{0,t}(x,y) + \int_{\mathbb{R}^d} |y^p| d\mu_0(y) < +\infty, \tag{6.1}$$

which gives that M_t is concentrated over $\mathcal{P}_p(\mathbb{R}^d)$. Property (ii) follows from Proposition 4.10. □

Theorem 6.2. *Let $p \geq 1$. Let $\boldsymbol{M} = (M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ and $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ an $L^p(M_t \otimes dt)$-non-local vector field satisfying:*

(1) for any $\mu_0,\mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$ there exists a W_p-optimal plan π between μ_0 and μ_1 such that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |b(t,x_0,\mu_0) - b(t,x_1,\mu_1)|^p d\pi(x_0,x_1) \leq L(t)W_p^p(\mu_0,\mu_1), \tag{6.2}$$

with $L \in L^1(0,T)$;

(2) $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ and $M_0 = \overline{M} \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ is concentrated over $\mathcal{P}_p(\mathbb{R}^d)$.

Then:

(i) there exists a unique $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ concentrated on $\mathrm{CE}(b)$ and such that $(\mathfrak{e}_0)_\sharp\Lambda = \overline{M}$;
(ii) there exists a unique $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ concentrated on $\mathrm{SPS}(b)$ and such that $(E_0)_\sharp\mathfrak{L} = \overline{M}$.

In particular, $\boldsymbol{M} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ is the unique solution of $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ satisfying $M_0 = \overline{M}$.

It is worth commenting on the existence for solutions of the continuity equation for random measures: it can be recovered under Lipschitz assumptions in the spirit of our theorem (see [CLOS22; BF21]) or even under just Carathéodory assumptions (see [BF24]). Indeed, if given a non-local vector field $b : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, we can prove existence for $\partial_t\mu_t + \mathrm{div}(b_t(\cdot,\mu_t)\mu_t) = 0$ for any starting measure $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$, then using a measurable selection argument and the disintegration theorem, for any $M_0 \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ we can prove existence for $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ concentrated over $\mathrm{CE}(b)$ and such that $(\mathfrak{e}_0)_\sharp\Lambda = M_0$. At this point, the superposition principle gives existence for: $(M_t)_{t\in[0,T]} \in C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$ solving $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ starting from M_0; and $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ concentrated over $\mathrm{SPS}(b)$ and such that $(E_0)_\sharp \mathfrak{L} = M_0$.

6.1. **Proof of Theorem 6.2.** Even if point (i) follows from point (ii), we prove it first using a more classical argument. Consider $\Lambda \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ given by Theorem 1.2. We will prove uniqueness for the trajectories

$$\partial_t \mu_t + \mathrm{div}(b_t(\cdot, \mu_t)\mu_t) = 0 \tag{6.3}$$

for any fixed starting point $\overline{\mu} \in \mathcal{P}_p(\mathbb{R}^d)$. Indeed, given $\boldsymbol{\mu}^0, \boldsymbol{\mu}^1 \in C_T(\mathcal{P}(\mathbb{R}^d))$ two solutions of (6.3), thanks to [AGS08, Theorem 8.4.7 and Remark 8.4.8], we can differentiate in time the quantity $W_p^p(\mu_t^0, \mu_t^1)$, to obtain that, given a W_p-optimal plan $\pi_t^{0,1}$ between μ_t^0 and μ_t^1

$$\begin{aligned}\frac{d}{dt} W_p^p(\mu_t^0, \mu_t^1) &= p \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_0 - x_1|^{p-2}(x_0 - x_1) \cdot \big(b(t, x_0, \mu_t^0) - b(t, x_1, \mu_t^1)\big) d\pi_t^{0,1}(x_0, x_1) \\ &\le p W_p^{p-1}(\mu_t^0, \mu_t^1) \left(\int |b(t, x_0, \mu_t^0) - b(t, x_1, \mu_t^1)|^p d\pi_t^{0,1}(x_0, x_1) \right)^{\frac{1}{p}} \le pL(t) W_p^p(\mu_t^0, \mu_t^1).\end{aligned}$$

Using Grönwall lemma, we conclude that for any $\overline{\mu} \in \mathcal{P}_p(\mathbb{R}^d)$ there exists a unique $\boldsymbol{\mu}_{\overline{\mu}} \in C_T(\mathcal{P}(\mathbb{R}^d))$ solution of (6.3). Thus, we have the following representation

$$\Lambda = \int_{\mathcal{P}_p(\mathbb{R}^d)} \delta_{\boldsymbol{\mu}_{\overline{\mu}}} d\overline{M}(\overline{\mu}), \tag{6.4}$$

implying the uniqueness result of the theorem.
Using this result, we can already prove the uniqueness of $\partial_t M_t + \mathrm{div}_{\mathcal{P}}(b_t M_t) = 0$ given a starting point $\overline{M} \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ concentrated over $\mathcal{P}_p(\mathbb{R}^d)$. Indeed, if we had two different solutions $\boldsymbol{M}^0$ and $\boldsymbol{M}^1$, Theorem 1.2 would give us two different $\Lambda^0, \Lambda^1 \in \mathcal{P}(C_T(\mathcal{P}(\mathbb{R}^d)))$ satisfying property (i) of Theorem 6.2, which is a contradiction.

We are left with the proof of (ii), that, similarly to the proof of (i), passes again through the uniqueness of superposition solutions $\lambda \in \mathrm{SPS}(b)$ with a fixed starting point $(e_0)_\sharp \lambda = \overline{\mu} \in \mathcal{P}_p(\mathbb{R}^d)$. In particular, we will see how the Lipschitz assumption in Theorem 6.2 implies a Lipschitz assumption in the variable space $x \in \mathbb{R}^d$, giving uniqueness of trajectories at the particle level. Before proceeding, we need some preliminary results, that are the extension to the case $p \ge 1$ of (a part of) [CSS25a, Lemma 6.1, Theorem 6.2 and Theorem 7.6].

Lemma 6.3. *Let $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$ and $\pi \in \Gamma(\mu_0, \mu_1)$. Assume μ_0 has finite support $S = \{\bar{x}_1, \ldots, \bar{x}_N\}$ with $\delta := \min\{|\bar{x}_i - \bar{x}_j| \ : \ i \neq j\}$ and*

$$\sup\big\{|y - x| \ : (x, y) \in \mathrm{supp}\,\pi\big\} \le \frac{\delta}{2}.$$

Then π is W_p-optimal, i.e. $W_p^p(\mu_0, \mu_1) = \int |y - x|^p d\pi(x, y)$.

Proof. It is sufficient to prove that the support of π satisfies the c-cyclical monotonicity property, with $c(x, y) := |x - y|^p$. Consider $\{(x_i, y_i)\}_{i=1}^n \subset \mathrm{supp}\,\pi$, with $x_0 := x_n$. Then

$$\sum_{i=1}^n |x_{i-1} - y_i|^p - |x_i - y_i|^p \ge \sum_{i=1}^n \left[\big||x_{i-1} - x_i| - |x_i - y_i|\big|^p - \left(\frac{\delta}{2}\right)^p \right] \ge 0,$$

because $(x_i, y_i) \in \mathrm{supp}\,\pi$, thus $|x_i - y_i| \le \delta/2$, and

$$|x_{i-1} - y_i| = |x_{i-1} - x_i + x_i - y_i| \ge |x_{i-1} - x_i| - |x_i - y_i| \ge \delta - \delta/2 = \delta/2.$$

□

Lemma 6.4. *Let $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$ be two measures with finite support, $\pi \in \Gamma(\mu_0, \mu_1)$ and $\mu_t := (\mathrm{x}^t)_\sharp \pi$, where $\mathrm{x}^t(x_0, x_1) := (1-t)x_0 + tx_1$. Then the following properties hold:*

(i) for every $s \in [0,1]$ there exists $\delta > 0$ such that for every $t \in [0,1]$ with $|t-s| \le \delta$ $\pi^{st} := (\mathrm{x}^s, \mathrm{x}^t)_\sharp \pi$ is a W_p optimal plan between μ_s and μ_t. Moreover

$$W_p^p(\mu_s, \mu_t) = |t-s|^p \int |x_0 - x_1|^p d\pi(x_0, x_1); \tag{6.5}$$

(ii) there exist $0 = t_0 < \cdots < t_K = 1$ such that for every $k = 1, \ldots, K$, $\mu|_{[t_{k-1}, t_k]}$ is a constant speed geodesic w.r.t. W_p and

$$W_p^p(\mu_s, \mu_r) = |r-s|^p \int |x_0 - x_1|^p d\pi(x_0, x_1) \quad \forall s, r \in [t_{k-1}, t_k]; \tag{6.6}$$

(iii) the length of the curve $t \mapsto \mu_t$, w.r.t. W_p, is $\big(\int |x_0 - x_1|^p d\pi(x_0, x_1)\big)^{1/p}$.

Proof. It is the very same of [CSS25a, Theorem 6.2]. □

Lemma 6.5. *Let $b : \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d$ be such that for all $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$ and some W_p-optimal plan $\pi \in \Gamma(\mu_0, \mu_1)$ it holds*

$$\int |b(x_0, \mu_0) - b(x_1, \mu_1)|^p d\pi(x_0, x_1) \le L \int |x_1 - x_0|^p d\pi(x_0, x_1), \tag{6.7}$$

for some $L \in (0, +\infty)$. Then (6.7) holds for any transport plan $\pi \in \Gamma(\mu_0, \mu_1)$.

Proof. The proof is divided in two steps: first we prove the result for measures that are supported on finite sets, and then we use an approximation procedure to extend the result to all measures.

