DIFFERENTIABLE MAPS BETWEEN WASSERSTEIN SPACES

Bernadette Lessel
Max Planck Institute for the History of Science
Berlin, Germany
blessel@mpiwg-berlin.mpg.de

Thomas Schick
Mathematical Institute
University of Göttingen
thomas.schick@math.uni-goettingen.de

October 5, 2020

ABSTRACT

A notion of differentiability for maps $F : W_2(M) \longrightarrow W_2(N)$ between Wasserstein spaces of order 2 is being proposed, where M and N are smooth, connected and complete Riemannian manifolds. Due to the nature of the tangent space construction on Wasserstein spaces, we only give a global definition of differentiability, i.e. without a prior notion of pointwise differentiability. With our definition, however, we recover the expected properties of a differential. Special focus is being put on differentiability properties of maps of the form $F = f_\#$, $f : M \longrightarrow N$ and on convex mixing of differentiable maps, with an explicit construction of the differential.

Contents

1 Introduction

Fundamental work has been done on the weak Riemannian manifold structure and second order analysis on Wasserstein spaces $W_2(M)$, most notably by Felix Otto [Ott01], John Lott [Lot07] and Nicola Gigli [Gig12]. However, to our knowledge, no notion of differentiability for maps between Wasserstein spaces has been proposed in the literature yet.

We begin with a reminder of Wasserstein spaces and its weak differentiable structure, to motivate the definitions we make later on. Our notion of differentiability for maps $F : W_2(M) \to W_2(N)$ between Wasserstein spaces is a global one, in the sense that it does not use a pointwise notion of differentiability. It seems to be the case that the latter is not possible in an immediate way due to the way tangent spaces are constructed in Wasserstein geometry: The basis for talking about tangent vectors along curves in $W_2(M)$ is constituted by the weak continuity equation $\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$, which can be seen as a differential characterization of absolutely continuous curves in $W_2(M)$ (see Theorem 9). The curve of minimal vector fields v_t that solves the continuity equation for an absolutely continuous curve μ_t is then seen as being tangential along μ_t. However, v_t is only defined for almost every t, so that a pointwise evaluation is not meaningful and therefore undermines the definition of a pointwise notion of differentiability in our approach. The differential of a map is, however, defined in a pointwise manner.

Our account on differentiable maps between Wasserstein spaces begins with the definition of *absolutely continuous maps* which map absolutely continuous curves to absolutely continuous curves. This definition is made in analogy to the theorem in differential geometry that a map $f : M \to N$ is differentiable if and only if it maps differentiable curves to differentiable curves. Absolutely continuous maps serve as a pre-notion to differentiability. An absolutely continuous map $F : W_2(M) \to W_2(N)$ is then said to be *differentiable* if every $\mu \in W_2(M)$ there exists a bounded linear map dF_μ between the tangent space at μ and the tangent space at $F(\mu)$ such that for every absolutely continuous curve μ_t the image curve $dF_{\mu_t}(v_t)$ of the curve of tangent vector fields v_t along μ_t is a curve of tangent vector fields along $F(\mu_t)$ (Definition 27). The collection of all these dF_μ, in the sense of a bundle map between tangent bundles, is then called the *differential* dF of F.

We show that dF unique up to a redefinition on a *negligible* set. Also, the usual properties of the differential are derived, such as the expected differential of the constant and of the indentity mapping, also of the composition of two differentiable maps and of the inverse of a differentiable map.

Special attention is payed to maps of the form $F = f_\#$, where measures are mapped to their image-measure with respect to $f : M \to N$, f being smooth and proper and where $\sup_{x \in M} \|df_x\| < \infty$. Maps of this kind are absolutely continuous, and an explicit formula is derived for a curve of vector fields satisfying the continuity equation together with $F(\mu_t)$, where μ_t is absolutely continuous. Unfortunately, it is not true in general that this curve of vector fields is actually tangent to μ_t, i.e. minimal. To enforce that, one can, however, apply a projector onto the respective tangent spaces, for almost every t, which in particular guarantees the existence of a differential for F.

Further focus is being put on the treatment of differentiability properties of convex mixings of maps between Wasserstein spaces, as they provide a class of non-trivial maps which are not given by a pushfoward of measures.

For background knowledge on Wasserstein geometry and optimal transport we refer to [AG13] and [Vil08].