Step 1: assume that μ_0, μ_1 have finite support and consider a generic transport plan $\pi \in \Gamma(\mu_0, \mu_1)$. Let $0 = t_0 < \cdots < t_K = 1$ be as in Lemma 6.4, so that $(\mathrm{x}^{t_{k-1}}, \mathrm{x}^{t_k})_\sharp \pi$ is a W_p-optimal plan between $(\mathrm{x}^{t_{k-1}})_\sharp \pi$ and $(\mathrm{x}^{t_k})_\sharp \pi$. It is also the unique optimal plan, see [AGS08, Lemma 7.2.1 and Theorem 7.2.2]. Then

$$\begin{aligned}
\Big(\int |b(x_0, \mu_0) - & b(x_1, \mu_1)|^p d\pi(x_0, x_1) \Big)^{1/p} \\
&\le \sum_{k=1}^K \Big(\int |b(\mathrm{x}^{t_{k-1}}, (\mathrm{x}^{t_{k-1}})_\sharp \pi) - b(\mathrm{x}^{t_k}, (\mathrm{x}^{t_k})_\sharp \pi)|^p d\pi(x_0, x_1) \Big)^{1/p} \\
&\le \sum_{k=1}^K L W_p\big((\mathrm{x}^{t_{k-1}})_\sharp \pi, \mathrm{x}^{t_k}_\sharp \pi\big) = \sum_{k=1}^K L(t_k - t_{k-1}) \int |x_1 - x_0|^p d\pi(x_0, x_1) \\
&= L \int |x_1 - x_0|^p d\pi(x_0, x_1).
\end{aligned}$$

Step 2: let $\mu_0^n \in \mathcal{P}(\mathbb{R}^d)$ (resp. $\mu_1^n \in \mathcal{P}(\mathbb{R}^d)$) have finite support and be such that $W_p(\mu_0^n, \mu_0) \to 0$ (resp. $W_p(\mu_1^n, \mu_1) \to 0$). Let $\pi_0^n \in \Gamma(\mu_0^n, \mu_0)$ and $\pi_1^n \in \Gamma(\mu_1, \mu_1^n)$ be W_p-optimal plans for which (6.7) is satisfied. Exploiting [ABS24, Proposition 8.6], let $\sigma_n \in \mathcal{P}\big((\mathbb{R}^d)^4\big)$ be such that

$$p^{12}_\sharp \sigma_n = \pi_0^n, \ p^{23}_\sharp \sigma_n = \pi, \ p^{34}_\sharp \sigma_n = \pi_1^n,$$

where p^{ij} is the projection on both i-th and j-th coordinates. In particular, $p^{14}_\sharp\sigma_n \in \Gamma(\mu_0^n, \mu_1^n)$ and converges to π w.r.t. W_p. Indeed, rearranging the coordinates of σ^n, we have a transport plan between π and $(p_1, p_4)_\sharp\sigma^n$, which is $(p_2, p_3, p_1, p_4)_\sharp\sigma^n \in \Gamma(\pi, (p_1, p_4)_\sharp\sigma^n)$, and

$$W_p^p(\pi, (p_1, p_4)_\sharp\sigma^n) \le \int_{(\mathbb{R}^d)^4} |y_2-y_1|^p+|y_3-y_4|^p d\sigma^n(y_1, y_2, y_3, y_4) = W_p^p(\mu_0^n, \mu_0)+W_p^p(\mu_1^n, \mu_1) \to 0.$$

Then, using the notation $(y_1, y_2, y_3, y_4) \in (\mathbb{R}^d)^4$, we have

$$\begin{aligned}\left(\int |b(x_0, \mu_0) - b(x_1, \mu_1)|^p d\pi(x_0, x_1)\right)^{1/p} &= \|b(y_2, \mu_0) - b(y_3, \mu_1)\|_{L^p(\sigma_n;\mathbb{R}^d)} \\ \le&\|b(y_2, \mu_0) - b(y_1, \mu_0^n)\|_{L^p} + \|b(y_1, \mu_0^n) - b(y_4, \mu_1^n)\|_{L^p} + \|b(y_4, \mu_1^n) - b(y_3, \mu_1)\|_{L^p} \\ \le& L^{1/p}\left[W_p(\mu_0^n, \mu_0) + W_p(\mu_1^n, \mu_1) + \left(\int |x_1 - x_0|^p dp^{14}_\sharp\sigma_n(x_0, x_1)\right)^{1/p}\right],\end{aligned}$$

where in the last inequality we used the fact that (6.7) holds for $p^{12}_\sharp\sigma_n = \pi_0^n$ and $p^{34}_\sharp\sigma_n = \pi_1^n$, and $p^{14}_\sharp\sigma_n$ is any transport plan between μ_0^n and μ_1^n, that are finitely supported so that (6.7) holds as well, thanks to Step 1. Then, we conclude passing to the limit as $n \to +\infty$. □

The next result shows how the Lipschitz property along all the possible transport plans implies a Lipschitz property in space. For a proof, we refer to [CSS25b, Theorem 4.8, (1)].

Lemma 6.6. *Let $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be satisfying the hypothesis of Lemma 6.5. Then, for all $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, the map $b(\cdot, \mu) : \operatorname{supp}\mu \to \mathbb{R}^d$ is L-Lipschitz.*

Now, we can proceed to prove point (ii) in Theorem 6.2. Notice that all the previous lemmas apply to any non-local vector field $(x, \mu) \mapsto b(t, x, \mu)$, for any t such that $L(t) < +\infty$ (in particular for a.e. $t \in (0, T)$). So, for a.e. $t \in (0, T)$ and for every $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, the map $b(t, \cdot, \mu) : \operatorname{supp}\mu \to \mathbb{R}^d$ is $L(t)$-Lipschitz.
Now, let $\lambda \in \mathrm{SPS}_b$ and consider $\gamma_0, \gamma_1 \in \operatorname{supp}\lambda$ that satisfy $\dot\gamma_i(t) = b(t, \gamma_i(t), (e_t)_\sharp\lambda)$, $i = 0, 1$. Thanks to Lemma 6.1, for $i = 0, 1$, we know that $(e_t)_\sharp\lambda \in \mathcal{P}_p(\mathbb{R}^d)$ for any $t \in (0, T)$ and $\gamma_i(t) \in \operatorname{supp}(e_t)_\sharp\lambda$, because $\gamma_i \in \operatorname{supp}\lambda$. Then, for a.e. $t \in (0, T)$

$$\begin{aligned}\frac{d}{dt}|\gamma_0(t) - \gamma_1(t)|^2 =&2\big(\gamma_0(t) - \gamma_1(t)\big) \cdot \big(b(t, \gamma_0(t), (e_t)_\sharp\lambda) - b(t, \gamma_1(t), (e_t)_\sharp\lambda)\big) \\ \le&2L(t)|\gamma_1(t) - \gamma_2(t)|^2.\end{aligned}$$

Thus, using Gronwall lemma and the continuity of the curves γ_0 and γ_1, we have $\gamma_0 = \gamma_1$, which implies that, defining $\overline{\mu} := (e_0)_\sharp\lambda$, it holds

$$\lambda = \lambda_{\overline{\mu}} := \int \delta_{\gamma_{\overline{x}}} d\overline{\mu}(\overline{x}), \tag{6.8}$$

where $\gamma_{\overline{x}} \in C_T(\mathbb{R}^d)$ is the unique curve in the support of λ solving $\dot\gamma(t) = b(t, \gamma(t), (e_t)_\sharp\lambda)$ starting from $\gamma(0) = \overline{x} \in \mathbb{R}^d$. Then, considering $\mathfrak{L} \in \mathcal{P}(\mathcal{P}(C_T(\mathbb{R}^d)))$ given by Theorem 1.2, $\mathfrak{L}$-a.e. $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ must be of the form (6.8), so that

$$\mathfrak{L} = \int_{\mathcal{P}(\mathbb{R}^d)} \delta_{\lambda_{\overline{\mu}}} d\overline{M}(\overline{\mu}), \tag{6.9}$$

giving the uniqueness of $\mathfrak{L}$ satisfying $(E_0)_\sharp\mathfrak{L} = \overline{M}$.

Remark 6.7. *The same proof works, in the case $p = 2$, only assuming monotonicity of the vector field b, i.e. that for any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a W_2-optimal plan π between μ_0 and μ_1 such that*

$$\int (x_0 - x_1) \cdot \big(b(t, x_0, \mu_0) - b(t, x_1, \mu_1)\big) d\pi(x_0, x_1) \le L(t) W_2^2(\mu_0, \mu_1), \tag{6.10}$$

with $L \in L^1(0, T)$. Indeed, thanks to the aforementioned [CSS25a; CSS25b], the same proofs work under this hypothesis.

Appendix A. Lusin sets, Souslin sets and measurable selection theorem

Here we collect some results from [Sch73] and [Bog07] about Lusin and Souslin sets, that are persistently used throughout the paper. In particular, we recall (without proof) the universal measurability of Souslin subsets and a measurable selection theorem.

Definition A.1 (Lusin sets)**.** *A subset $L \subset X$ of a Hausdorff topological space (X, τ), is said to be a Lusin set if there exists a Polish space (Y, τ') and a continuous and bijective map $i : Y \to L$. The space X is said to be a Lusin space if it is a Lusin set.*

Lusin spaces shares many interesting properties. Here we list the ones that are useful for our presentation and we suggest [Sch73, Chapter 2] for further reading.

Lemma A.2. *Let (X, τ) be a Lusin space. Then $L \subset X$ is a Lusin set if and only if it is a Borel set. In particular, the same holds when (X, τ) is Polish.*

A direct consequence of the definition of Lusin set and the previous lemma is the next corollary.

Corollary A.3. *Let (X, τ) and (Y, τ') be Lusin spaces and $f : Y \to X$ continuous and injective. Then, $f(B) \subset X$ is Borel for any $B \subset Y$ Borel.*

Lemma A.4. *Let (X, τ) be a Hausdorff topological space and $L_n \subset X$ a sequence of Lusin sets of X. Then $\bigcup L_n$ is a Lusin set.*

Definition A.5 (Souslin sets)**.** *A subset $S \subset X$ of a Hausdorff topological space (X, τ) is said to be a Souslin set if there exist a Polish space Y and a continuous map $f : Y \to X$ such that $f(Y) = S$. The space X is said to be a Souslin space if it is a Souslin set.*
Let $\tilde{\mathcal{S}}(X)$ be the class of all Souslin subsets of X and $\mathcal{S}(X)$ be the σ-algebra generated by $\tilde{\mathcal{S}}(X)$.

An important tool to better understand the structure of the Souslin subsets of a given space, is the so-called A-operation (or Souslin operation). In the following theorem, we see how they are connected.

Theorem A.6. *Let (X, τ) be a Hausdorff topological space. Every Souslin subset $S \in \tilde{\mathcal{S}}(X)$ can be obtained from closed sets by means of the A-operation, i.e. for any $S \in \tilde{\mathcal{S}}(X)$ there exists a class of closed sets $\{C_{n_1,\dots,n_k}\}$, where $(n_1, \dots, n_k)$ is any possible finite sequence, such that*

$$S = \bigcup_{(n_i) \in \mathbb{N}^\infty} \bigcap_{k=1}^{+\infty} C_{n_1,\dots,n_k}. \tag{A.1}$$

Moreover, the set $\tilde{\mathcal{S}}(X)$ is closed under the A-operation.

Remark A.7. *Within this theorem, one can think that $\tilde{\mathcal{S}}(X) = \mathcal{S}(X)$, but it is known to be false. A consequence of the theorem is that $\tilde{\mathcal{S}}(X)$ is closed under countable union and intersection. The problems are given by the complement operation: indeed, [Bog07, Corollary 6.6.10], if both*

S and S^c are Souslin, then S is a Borel subset, and by [Bog07, Theorem 6.7.10] we know that there exists a Souslin set that is not Borel.

Now, we show that the Souslin-measurable sets are *universally measurable.*

Proposition A.8. *Let (X, τ) be a Hausdorff space and μ a Borel, positive and finite measure over X. Then, the σ-algebra $\mathcal{B}_\mu$ of all the μ-measurable subsets is closed under the A-operation. In particular, $\mathcal{S}(X) \subset \mathcal{B}_\mu$ for all Borel, positive and finite measure μ, and we say that the Souslin σ-algebra is universally measurable.*

In particular, we can use any Souslin-Borel measurable map to define a push-forward of a Borel measure and obtain a Borel measure, as the next Corollary shows.

Corollary A.9. *Let X, Y be two Souslin topological spaces. Let $\mu \in \mathcal{M}_+(X)$ be Borel measure on X and $f : X \to Y$ be a Souslin-Borel measurable map, i.e. $f^{-1}(B) \in \mathcal{S}(X)$ for any $B \in \mathcal{B}(Y)$. Then $\nu := f_\sharp \mu \in \mathcal{M}_+(Y)$ is a well-defined measure over Y.*

Finally, we state a version of the *measurable selection theorem*, see [Bog07, Theorem 6.9.1].

Theorem A.10. *Let X and Y be two Souslin topological spaces. Let $F : X \to Y$ be a surjective Borel map. Then, there exists a $(\mathcal{S}(Y), \mathcal{B}(X))$-measurable map $G : Y \to X$ that is a right-inverse of F, i.e. $F(G(y)) = y$ for any $y \in Y$. In addition, the image of G belongs to $\mathcal{S}(X)$.*

Appendix B. The extended metric-topological structure of $\mathbb{R}^\infty$

In this section, we describe some properties of the space $\mathbb{R}^\infty$, i.e. the space of real sequences $\mathrm{x} = (x_n)_{n\in\mathbb{N}}$, endowed with two different structures:

- the metric
$$D_\infty(\mathrm{x}, \mathrm{y}) := \sup_{n\in\mathbb{N}} |x_n - y_n| \wedge 1, \tag{B.1}$$
inducing the uniform convergence;
- the topology τ_w induced by the element-wise convergence, which is the topology given by the metric
$$d_\infty(\mathrm{x}, \mathrm{y}) := \sum_{n\in\mathbb{N}} \frac{|x_n - y_n| \wedge 1}{2^n}, \tag{B.2}$$
that is complete, and thus the topological space $(\mathbb{R}^\infty, \tau_w)$ is Polish. We will always consider on $\mathbb{R}^\infty$ the Borel σ-algebra generated by τ_w.

This structure is related to $(\mathcal{P}(\mathbb{R}^d), \hat{W}_1)$. Let $\mathcal{A} = \{\phi_1, \phi_2, \dots\} \subset C^1_c(\mathbb{R}^d)$ be satisfying:

(i) ϕ_k 1-Lipschitz w.r.t. $|\cdot| \wedge 1$, in particular $\|\nabla\phi_k\|_\infty \le 1$ for all $k \in \mathbb{N}$;
(ii) $\mathrm{Span}(\mathcal{A})$ dense in $C^1_0(\mathbb{R}^d)$;
(iii) $\hat{W}_1(\mu, \nu) = \sup_k \int_{\mathbb{R}^d} \phi_k d(\mu - \nu)$.

The existence of such family is justified from the fact that the countable class of functions
$$\mathcal{A}' := \Big\{\phi_{k,n}(x) = \big(|x - x_k| \wedge 1\big) * \rho_{\frac{1}{n}}, \quad k, n \in \mathbb{N}\Big\}$$
satisfies the third condition, where $\{x_k\} \subset \mathbb{R}^d$ is a countable and dense subset of $\mathbb{R}^d$, $\rho(x)$ is a mollifier and $\rho_\varepsilon(x) = \frac{1}{\varepsilon^d}\rho(\frac{x}{\varepsilon})$. On the other hand, the unit ball of $C^1_0(\mathbb{R}^d)$, endowed with the C^1-norm, is separable. Now, define
$$\iota : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^\infty, \quad \iota(\mu) = \big(L_{\phi_1}(\mu), L_{\phi_2}(\mu), \dots\big). \tag{B.3}$$

Lemma B.1. *The map ι is an isometry between $(\mathcal{P}(\mathbb{R}^d), \hat{W}_1)$ and $(\iota(\mathcal{P}(\mathbb{R}^d)), D_\infty)$. In particular, $\iota(\mathcal{P}(\mathbb{R}^d))$ is bounded and closed w.r.t. D_∞, while it is Borel but not closed with respect to τ_w.*

Proof. The fact that ι is an isometry comes from (B) of $\mathcal{A}$, which tells us

$$D_\infty(\iota(\mu), \iota(\nu)) = \sup_n |L_{\phi_n}(\mu) - L_{\phi_n}(\nu)| \wedge 1 = \sup_n |L_{\phi_n}(\mu) - L_{\phi_n}(\nu)| = \hat{W}_1(\mu, \nu).$$

Then, the set $\iota(\mathcal{P}(\mathbb{R}^d)$ must be bounded and closed w.r.t. D_∞. On the other hand, consider the sequence of measures $\mu_n := \delta_{x_n}$, with $x_n \in \mathbb{R}^d$ such that $|x_n| \to +\infty$. Then, $\iota(\mu_n) \to 0$, i.e. $\iota(\mu_n)$ converges element-wise to 0, but the sequence given by all zeros is not in $\iota(\mathcal{P}(\mathbb{R}^d))$, thus $\iota(\mathcal{P}(\mathbb{R}^d))$ is not closed in τ_w. Anyway, ι is continuous when endowing $\mathcal{P}(\mathbb{R}^d)$ with the narrow topology and $\mathbb{R}^\infty$ with τ_w. Since it is also injective and $(\mathbb{R}^\infty, \tau_w)$ is Polish, then $\iota(\mathcal{P}(\mathbb{R}^d))$ is Borel because it is Lusin (see Lemma A.2 and Corollary A.3). □

We proceed by introducing the cylinder functions for $\mathbb{R}^\infty$, the continuity equation over $\mathbb{R}^\infty$ and the superposition principle. For the proofs, we refer to [AT14, Chapter 7].

Definition B.2 (Cylinder functions of $\mathbb{R}^\infty$)**.** *We say that a function $F : \mathbb{R}^\infty \to \mathbb{R}$ is a cylinder function, and we write $F \in \mathrm{Cyl}^1(\mathbb{R}^\infty)$, if there exists $k \in \mathbb{N}$ and $\Psi \in C^1(\mathbb{R}^k)$ bounded and with bounded derivatives, such that*

$$F(\mathrm{x}) = \Psi(\pi_k(\mathrm{x})) = \Psi(x_1, \dots, x_k) \quad \forall x \in \mathbb{R}^\infty. \tag{B.4}$$

Its gradient is then defined as

$$\nabla F(\mathrm{x}) = (\partial_1 \Psi(\pi_k(\mathrm{x})), \dots, \partial_k \Psi(\pi_k(\mathrm{x})), 0, 0, \dots). \tag{B.5}$$

Definition B.3 (Continuity equation over $\mathbb{R}^\infty$)**.** *Let $v : [0,T] \times \mathbb{R}^\infty \to \mathbb{R}^\infty$ be a Borel vector field (w.r.t. τ_w) and $(\mathtt{m}_t)_{t\in[0,T]} \subset \mathcal{P}(\mathbb{R}^\infty)$ a weakly* continuous curve of probability measures over $\mathbb{R}^\infty$. Then, we say that the continuity equation $\partial_t \mathtt{m}_t + \mathrm{div}(v_t \mathtt{m}_t) = 0$ holds if*

$$\int_0^T \int |v_t^{(k)}| d\mathtt{m}_t dt < +\infty \quad \forall k \in \mathbb{N}, \tag{B.6}$$

where the superscript k indicates the k-th component of v, and

$$\frac{d}{dt} \int F d\mathtt{m}_t = \int \nabla F \cdot v_t d\mathtt{m}_t, \quad \forall F \in \mathrm{Cyl}^1(\mathbb{R}^\infty) \tag{B.7}$$

in the sense of distribution of $(0,T)$.