2 Wasserstein geometry

Wasserstein geometry is a dynamical structure on Wasserstein spaces, which basically are sets of probability measures together with the Wasserstein distance.

Let thus (X, d) be a Polish space, where d metrizes the topology of X, and $\mathcal{P}(X)$ the set of all probability measures on X with respect to the Borel σ-algebra $\mathcal{B}(X)$. Instead of (X, d) we will often just write X. A measurable map between two Polish spaces $T : X \to Y$ induces a map between the respective spaces of probability measures via the *pushforward* $T_\#$ of measures: $T_\# : \mathcal{P}(X) \to \mathcal{P}(Y)$, $\mu \mapsto T_\#\mu$, where $T_\#\mu(A) := \mu(T^{-1}(A))$, for $A \in \mathcal{B}(Y)$. The *support* of a measure μ is defined by $supp(\mu) := \{x \in X \mid \text{ every open neighbourhood of } x \text{ has positive } \mu\text{-measure}\}$. The Lebesgue measure on $\mathbb{R}^n$ is denoted by λ.

2.1 Wasserstein spaces $W_p(X)$

We denote the set of probability measures which have finite p-th moment by $\mathcal{P}_p(X)$, where $p \in [1, \infty)$:

$$\mathcal{P}_p(X) := \{\mu \in \mathcal{P}(X) \mid \int_X d^p(x_0, x)\, d\mu(x) < \infty\}.$$

Note that $\mathcal{P}_p(X)$ is independent of the choice of $x_0 \in X$. Furthermore, we define

$$Adm(\mu, \nu) := \{\gamma \in \mathcal{P}(X \times Y) \mid \pi^X_\# \gamma = \mu, \pi^Y_\# \gamma = \nu\},$$

the so called *admissible transport plans* between μ and ν. Here, $\pi^X : X \times Y \to X$, $\pi^X(x, y) = x$, similarly π^Y.

Definition 1 (***Wasserstein distances and Wasserstein spaces***)**.** Let (X,d) be a Polish space and $p \in [0,\infty)$, then

$$\begin{aligned} W_p : \mathcal{P}_p(X) \times \mathcal{P}_p(X) &\to X \\ (\mu,\nu) &\mapsto \left(\inf_{\gamma \in Adm(\mu,\nu)} \int_{X\times X} d^p(x,y)\, d\gamma(x,y) \right)^{1/p} \end{aligned}$$

is called the *p-th Wasserstein distance*, or Wasserstein distance of order p. The tuple $(\mathcal{P}_p(X), W_p)$ is called *Wasserstein space* and is denoted by the symbol $W_p(X)$.

The fact that W_p is indeed a metric distance is a problem treated in optimal transport, where it is established that a minimizer for

$$\inf_{\gamma \in Adm(\mu,\nu)} \int_{X\times X} d^p(x,y)\, d\gamma(x,y)$$

actually exists. Such a minimizer is called *optimal transport plan*. In case a plan $\gamma \in Adm(\mu,\nu)$ is induced by a measurable map $T : X \to Y$, i.e. in case $\gamma = (Id, T)_\# \mu$, T is called *transport map*. Then, $T_\# \mu = \nu$.

One can show that $W_p(X)$ is complete and separable. Furthermore, W_p metrizes the weak convergence in $\mathcal{P}_p(X)$.

Definition 2 (***Weak convergence in*** $\mathcal{P}_p(X)$)**.** A sequence $(\mu_k)_{k\in\mathbb{N}} \subset \mathcal{P}(X)$ is said to *converge weakly* to $\mu \in \mathcal{P}_p(X)$ if and only if $\int \varphi d\mu_k \to \int \varphi d\mu$ for any bounded continuous function φ on X. This is denoted by $\mu_k \rightharpoonup \mu$. A sequence $(\mu_k)_{k\in\mathbb{N}} \subset \mathcal{P}_p(X)$ is said to converge weakly to $\mu \in \mathcal{P}_p(X)$ if and only if for $x_0 \in X$ it is:

1) $\mu_k \rightharpoonup \mu$ and
2) $\int d^p(x_0,x) d\mu_k(x) \to \int d^p(x_0,x) d\mu(x)$.

This is denoted by $\mu_k \rightharpoonup \mu$.

An important class of curves in Wasserstein space that we will need later on are constant speed geodesics.