Definition B.4. *The set $AC_w([0,T], \mathbb{R}^\infty) \subset C([0,T], (\mathbb{R}^\infty, \tau_w))$ is the set of τ-continuous curves $\gamma : [0,T] \to \mathbb{R}^\infty$ that are element-wise absolutely continuous.*

Theorem B.5 (Superposition principle over $\mathbb{R}^\infty$)**.** *Let $v : [0,T] \times \mathbb{R}^\infty \to \mathbb{R}^\infty$ and $(\mathtt{m}_t)_{t\in[0,T]}$ as in the previous definition, satisfying $\partial_t \mathtt{m}_t + \mathrm{div}(v_t \mathtt{m}_t) = 0$. Then there exists a probability measure $\mathtt{L} \in \mathcal{P}(C_w([0,T], \mathbb{R}^\infty))$ such that $(e_t)_\sharp \mathtt{L} = \mathtt{m}_t$, it is concentrated over $AC_w([0,T], \mathbb{R}^\infty)$ and for λ-a.e. $\tilde{\gamma}$, it holds*

$$\frac{d}{dt} \tilde{\gamma}^{(k)}(t) = v_t^{(k)}(\tilde{\gamma}_t) \quad \forall k \in \mathbb{N}. \tag{B.8}$$

Appendix C. Measurability in the space of curves

We collect here some results concerning measurability properties of subsets of $C_T(Y)$, where Y is a Polish space and $C_T(Y)$ is endowed with the distance D_d where d is a distance on Y generating its topology.

Lemma C.1. *For any $p \in [1, +\infty)$, the space $AC^p_T(Y)$ is a Borel subsets of $C_T(Y)$, both endowed with the sup distance.*

Proof. By Theorem 10.2, [ABS24], the functional a_p is lower semicontinuous (their proof can be easily extended to the case $p \in (1, +\infty)$ using again Lemma 10.1, [ABS24]). Then, the sublevel sets are closed, and by the definitions above we can write AC^p as the union of the sublevel sets at level $n \in \mathbb{N}$. Then it is F_σ, and so it is Borel. Regarding the case $p = 1$, we refer to [AGS14, §2.2] □

Lemma C.2. *Let $D : [0, T] \times C_T(\mathbb{R}^d) \to \big(\mathbb{R} \cup \{\pm\infty\}\big)^d$ defined as the pointwise and component-wise* lim sup *of the discrete derivatives, i.e.*

$$\big(D(t, \gamma)\big)_j = \limsup_{h \to 0} \frac{\gamma_j(t + h) - \gamma_j(t)}{h}, \quad j = 1, \dots, d.$$

Then, the function D is Borel measurable and for all $\lambda \in \mathcal{P}(C_T(\mathbb{R}^d))$ concentrated over $AC_T(\mathcal{P}(\mathbb{R}^d))$, $D(t, \gamma) \in \mathbb{R}^d$ for $\mathcal{L}^1_T \otimes \lambda$-a.e. (t, γ).

Proof. Being the lim sup of continuous functions, D is Borel. The last property follows from the fact that for λ-a.e. γ and a.e. $t \in (0, T)$, the derivative of γ in t exists and is finite. □

C.1. **Curves in $\mathbb{R}^\infty$.** Here we show a measurability result, linking the two spaces of continuous curves on $\mathbb{R}^\infty$ with respect to the two topologies presented in Appendix B, i.e. $C_T(\mathbb{R}^\infty, D_\infty)$ and $C_T(\mathbb{R}^\infty, \tau_w)$. In particular, we will need the following result.

Lemma C.3. *Let $\iota : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^\infty$ be as in* (B.3)*. The set $C_T(\iota(\mathcal{P}(\mathbb{R}^d)), D_\infty)$ is a Borel subset of $C_T(\mathbb{R}^\infty, \tau_w)$. Moreover, the Borel σ-algebra generated by its subspace topology induced by $C_T(\mathbb{R}^\infty, D_\infty)$ coincides with the Borel σ-algebra generated by the subspace topology induced by $C_T(\mathbb{R}^\infty, \tau_w)$.*

Proof. The space $(\iota(\mathcal{P}(\mathbb{R}^d)), D_\infty)$ is Polish, since it is isometric to $(\mathcal{P}(\mathbb{R}^d), \hat{W}_1)$, then the same holds for $C_T(\iota(\mathcal{P}(\mathbb{R}^d)), D_\infty)$, with its natural compact-open topology, that we call $\mathcal{T}$. Such a topology is clearly stronger than the subspace topology induced by the larger (Polish) space $C_T(\mathbb{R}^\infty, \tau_w)$, that we denote by $\mathcal{T}_w$, so that the map

$$\mathrm{id} : \big(C_T(\iota(\mathcal{P}), D_\infty), \mathcal{T}\big) \to \big(C_T(\mathbb{R}^\infty, \tau_w), \mathcal{T}_w\big)$$

is continuous and injective, i.e. $C_T(\iota(\mathcal{P}), D_\infty)$, as a topological subspace of $C_T(\mathbb{R}^\infty, \tau_w)$ is Lusin by Definition A.1, and then Borel by Lemma A.2. To conclude, notice that both the topology over $C_T(\iota(\mathcal{P}), D_\infty)$ makes it a Lusin space and they are comparable; then the induced Borel σ-algebras coincide by [Sch73, Corollary 2, pp. 101]. □

C.2. **Curve of random measures.**

Lemma C.4. *For any* $\mathtt{m}_0, \mathtt{m}_1 \in \mathcal{P}(\mathbb{R}^\infty)$ *it holds*

(C.1)
$$W_{1,D_\infty}(\mathtt{m}_0,\mathtt{m}_1) = \sup\left\{ \int_{\mathbb{R}^\infty} F d(\mathtt{m}_0 - \mathtt{m}_1) \ : \ F(x) \in \mathrm{Cyl}^1_c(\mathbb{R}^\infty),\ F \text{ 1-Lip } \textit{w.r.t. } D_\infty \right\}$$
$$= \sup\left\{ \int_{\mathbb{R}^\infty} \Psi\circ\pi_k d(\mathtt{m}_0 - \mathtt{m}_1) : k \in \mathbb{N},\ \Psi \in C^1_c(\mathbb{R}^k),\ \|\Psi\|_\infty \le 1/2,\ \Big\| \sum_{i=1}^k |\partial_i \Psi| \Big\|_\infty \le 1 \right\}.$$

Proof. We prove the first inequality, while the second one will be a byproduct of our argument. The $\geq$ inequality is trivial, so we focus on the other one. Define the projection functions

$$\pi^n : \mathbb{R}^\infty \to \mathbb{R}^\infty, \quad \pi^n(\mathrm{x}) = (x_1, \dots, x_n, 0, 0, \dots)$$
$$p^n : \mathbb{R}^\infty \to \mathbb{R}^n, \quad p^n(\mathrm{x}) = (x_1, \dots, x_n).$$

Notice that $W_{1,D_\infty}(\pi^n_\sharp \mathtt{m}_0, \pi^n_\sharp \mathtt{m}_1) \le W_{1,D_\infty}(\mathtt{m}_0,\mathtt{m}_1)$. Moreover, for all $\mathtt{m} \in \mathcal{P}(\mathbb{R}^\infty)$, we have $\pi^n_\sharp \mathtt{m} \to \mathtt{m}$ weakly in duality with $C_b(\mathbb{R}^\infty, \tau_w)$. Indeed, for any $\mathrm{x} \in \mathbb{R}^\infty$, $\pi^n(\mathrm{x}) \xrightarrow{\tau} \mathrm{x}$, so for any $F \in C_b(\mathbb{R}^\infty, \tau_w)$, by the dominated convergence theorem, it holds

$$\int F(\mathrm{x}) d(\pi^n_\sharp \mathtt{m})(\mathrm{x}) = \int F(\pi^n(\mathrm{x})) d\mathtt{m}(\mathrm{x}) \to \int F(\mathrm{x}) d\mathtt{m}(x).$$

Moreover, D_∞ is $\tau_w \otimes \tau_w$ l.s.c.: indeed consider $\mathrm{x}^{(n)}, \mathrm{y}^{(n)} \in \mathbb{R}^\infty$ respectively, converging component-wise to x and y, then

$$\liminf_{n\to+\infty} D_\infty(\mathrm{x}^{(n)}, \mathrm{y}^{(n)}) = \liminf_{n\to+\infty} \sup_{j\in\mathbb{N}} |x_j^{(n)} - y_j^{(n)}| \wedge 1$$
$$\ge \liminf_{n\to+\infty} \sup_{j\le m} |x_j^{(n)} - y_j^{(n)}| \wedge 1 = \sup_{j \le m} |x_j - y_j| \wedge 1,$$

for any $m \in \mathbb{N}$, and for its arbitrariness we conclude.

Consider a sequence of W_{1,D_∞}-optimal plans $\Pi_n \in \Gamma_0(\pi^n_\sharp \mathtt{m}_0, \pi^n_\sharp \mathtt{m}_1)$. By the convergence of the marginals and the fact that $(\mathbb{R}^\infty, \tau_w)$ is a Polish space, we have that the set of measures $\{\Pi_n\}$ is tight, which implies that

$$\forall (n_k)\ \exists (n_{k_j}) \ : \ \Pi_{n_{k_j}} \to \Pi \in \Gamma(\mathtt{m}_0, \mathtt{m}_1),$$

where the convergence is weakly in duality with $C_b(\mathbb{R}^\infty \times \mathbb{R}^\infty, \tau\otimes\tau)$. Then, by $\tau_w \otimes \tau_w$-l.s.c. of D_∞ we have

$$\liminf_{j\to+\infty} W_{1,D_\infty}(\pi^{n_{k_j}}_\sharp \mathtt{m}_0, \pi^{n_{k_j}}_\sharp \mathtt{m}_1) = \liminf_{j\to+\infty} \int D_\infty(\mathrm{x},\mathrm{y}) d\Pi_{n_{k_j}}(\mathrm{x},\mathrm{y})$$
$$\ge \int D_\infty(\mathrm{x},\mathrm{y}) d\Pi(\mathrm{x},\mathrm{y}) \ge W_{1,D_\infty}(\mathtt{m}_0,\mathtt{m}_1),$$

and by arbitrariness of (n_k) we conclude that $\liminf_{n\to+\infty} W_{1,D_\infty}(\pi^n_\sharp \mathtt{m}_0, \pi^n_\sharp \mathtt{m}_1) \ge W_{1,D_\infty}(\mathtt{m}_0,\mathtt{m}_1)$. Together with what was proved before, we have that

$$\lim_{n\to+\infty} W_{1,D_\infty}(\pi^n_\sharp \mathtt{m}_0, \pi^n_\sharp \mathtt{m}_1) = W_{1,D_\infty}(\mathtt{m}_0,\mathtt{m}_1). \tag{C.2}$$

Now, for any $x, y \in \mathbb{R}^n$ define $D_n(x,y) := \sup_{j\le n} |x_j - y_j| \wedge 1$ and notice that

$$W_{1,D_\infty}(\mathtt{m}_0,\mathtt{m}_1) = \sup_{n\in\mathbb{N}} W_{1,D_\infty}(\pi^n_\sharp \mathtt{m}_0, \pi^n_\sharp \mathtt{m}_1) = \sup_{n\in\mathbb{N}} W_{1,D_n}(\pi^n_\sharp \mathtt{m}_0, \pi^n_\sharp \mathtt{m}_1)$$
$$= \sup_{n\in\mathbb{N}} \sup_{\{\Psi\in C^1_c(\mathbb{R}^n) : \Psi \ D_n \ 1-\mathrm{Lip}\}} \left\{ \int_{\mathbb{R}^n} \Psi(x) d\pi^n_\sharp \mathtt{m}_0(x) - \int_{\mathbb{R}^n} \Psi(y) d\pi^n_\sharp \mathtt{m}_1(y) \right\}$$

$$= \sup_{n,\Psi \text{ as above}} \int \Psi(\pi^n(\mathrm{x}))d(\mathtt{m}_0 - \mathtt{m}_1)(\mathrm{x}).$$

The proof is then concluded by the fact that $\mathbb{R}^\infty \ni x \mapsto \Psi(\pi^n(x))$ is the general expression for a cylinder function from $\mathbb{R}^\infty \to \mathbb{R}$, and it is D_∞ 1-Lipschitz if and only if $\mathbb{R}^n \ni x \to \Psi(x)$ is D_n 1-Lipschitz. This is equivalent to ask that

$$\left\| \sum_{i=1}^n |\partial_i \Psi| \right\|_\infty \leq 1 \quad \text{and} \quad \operatorname{osc} \Psi := \max \Psi - \min \Psi \leq 1.$$

Moreover, up to considering a translation, we can substitute the condition on oscillation with $\|\Psi\|_\infty \leq 1/2$, which concludes the proof. □

In particular, the previous lemma characterizes Lipschitzianity with respect to D_∞ of cylinder functions over $\mathbb{R}^\infty$ through conditions on the function Ψ that represents it, that is

$$F = \Psi \circ \pi^n \text{ is } D_\infty \text{ 1-Lipschitz} \iff \left\| \sum_{i=1}^n |\partial_i \Psi| \right\|_\infty \leq 1 \text{ and } \operatorname{osc} \Psi \leq 1.$$

The following shows how to use the previous lemma to have a duality formula for $\hat{\mathcal{W}}_1$ using only duality with cylinder functions.

Proposition C.5. *The following duality formula holds for the distance $\hat{\mathcal{W}}_1 = W_{1,\hat{W}_1}$: for all $M, N \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ we have*

$$\hat{\mathcal{W}}_1(M,N) = \sup_{F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d)),\ \mathrm{Lip}(F) \leq 1} \int_{\mathcal{P}(\mathbb{R}^d)} F(\mu)dM(\mu) - \int_{\mathcal{P}(\mathbb{R}^d)} F(\nu)dN(\nu), \tag{C.3}$$

where the Lipschitz constant has to be intended w.r.t. the $\hat{\mathcal{W}}_1$ distance.

Proof. It follows easily by Lemma C.4, indeed, thanks to (C.1), we can conclude

$$\begin{aligned} W_{1,\hat{W}_1}(M,N) &= W_{1,D_\infty}(\iota_\sharp M, \iota_\sharp N) = \sup \int_{\mathbb{R}^\infty} F d(\iota_\sharp M - \iota_\sharp N) \\ &= \sup \left\{ \int_{\mathcal{P}(\mathbb{R}^d)} \Psi(L_{\phi_1}(\mu), \dots, L_{\phi_n}(\mu))dM(\mu) - \int_{\mathcal{P}(\mathbb{R}^d)} \Psi(L_{\phi_1}(\nu), \dots, L_{\phi_n}(\nu))dM(\nu) \right\}, \end{aligned}$$

noticing that if $F \in \mathrm{Cyl}^1(\mathbb{R}^\infty)$ and it is D_∞ 1-Lipschitz then, again by the fact that ι is an isometry, we have that $\Psi(L_{\phi_1}(\mu), \dots, L_{\phi_n}(\mu)) \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ is $\hat{\mathcal{W}}_1$ 1-Lipschitz. □

Remark C.6. *We actually proved a stronger result: the Wasserstein distance W_{1,W_1} between M and N can be recovered by the cylinder functions depending only on the functions $(L_{\phi_n})_{n\in\mathbb{N}}$, where $(\phi_n)_{n\in\mathbb{N}} \subset C^1_c(\mathbb{R}^d)$ are 1-Lipschitz functions from $\mathbb{R}^d$ to $\mathbb{R}$ (w.r.t. $|\cdot| \wedge 1$) such that*

$$W_{1,|\cdot|\wedge 1}(\mu,\nu) = \sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \phi_n(x)d(\mu - \nu)(x).$$

Moreover, the last equality in the proof of Lemma C.4

$$W_{1,D_\infty}(\mathtt{m}_0, \mathtt{m}_1) = \sup_{n,\Psi} \int \Psi(p^n(\mathrm{x}))d(\mathtt{m}_0 - \mathtt{m}_1)(\mathrm{x}),$$

is telling us that for any $n \in \mathbb{N}$, we can consider a countable family $\mathcal{F}_n$ of 1-Lipschitz functions (w.r.t. D_n) such that the sup *is realized taking $F \in \mathcal{F}_n$. Defining the countable family $\mathcal{F} := \cup \mathcal{F}_n$ and taking in considerations what has been said above, we have that*

$$\hat{\mathcal{W}}_1(M, N) = \sup_{F \in \mathcal{F}} \int_{\mathcal{P}(\mathbb{R}^d)} F(L_{\psi_1}(\mu), \ldots, L_{\psi_n}(\mu)) d(M - N)(\mu). \tag{C.4}$$

Lemma C.7. *Let $(M_t)_{t \in [0,T]} \subset \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ and $(B_t)_{t \in [0,T]}$ be an $L^1(M_t \otimes dt)$-derivation such that* (4.7) *holds. Then there exists a curve $t \mapsto \overline{M}_t$ such that $\overline{M}_t = M_t$ for a.e. $t \in [0, T]$ and $(\overline{M}_t) \in AC_T\big(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)), \hat{\mathcal{W}}_1\big) \subset C_T(\mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$.*

Proof. For any $F \in \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$, the map $[0, T] \ni t \mapsto \int F(\mu) dM_t(\mu)$ is $W^{1,1}(0, T)$ with distributional derivative $\int B_t[F](\mu) dM_t(\mu)$. In particular, there exists $I_F \subset (0, T)$ of full Lebesgue measure such that for all $s, t \in I_F$

$$\int F dM_t - \int F dM_s = \int_s^t \int B_r[F](\mu) dM_r(\mu) dr.$$

Consider $\mathcal{C} \subset \mathrm{Cyl}^1_c(\mathcal{P}(\mathbb{R}^d))$ the countable set in the supremum of (C.4) and take the full Lebesgue measure set $I = \cap_{F \in \mathcal{C}} I_F$. Then for any $s, t \in I$ we have

$$\begin{aligned}
\hat{\mathcal{W}}_1(M_t, M_s) &= \sup_{F \in \mathcal{C}} \int F\big(L_{\psi_1}, \ldots, L_{\psi_n}\big) d(M_t - M_s) \\
&= \sup_{F \in \mathcal{C}} \int_s^t \frac{d}{dr} \int F\big(L_{\psi_1}(\mu), \ldots, L_{\psi_n}(\mu)\big) dM_r(\mu) dr \\
&= \sup_{F \in \mathcal{C}} \int_s^t \int B_r[F](\mu) dM_r(\mu) dr \le \int_s^t \int c_r(\mu) dM_r dr,
\end{aligned}$$

where the last inequality follows from the fact that $|F(x) - F(y)| \le D_n(x, y) \le \sup_{i \le n} |x_i - y_i|$, which implies that the 1-norm in $\mathbb{R}^n$ of the gradient of F is bounded by 1, i.e.

$$\sum_{i=1}^n |\partial_i F(x)| \le 1 \quad \forall x \in \mathbb{R}^n.$$

This implies that $\forall x \in \mathbb{R}^d$ and $\forall \mu \in \mathcal{P}(\mathbb{R}^d)$

$$|\nabla_W F(x, \mu)| \le \sum_{i=1}^n |\partial_i F\big(L_{\psi_1}(\mu), \ldots, L_{\psi_n}(\mu)\big)| |\nabla \psi_i| \le \sum_{i=1}^n |\partial_i F\big(L_{\psi_1}(\mu), \ldots, L_{\psi_n}(\mu)\big)| \le 1,$$

following also from the fact that for all $n \in \mathbb{N}$, ψ_n is $|\cdot| \wedge 1$ 1-Lipschitz, so in particular they are 1-Lipschitz w.r.t. the usual Euclidean norm $|\cdot|$ in $\mathbb{R}^d$.
Then, there exists a curve of random measures $t \mapsto \overline{M}_t$ that is absolutely continuous w.r.t. $W_{1, \hat{W}_1}$ and $M_t = \overline{M}_t$ for all $t \in I$. □

Appendix D. Measurability in the spaces of measures

In this section, we state some general results of Borel measurability in the spaces $\mathcal{M}_+(Y)$, $\mathcal{M}(Y, \mathbb{R}^d)$ and their product, where Y is a Polish space. In $\mathcal{M}_+(Y)$ and $\mathcal{M}(Y, \mathbb{R}^d)$ we consider the weak topology in duality with C_b functions, and the product topology over product spaces.