Definition 3 (***Constant speed geodesic***)**.** A curve $(\gamma_t)_{t\in[0,1]}$, $\gamma_0 \neq \gamma_1$, in a metric space (X,d) is called a *constant speed geodesic* or *metric geodesic* in case that

$$d(\gamma_t, \gamma_s) = |t-s| d(\gamma_0,\gamma_1) \;\; \forall t,s \in [0,1]. \tag{1}$$

We will often abbreviate curves $(\gamma_t)_{t\in[0,1]}$ by writing γ_t instead.

Definition 4 (***Geodesic space***)**.** A metric space (X,d) is called *geodesic* if for every $x,y \in X$ with $x \neq y$, there exists a constant speed geodesic γ_t with $\gamma_0 = x$ and $\gamma_1 = y$.

If (X,d) is geodesic, then $W_2(X)$ is geodesic as well ([AG13]).

2.2 The continuity equation on $W_2(M)$

In the upcoming section, we will only be concerned with $W_2(M)$, where M is a smooth, connected and complete Riemannian manifold with Riemannian metric tensor h and associated Riemannian measure μ. We will often write $W(M)$ instead of $W_2(M)$. Furthermore, we equip the set of measurable sections of TM, which we will denote by $\Gamma(TM)$, with an L^2-topology. That means, for $v \in \Gamma(TM)$ we define

$$\|v\|_{L^2(\mu)} := \sqrt{\int_M h(v,v)\, d\mu}$$

and

$$L^2(TM,\mu) := \{ v \in \Gamma(TM) \mid \|v\|_{L^2(\mu)} < \infty \} /_\sim .$$

Here, two vector fields are considered to be equivalent in case they differ only on a set of μ-measure zero. $L^2(TM,\mu)$ is a Hilbert space with the canonical scalar product. We will often write $L^2(\mu)$ if it is clear to which manifold M it is referred to.

The (infinite dimensional) manifold structure that is commonly used on $W(M)$ is not a smooth structure in the sense of e.g. [KM97] where infinite dimensional manifolds are modeled on convenient vector spaces. The differentiable structure on $W(M)$, that will be introduced below, rather consists of ad hoc definitions accurately tailored to optimal transport and the Wasserstein metric structure which only mimic conventional differentiable and Riemannian behavior.

Instead of starting with a smooth manifold structure, on Wasserstein spaces one starts with the notion of a tangent space. Traditionally, the basic idea of a tangent vector at a given point is that it indicates the direction a (smooth) curve will be going infinitesimally from that point. Then, the set of all such vectors which can be found to be tangent to some curve at a given fixed point are collected in the tangent space at that point. On $W(M)$, however, there is no notion of smooth curves. But there is a notion of metric geodesics. In case the transport plan for the optimal transport between two measures is induced by a map T, the interpolating geodesic on Hilbert spaces can be written as $\mu_t = ((1-t)Id + tT)_{\#}\mu_0$, thus being of the form $\mu_t = F_{t\#}\mu_0$. More generally, on Riemannian manifolds optimal transport between μ_0 and μ_t can be achieved by $\mu_t = F_{t\#}\mu_0$, $F_t = \exp(t\nabla\varphi)$(see e.g [Vil08], Chapter 12). In these cases, F_t is injective and locally Lipschitz for $0 < t < 1$ ([Vil03], Subsubsection 5.4.1). It is known from the theory of characteristics for partial differential equations that curves of this kind solve the weak continuity equation, together with the vector field to which integral lines F_t corresponds.

Definition 5 (***Continuity equation***)**.** Given a family of vector fields $(v_t)_{t\in[0,T]}$, a curve $\mu_t : [0,T] \to W_2(M)$ is said to solve the *weak continuity equation*

$$\partial_t\mu_t + \nabla\cdot(v_t\mu_t) = 0, \tag{2}$$

if

$$\int_0^T \int_M \left(\frac{\partial}{\partial t}\varphi(x,t) + h(\nabla\varphi(x,t), v_t(x))\right) d\mu_t(x)dt = 0 \tag{3}$$

holds true for all $\varphi \in C_c^\infty((0,T)\times M)$.