D.1. **Equivalence between σ-algebras over $\mathcal{M}_+(Y)$.** Let (Y,τ) be a Polish space. The goal of this subsection is to prove that the Borel σ-algebra of $\mathcal{M}_+(Y)$, endowed with the narrow topology, is the same as the smallest σ-algebra $\mathcal{S}$ that makes measurable the evaluation on Borel sets, i.e. such that for any Borel set $A \subset Y$ the map $\mathcal{M}_+(Y) \ni \mu \mapsto \mu(A) \in \mathbb{R}$ is $\mathcal{S}$-measurable.

Lemma D.1 (Measurability of $\mu \mapsto \int g d\mu$)**.** *Let Y be a Polish space. For each Borel map $g : Y \to [0,+\infty]$*

$$G : \mathcal{M}_+(Y) \to [0,+\infty], \quad G(\mu) := \int_Y g \, d\mu$$

is a Borel function. In particular, the set

$$\mathcal{P}_g(Y) := \left\{ \mu \in \mathcal{P}(Y) \ : \ \int_Y g d\mu < +\infty \right\}$$

is a Borel subset of $\mathcal{P}(Y)$.

Proof. Define

$$\mathcal{H} := \{h : Y \to \mathbb{R} \ : \ h \text{ is Borel and bounded, } \mu \mapsto \int_Y h \, d\mu \text{ is Borel}\}.$$

Obviously $\mathcal{H}$ contains $C_b(Y)$. Moreover, $\mathcal{H}$ is closed under monotone limits, indeed if $\mathcal{H} \ni h_n \to h$ monotonically, then by dominated convergence theorem, for any $\mu \in \mathcal{M}_+(Y)$

$$H(\mu) := \int_Y h \, d\mu = \lim_{n \to +\infty} \int_Y h_n d\mu,$$

so H is the pointwise limit of Borel functions, thus it is Borel, which implies that $h \in \mathcal{H}$. Then we can apply [Bog07, Theorem 2.12.9, (iii)] to conclude that $h \in \mathcal{H}$ for any bounded and Borel function $h : Y \to \mathbb{R}$. Then, by monotone convergence theorem, we conclude that G is a Borel function approximating g pointwise with $g_n := g \wedge n$. □

Lemma D.2. *The narrow topology over $X := \mathcal{M}_+(Y)$ is metrizable.*

Proof. Thanks to [Bog07], Theorem 8.3.2, the narrow topology over $\mathcal{M}_+(Y)$ is induced by the norm

$$\|\mu\|_{BL} := \sup \left\{ \int_Y \phi d\mu \ : \ \phi \in \mathrm{Lip}_b(Y), \ \mathrm{LIP}(\phi) \leq 1 \right\}. \tag{D.1}$$

□

Before moving to the main goal of this section, we need a general lemma.

Lemma D.3. *Let Z be a Polish space. Then $\mathcal{B}(Z)$ is the smallest σ-algebra under which continuous functions from Z to $\mathbb{R}$ are measurable.*

Proof. Let $C \subset Z$ be a closed subset. Define the continuous function $\delta_C(z) := \inf\{d_Z(y,z) : y \in C\}$. Notice that $C = \delta_C^{-1}(\{0\})$. Let $\mathcal{S}$ be the smallest σ-algebra on Z such that all continuous functions from Z to $\mathbb{R}$ are measurable. For sure $\mathcal{S} \subset \mathcal{B}(Z)$. On the other hand, for any closed set $C \subset Z$, we have that $C = \delta_C^{-1}(\{0\}) \in \mathcal{S}$, so $\mathcal{B}(Z) \subset \mathcal{S}$. □

We will use this lemma with $Z = \mathcal{M}_+(Y)$, where Y is a Polish space. To this aim, define the class of functions

$$\mathcal{F} := \{F : \mathcal{M}_+(Y) \to \mathbb{R} \ : \ F(\mu) = \int f d\mu, \ f \in C_b(Y)\}.$$

Lemma D.4. *Let $\mathcal{C}$ be the collection of all narrowly continuous functions from $\mathcal{M}_+(Y)$ to $\mathbb{R}$. Let $\mathcal{S}_{\mathcal{F}}$ (resp. $\mathcal{S}_{\mathcal{C}}$) be the smallest σ-algebra over $\mathcal{M}_+(Y)$ that makes measurable the functions in $\mathcal{F}$ (resp. $\mathcal{C}$). Then $\mathcal{S}_{\mathcal{F}} = \mathcal{S}_{\mathcal{C}} = \mathcal{B}(\mathcal{M}_+(Y))$.*

Proof. The fact that $\mathcal{B}(\mathcal{M}_+(Y)) = \mathcal{S}_{\mathcal{C}}$ follows from the previous lemma. Regarding $\mathcal{S}_{\mathcal{F}}$, trivially we have $\mathcal{S}_{\mathcal{F}} \subset \mathcal{S}_{\mathcal{C}}$. On the other hand, notice that the countable class of functions $\mathcal{C}_0 \subset \mathcal{F}$, introduced in [AGS08, Remeark 5.1.1], is sufficient to describe the narrow topology $\mathcal{T}$ of $\mathcal{M}_+(Y)$. This means that $\mathcal{T}$ coincides with the smallest topology that makes continuous the functions in $\mathcal{C}_0$. Then

$$\mathcal{B}(\mathcal{M}_+(Y)) = \sigma(\tau) = \sigma\left(\left\{F^{-1}((a,b)) \ : \ F \in \mathcal{C}_0, \ a < b, \ a, b \in \mathbb{Q}\right\}\right) \subset \mathcal{S}_{\mathcal{F}}.$$

□

Proposition D.5. *Let Y be a Polish space and $\mathcal{S}$ be the smallest σ-algebra on $\mathcal{M}_+(Y)$ that makes measurable the functions $\mu \mapsto \mu(A)$ for any $A \in \mathcal{B}(Y)$. Then $\mathcal{S}$ coincides with $\mathcal{B}(\mathcal{M}_+(Y))$.*

Proof. Thanks to Lemma D.1, it holds that $\mathcal{S} \subset \mathcal{B}(\mathcal{M}_+(Y))$. On the other hand, the integral of step functions is $\mathcal{S}$-measurable, and then also $\mu \mapsto \int f d\mu$ is measurable for any $f \in C_b(Y)$. This implies that $\mathcal{S}_{\mathcal{F}} \subset \mathcal{S}$, and thanks to the previous lemma we are done. □

Corollary D.6. *Let X be a topological space and $X \ni x \mapsto \mu_x \in \mathcal{M}_+(Y)$ be a map taking values in $\mathcal{M}_+(Y)$. Then*

$$x \mapsto \mu_x \text{ is Borel } \iff x \mapsto \mu_x(A) \text{ is Borel } \forall A \in \mathcal{B}(Y).$$

This corollary directly shows the measurability of the family of measures given by the *disintegration theorem*, that we recall here for completeness.

Theorem D.7. *Let Y, X be Polish spaces, $\mu \in \mathcal{M}_+(Y)$ and $e : Y \to X$ a Borel function. Define $\theta := e_\sharp \mu \in \mathcal{M}_+(X)$. Then there exists a family $\{\mu_x\}_{x \in X} \subset \mathcal{P}(Y)$ such that*

(i) $x \mapsto \mu_x(A)$ is Borel measurable for any $A \in \mathcal{B}(Y)$;
(ii) $\mu(dz) = \int_X \mu_x(dy) d\theta(x)$;
(iii) μ_x is concentrated on $e^{-1}(\{x\})$ for θ-a.e. $x \in X$.

Moreover, such a disintegration is unique, in the sense that if another family $\{\mu'_x\}_{x \in X}$ satisfies these properties, then $\mu_x = \mu'_x$ for θ-a.e. $x \in X$.

D.2. **Measurability of sets and maps.**

Proposition D.8 (Measurability of $\mu \mapsto f_\sharp \mu$)**.** *Let X, Y be two Polish spaces and $f : X \to Y$ be a Borel measurable map. Then $f_\sharp : \mathcal{M}_+(X) \to \mathcal{M}_+(Y)$ is Borel measurable. The same holds for the restriction $f_\sharp : \mathcal{P}(X) \to \mathcal{P}(Y)$. If, additionally, f is continuous, we have that $f_\sharp : \mathcal{M}(X; \mathbb{R}^d) \to \mathcal{M}(Y; \mathbb{R}^d)$ is continuous, and in particular Borel measurable.*

Proof. By Corollary D.6, the map $\mu \mapsto f_\sharp \mu$ is measurable if and only if for all $B \in \mathcal{B}(Y)$ the map $\mu \mapsto f_\sharp \mu(B)$ is measurable. By Lemma D.1 we conclude, since $A = f^{-1}(B) \in \mathcal{B}(X)$ and

$$f_\sharp \mu(B) = \mu(f^{-1}(B)) = \int_X \mathbb{1}_A d\mu$$

is measurable. For the second part, simply notice that $f_\sharp|_{\mathcal{P}(X)}$ maps $\mathcal{P}(X)$ to $\mathcal{P}(Y)$.

Thanks to [Bré11, Proposition 3.2], we only need to show that $\mathcal{M}(X; \mathbb{R}^d) \ni \mu \mapsto \int_Y \phi d(f_\sharp \mu) \in \mathbb{R}$ is continuous for all $\phi \in C_b(Y; \mathbb{R}^d)$. We conclude noticing that $\int_Y \phi d(f_\sharp \mu) = \int_X \phi \circ f d\mu$, which is continuous since $\phi \circ f \in C_b(X; \mathbb{R}^d)$. □

Theorem D.9 (Measurability of $\mu \mapsto f\mu$)**.** *Let Y be a Polish space. Let $f : Y \to [0, +\infty]$ be a bounded and Borel map. Then the map*

$$F : \mathcal{M}_+(Y) \to \mathcal{M}_+(Y), \quad F(\mu) = f\mu,$$

is Borel, endowing $\mathcal{M}_+(Y)$ with the weak topology w.r.t. the duality with $C_b(Y)$ functions.

Proof. Thanks to Corollary D.6, it suffices to prove that for all $A \in \mathcal{B}(Y)$, the map $\mu \mapsto \int_A f d\mu$ is measurable. We are done by Lemma D.1, with $g(y) = f(y)\mathbb{1}_A(y)$. □

Corollary D.10. *Assume that $f : Y \to [0, +\infty]$ is Borel. Then the map*

$$F : \mathcal{M}_{+,f}(Y) \to \mathcal{M}_+(Y), \quad F(\mu) = f\mu,$$

is Borel, where $\mathcal{M}_{+,f}(Y) := \{\mu \in \mathcal{M}_+(Y) \ : \ f \in L^1(\mu)\}$ is endowed with the subspace topology.

Proof. We know that $\mathcal{M}_{+,f}(Y)$ is a Borel set. Define $F_k(\mu) := (f \wedge k)\mu$ for all $\mu \in \mathcal{M}_{+,f}(Y)$, $k \in \mathbb{N}$. Then F_k pointwise converges to F; indeed, $(f \wedge k)\mu \rightharpoonup f \wedge \mu$ for all $\mu \in \mathcal{M}_{+,f}(Y)$. □

Corollary D.11. *Assume $f : Y \to \overline{R}$ is Borel. Then the following map is Borel*

$$F : \mathcal{M}_{+,f}(Y) \to \mathcal{M}(Y), \quad F(\mu) = f\mu.$$

Proof. Notice that for all $\mu \in \mathcal{M}_{+,f}(Y)$, $F(\mu) = F_+(\mu) - F_-(\mu)$, where $F_\pm(\mu) = f_\pm \mu$. □

Corollary D.12. *Assume $f : Y \to \mathbb{R}^d$ is Borel. Then the following map is Borel*

$$F : \mathcal{M}_{+,f}(Y) \to \mathcal{M}(Y; \mathbb{R}^d), \quad F(\mu) = f\mu.$$

Proof. Notice that $\mathcal{M}(Y; \mathbb{R}^d)$ is homeomorphic to $\mathcal{M}(Y)^d$. □

Lemma D.13 (Measurability of the condition $\nu = f\mu$)**.** *Let Y be a Polish space. For each Borel map $f : Y \to \mathbb{R}^d$ and any $p \geq 1$, the following set is Borel measurable:*

$$\mathfrak{D}_f := \{(\mu, \nu) \in \mathcal{M}_+(Y) \times \mathcal{M}(Y; \mathbb{R}^d) \ : \ f \in L^p(\mu), \ \nu = f\mu\}.$$

Proof. The condition $f \in L^p(\mu)$ is a Borel condition thanks to the previous lemma.
Regarding the condition $\nu = f\mu$, thanks to the equivalence (i) $\iff$ (ii) in [Bog07, Theorem 8.10.39], we have that $\mathcal{M}(Y, \mathbb{R}^d)$ is countably separated, i.e. there exists a countable set of continuous and bounded functions $\{h_n : Y \to \mathbb{R}\}$ such that for all $\nu_1, \nu_2 \in \mathcal{M}(X, \mathbb{R}^d)$

$$\int_X h_n \cdot d\nu_1 = \int_X h_n \cdot d\nu_2 \ \forall n \in \mathbb{N} \implies \nu_1 = \nu_2.$$

Then, we can rewrite

$$\mathfrak{D}_f = \{(\mu, \nu) \in \mathcal{M}_+(Y) \times \mathcal{M}(Y; \mathbb{R}^d) \ : \ f \in L^p(\mu), \ \int_Y h_n \cdot d\nu = \int_Y h_n \cdot f \ d\mu \ \forall n\}.$$

Notice that the function $\mathcal{M}(Y, \mathbb{R}^d) \ni \nu \mapsto \int h_n \cdot d\nu$ is continuous, while on the function $\mathcal{M}_+(X) \ni \mu \mapsto \int h_n \cdot f \ d\mu$ some comments must be done: written like this it is not well defined, because f could be not bounded. Replace it with the function

$$H_{n,f}(\mu) := \begin{cases} \int (h_n \cdot f)_+ - (h_n \cdot f)_- \ d\mu & \text{if } f \in L^1(\mu) \\ +\infty & \text{otherwise} \end{cases}$$

This is a Borel function, because $\{\mu \ : \ f \in L^1(\mu)\}$ is a Borel set and thanks to the previous lemma the two functions $\int (h_n \cdot f)_\pm \ d\mu$ are Borel and we are done. □

Proposition D.14 (Measurability of the condition $\nu \ll \mu$)**.** *Let Y be Polish. Then*

$$\{(\mu,\nu) \in \mathcal{M}_+(Y) \times \mathcal{M}_+(Y) \ : \nu \ll \mu\}$$

is a Borel subset of $\mathcal{M}_+(Y) \times \mathcal{M}_+(Y)$.

Proof. The proof follows the same line of Lemma C.1, with some more refinements. Recall that the condition $\nu \ll \mu$ is equivalent to

$$\forall \varepsilon > 0 \exists \delta > 0 \ : \ \forall B \in \mathcal{B}(X),\ \mu(B) \le \delta \implies \nu(B) \le \varepsilon.$$

Thanks to the outer regularity of any finite Borel measure, it is equivalent to require

$$\forall \varepsilon > 0 \exists \delta > 0 \ : \ \forall A \subset X \text{ open},\ \mu(B) \le \delta \implies \nu(B) \le \varepsilon.$$

Define the functional

$$\mathcal{F}_\delta(\nu|\mu) := \sup_{A \subset X \text{ open}, \mu(A) \le \delta} \nu(A),$$

and notice that, thank to what has already been said,

$$\nu \ll \mu \iff \inf_{n \in \mathbb{N}} \mathcal{F}_{\frac{1}{n}}(\nu|\mu) = 0.$$

Then, we are done if we prove that $(\mu,\nu) \mapsto \mathcal{F}_\delta(\nu|\mu)$ is Borel. We prove that, for all $\delta > 0$, such function is actually l.s.c. in the couple (μ,ν). Notice that such functional can be rewritten as

$$\mathcal{F}_\delta(\nu|\mu) = \sup_{A \subset X \text{ open}} \left(\nu(A) + \chi_{(-\infty,\delta]}(\mu(A))\right), \qquad \chi_I(x) = \begin{cases} -\infty & \text{if } x \notin I \\ 0 & \text{if } x \in I \end{cases}$$

Now notice that

- $\nu \mapsto \nu(A)$ is l.s.c. for any $A \subset X$ open:
- the function $x \mapsto \chi_{(-\infty,\delta]}(x)$ is l.s.c. and non-decrasing. This, together with the lower semi-continuity of $\mu \mapsto \mu(A)$, implies that $\mu \mapsto \chi_{(\infty,\delta]}(\mu(A))$ is lower semicontinuous.

Then, $\mathcal{F}_\delta(\nu|\mu)$ is the supremum of l.s.c. functionals, which implies that it is l.s.c. as well. In particular, it is a Borel map. □

Corollary D.15. *Given X Polish, the set*

$$\{(\mu,\nu) \in \mathcal{M}_+(X) \times \mathcal{M}(X;\mathbb{R}^d) \ : \nu \ll \mu\}$$

is a Borel subset of $\mathcal{M}_+(X) \times \mathcal{M}(X;\mathbb{R}^d)$.

Proof. The condition $\nu \ll \mu$ is equivalent to $|\nu| \ll \mu$, where ν is the total variation measure of ν. We conclude thanks to the previous Lemma and the fact that $\mathcal{M}(X;\mathbb{R}^d) \ni \nu \mapsto |\nu| \in \mathcal{M}_+(X)$ is Borel (see Remark 2.4, [AILP24]). □

References

[ABS24] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on optimal transport.* Springer, 2024. ISBN: 978-3-031-76833-0. DOI: 10.1007/978-3-031-76834-7.

[AC08] Luigi Ambrosio and Gianluca Crippa. "Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields". In: *In: Transport Equations and Multi-D Hyperbolic Conservation Laws, Lecture Notes of the Unione Matematica Italiana* 5 (2008). DOI: 10.1007/978-3-540-76781-7_1.

[AC14] Luigi Ambrosio and Gianluca Crippa. "Continuity equations and ODE flows with non-smooth velocity". In: *Proceedings of the Royal Society of Edinburgh: Section A Mathematics* 144.6 (2014), pp. 1191–1244. DOI: 10.1017/S0308210513000085.

[AF09] Luigi Ambrosio and Alessio Figalli. "On flows associated to Sobolev vector fields in Wiener spaces: An approach à la DiPerna–Lions". In: *Journal of Functional Analysis* 256.1 (2009), pp. 179–214. ISSN: 0022-1236. DOI: https://doi.org/10.1016/j.jfa.2008.05.007.

[AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. "Functions of Bounded Variation". In: *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, Mar. 2000. ISBN: 9780198502456. DOI: 10.1093/oso/9780198502456.003.0003.

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Second. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008, pp. x+334. ISBN: 978-3-7643-8721-1.

[AGS13] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. "Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces". In: *Rev. Mat. Iberoam.* 29.3 (2013), pp. 969–996. ISSN: 0213-2230. DOI: 10.4171/RMI/746.

[AGS14] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below". In: *Invent. Math.* 195.2 (2014), pp. 289–391. ISSN: 0020-9910. DOI: 10.1007/s00222-013-0456-1.

[AILP24] Luigi Ambrosio, Toni Ikonen, Danka Lučić, and Enrico Pasqualetto. *Metric Sobolev spaces I: equivalence of definitions*. cvgmt preprint, http://cvgmt.sns.it/paper/6545/. 2024.

[AKPR25] Beatrice Acciaio, Daniel Kršek, Gudmund Pammer, and Marco Rodrigues. "Absolutely Continuous Curves of Stochastic Processes". In: *arXiv preprint arXiv:2506.13634* (2025).

[Amb04] Luigi Ambrosio. "Transport equation and Cauchy problem for BV vector fields". In: *Invent. Math.* 158.2 (2004), pp. 227–260. ISSN: 0020-9910. DOI: 10.1007/s00222-004-0367-2.

[AT14] Luigi Ambrosio and Dario Trevisan. "Well-posedness of Lagrangian flows and continuity equations in metric measure spaces". In: *Anal. PDE* 7.5 (2014), pp. 1179–1234. ISSN: 2157-5045. DOI: 10.2140/apde.2014.7.1179.

[BB00] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numerische Mathematik* 84.3 (2000), pp. 375–393. DOI: 10.1007/s002119900117.

[BBP24] Daniel Bartl, Mathias Beiglböck, and Gudmund Pammer. "The Wasserstein space of stochastic processes". In: *J. Eur. Math. Soc.* (2024). DOI: 10.4171/JEMS/1554.

[BCD21] Elia Brué, Maria Colombo, and Camillo De Lellis. "Positive solutions of transport equations and classical nonuniqueness of characteristic curves". In: *Archive for Rational Mechanics and Analysis* 240 (2021), pp. 1055–1090. DOI: 10.1007/s00205-021-01628-5.

[BCK24] Elia Bruè, Maria Colombo, and Anuj Kumar. "Sharp Nonuniqueness in the Transport Equation with Sobolev Velocity Field". In: (2024). arxiv preprint. DOI: 10.48550/arXiv.2405.01670.

[BF21] Benoît Bonnet-Weill and Hélène Frankowska. "Differential inclusions in Wasserstein spaces: the Cauchy-Lipschitz framework". In: *J. Differential Equations* 271 (2021), pp. 594–637. ISSN: 0022-0396. DOI: 10.1016/j.jde.2020.08.031.

[BF24] Benoît Bonnet-Weill and Hélène Frankowska. "Carathéodory theory and a priori estimates for continuity inclusions in the space of probability measures". In: *Nonlinear Anal.* 247 (2024), Paper No. 113595, 32. ISSN: 0362-546X,1873-5215. DOI: 10.1016/j.na.2024.113595.

[Bog07] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007, Vol. I: xviii+500 pp., Vol. II: xiv+575. ISBN: 978-3-540-34513-8; 3-540-34513-2. DOI: 10.1007/978-3-540-34514-5.

[BPS25] Mathias Beiglböck, Gudmund Pammer, and Stefan Schrott. "A Brenier Theorem on $(P_2(\dots P_2(H)\dots), W_2)$ and Applications to Adapted Transport". In: *arXiv preprint arXiv:2509.03506* (2025).

[Bré11] Haim Brézis. *Functional analysis, Sobolev spaces and partial differential equations.* Vol. 2. Springer, 2011. ISBN: 978-0-387-70913-0. DOI: 10.1007/978-0-387-70914-7.

[BVK25] Clément Bonet, Christophe Vauthier, and Anna Korba. "Flowing Datasets with Wasserstein over Wasserstein Gradient Flows". In: (2025). arxiv preprint. DOI: https://doi.org/10.48550/arXiv.2506.07534.

[CD08] Gianluca Crippa and Camillo De Lellis. "Estimates and regularity results for the DiPerna-Lions flow". In: *J. Reine Angew. Math.* (2008). DOI: 10.1515/CRELLE.2008.016.

[CD18] René Carmona and François Delarue. *Probabilistic theory of mean field games with applications I-II.* Vol. 3. Springer, 2018. ISBN: 978-3-030-13259-0. DOI: https://doi.org/10.1007/978-3-319-56436-4.

[CL24] Marta Catalano and Hugo Lavenant. "Hierarchical Integral Probability Metrics: A distance on random probability measures with low sample complexity". In: *ICML'24: Proceedings of the 41st International Conf. on Machine Learning* (2024).

[CLOS22] Giulia Cavagnari, Stefano Lisini, Carlo Orrieri, and Giuseppe Savaré. "Lagrangian, Eulerian and Kantorovich formulations of multi-agent optimal control problems: Equivalence and Gamma-convergence". In: *J. Differential Equations* 322 (2022), pp. 268–364. ISSN: 0022-0396. DOI: 10.1016/j.jde.2022.03.019.

[CSS23] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. "Dissipative probability vector fields and generation of evolution semigroups in Wasserstein spaces". In: *Probab. Theory Related Fields* 185.3-4 (2023), pp. 1087–1182. ISSN: 0178-8051,1432-2064. DOI: 10.1007/s00440-022-01148-7.

[CSS25a] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. *A Lagrangian approach to totally dissipative evolutions in Wasserstein spaces.* arxiv preprint, https://arxiv.org/abs/2305.05211. 2025.

[CSS25b] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. "Extension of monotone operators and Lipschitz maps invariant for a group of isometries". In: *Canad. J. Math.* 77.1 (2025), pp. 149–186. ISSN: 0008-414X,1496-4279. DOI: 10.4153/S0008414X23000846.

[Del22] Lorenzo Dello Schiavo. "The Dirichlet–Ferguson diffusion on the space of probability measures over a closed Riemannian manifold". In: *The Annals of Probability* 50.2 (2022), pp. 591–648.

[Del24] Lorenzo Dello Schiavo. "Massive particle systems, Wasserstein Brownian motion and the Dean-Kawasaki equation". In: *to appear* (2024+).

[Di 14] Simone Di Marino. "Recent advances on BV and Sobolev Spaces in metric measure spaces". In: (2014). PhD thesis, https://cvgmt.sns.it/paper/2568/.

[DL89] R. J. DiPerna and P.-L. Lions. "Ordinary differential equations, transport theory and Sobolev spaces". In: *Invent. Math.* 98.3 (1989), pp. 511–547. ISSN: 0020-9910. DOI: 10.1007/BF01393835.

[EP25] Pedram Emami and Brendan Pass. "Optimal transport with optimal transport cost: the Monge-Kantorovich problem on Wasserstein spaces". In: *Calc. Var. Partial Differential Equations* 64.2 (2025), Paper No. 43, 11. ISSN: 0944-2669,1432-0835. DOI: 10.1007/s00526-024-02905-3.

[Fig08] Alessio Figalli. "Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients". In: *Journal of Functional Analysis* 254.1 (2008), pp. 109–153. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2007.09.020.

[HM25] Martin Huesmann and Bastian Müller. "A Benamou–Brenier formula for transport distances between stationary random measures". In: *Stochastic Processes and their Applications* 185 (2025), p. 104633. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2025.104633.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. "The Variational Formulation of the Fokker–Planck Equation". In: *SIAM Journal on Mathematical Analysis* 29.1 (1998), pp. 1–17. DOI: 10.1137/S0036141096303359.

[Lis07] Stefano Lisini. "Characterization of absolutely continuous curves in Wasserstein spaces". In: *Calc. Var. Partial Differential Equations* 28.1 (2007), pp. 85–120. ISSN: 0944-2669.

[LSZ22] Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. "Superposition and mimicking theorems for conditional McKean–Vlasov equations". In: *Journal of the European Mathematical Society* 25.8 (2022), pp. 3229–3288. DOI: 10.4171/JEMS/1266.

[Ngu16] XuanLong Nguyen. "Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure". In: *Bernoulli* (2016). DOI: 10.3150/15-BEJ703.

[Pin25a] Alessandro Pinzi. "A study of the metric measure space of probability measures via a purely atomic superposition principle". In: *arXiv, arxiv.org/abs/2511.21204* (2025).

[Pin25b] Alessandro Pinzi. "First order equation on random measures as superposition of weak solutions to the McKean-Vlasov equation". In: *arXiv, arxiv.org/abs/2510.07542* (2025).

[PS12] Emanuele Paolini and Eugene Stepanov. "Decomposition of acyclic normal currents in a metric space". In: *J. Funct. Anal.* 263.11 (2012), pp. 3358–3390. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2012.08.009.

[PS13] Emanuele Paolini and Eugene Stepanov. "Structure of metric cycles and normal one-dimensional currents". In: *J. Funct. Anal.* 264.6 (2013), pp. 1269–1295. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2012.12.007.

[PS25] Alessandro Pinzi and Giuseppe Savaré. "Totally convex functions and solution to the L^2-Monge problem for random measures". In: *arXiv, arxiv.org/abs/2510.07542* (2025).

[Sch73] L Schwartz. "Radon measures on arbitrary topological spaces and cylindrical measures". In: *Oxford Univ.* (1973).

[Smi93] S. K. Smirnov. "Decomposition of solenoidal vector charges into elementary solenoids, and the structure of normal one-dimensional flows". In: *Algebra i Analiz* 5.4 (1993), pp. 206–238. ISSN: 0234-0852.

[Sod23] Giacomo Enrico Sodini. "The general class of Wasserstein Sobolev spaces: density of cylinder functions, reflexivity, uniform convexity and Clarkson's inequalities". In: *Calculus of Variations and Partial Differential Equations* 62.7 (2023), p. 212. DOI: 10.1007/s00526-023-02543-1.

[Sri08] Sashi Mohan Srivastava. *A course on Borel sets*. Vol. 180. Springer Science & Business Media, 2008. ISBN: 978-0-387-98412-4. DOI: 10.1007/b98956.

[ST17] Eugene Stepanov and Dario Trevisan. "Three superposition principles: currents, continuity equations and curves of measures". In: *J. Funct. Anal.* 272.3 (2017), pp. 1044–1103. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2016.10.025.

[Tre15] Dario Trevisan. "Lagrangian flows driven by BV fields in Wiener spaces". In: *Probability Theory and Related Fields* 163.1 (2015), pp. 123–147. DOI: 10.1007/s00440-014-0589-1.

[Tre16] Dario Trevisan. "Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients". In: *Electronic Journal of Probability* 21 (2016), pp. 1–41. DOI: 10.1214/16-EJP4453.

[Wea00] Nik Weaver. "Lipschitz Algebras and Derivations II. Exterior Differentiation". In: *Journal of Functional Analysis* 178.1 (2000), pp. 64–112. ISSN: 0022-1236. DOI: https://doi.org/10.1006/jfan.2000.3637.

(Alessandro Pinzi) Università Commerciale Luigi Bocconi, Dipartimento di Scienze delle Decisioni, via Roentgen 1, 20136 Milano, Italy
e-mail address: alessandro.pinzi@phd.unibocconi.it
Orcid: https://orcid.org/0009-0007-9146-5434

(Giuseppe Savaré) Università Commerciale Luigi Bocconi, Dipartimento di Scienze delle Decisioni and BIDSA,
via Roentgen 1, 20136 Milano, Italy
e-mail address: giuseppe.savare@unibocconi.it
Orcid: https://orcid.org/0000-0002-0104-4158