Theorem 6 (***[Vil03], Theorem 5.34***)**.** *Let $(F_t)_{t\in[0,T)}$ be a family of maps on M such that $F_t : M \to M$ is a bijection for every $t \in [0,T)$, $F_0 = Id$ and both $(t,x) \mapsto F_t(x)$ and $(t,x) \mapsto F_t^{-1}(x)$ are locally Lipschitz on $[0,T)\times M$. Let further $v_t(x)$ be a family of velocity fields on M such that its integral lines correspond to the trajectories F_t, and μ be a probability measure. Then $\mu_t = F_{t\#}\mu$ is the unique weak solution in $\mathcal{C}([0,T),\mathcal{P}(M))$ of $\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0$ with initial condition $\mu_0 = \mu$. Here, $\mathcal{P}(M)$ is equipped with the weak topology.*

It is possible to characterize the class of curves on $W(M)$ that admit a velocity in the manner of Definition 5 ([AG13]) in the following way.

Definition 7 (***Absolutely continuous curve***)**.** Let (E,d) be an arbitrary metric space and I an interval in $\mathbb{R}$. A function $\gamma : I \to E$ is called *absolutely continuous (a.c.)*, if there exists a function $f \in L^1(I)$ such that

$$d(\gamma(t),\gamma(s)) \le \int_t^s f(r)dr, \quad \forall s,t \in I, t \le s. \tag{4}$$

Definition 8 (***Metric derivative***)**.** The *metric derivative* $|\dot\gamma|(t)$ of a curve $\gamma : [0,1] \to E$ at $t \in (0,1)$ is given as the limit

$$|\dot\gamma|(t) = \lim_{h\to 0}\frac{d(\gamma(t+h),\gamma(t))}{|h|}. \tag{5}$$

Every constant speed geodesic is absolutely continuous and $|\dot\gamma|(t) = d(\gamma(0),\gamma(1))$.

It is known that for absolutely continuous curves γ, the metric derivative exists for a.e. t. It is an element of $L^1(0,1)$ and, up to sets of zero Lebesgue-measure, the minimal function satisfying equation (4) for γ. In this sense absolutely continuous functions enable a generalization of the fundamental theorem of calculus to arbitrary metric spaces.

Theorem 9 (***Differential characterization of a.c. curves***)**.** *Let $\mu_t : [0,1] \to W_2(M)$ be an a.c. curve. Then there exists a Borel family of vector fields $(v_t)_{t\in[0,1]}$ on M such that the continuity equation* (3) *holds and*

$$\|v_t\|_{L^2(\mu_t)} \le |\dot\mu_t| \textit{ for a.e. } t \in (0,1).$$

Conversely, if a curve $\mu_t : [0,1] \to W_2(M)$ is such that there exists a Borel family of vector fields $(v_t)_{t\in[0,1]}$ with $\|v_t\|_{L^2(\mu_t)} \in L^1(0,1)$, together with which it satisfies (3)*, then there exists an a.c. curve $\tilde\mu_t$ being equal to μ_t for a.e. t and satisfying*

$$|\dot{\tilde\mu}_t| \le \|v_t\|_{L^2(\tilde\mu_t)} \textit{ for a.e. } t \in (0,1).$$

□

2.3 The tangent space $T_\mu W(M)$

As seen in Theorem 9, every absolutely continuous curve in $W(M)$ admits an $L^1(dt)$-family of $L^2(\mu_t)$-vector fields v_t, i.e. $\|v_t\|_{L^2(\mu_t)} \in L^1(0,1)$, together with which the continuity equation is satisfied. In the following, we will call every such pair (μ_t, v_t) an *a.c. couple*. We further want to call v_t an *accompanying vector field* for μ_t.

Vector fields v_t satisfying the continuity equation with a given μ_t are, however, not unique: there are many vector fields which allow for the same motion of the density: Adding another family w_t with the (t-independent) property $\nabla(w_t \mu_t) = 0$ to v_t does not alter the equation. Theorem 9 provides a natural criterion to choose a unique element among the $v_t's$. According to this theorem, there is at least one $L^1(dt)$- family v_t such that $|\dot{\mu}_t| = \|v_t\|_{L^2(\mu_t)}$ for almost all t, i.e. that is of minimal norm for almost all t. Linearity of (4) with respect to v_t and the strict convexity of the L^2-norms ensure the uniqueness of this choice, up to sets of zero measure with respect to t. We want to call such a couple (μ_t, v_t), where v_t is the unique minimal accompanying vector field for an a.c. curve μ_t, a *tangent couple*.

It then seems reasonable to define the tangent space at point μ as the set of $v \in L^2(TM, \mu)$ with $\|v\|_\mu \leq \|v + w\|_\mu$ for all $w \in L^2(TM, \mu)$ such that $\nabla(w\mu) = 0$. This condition for $v \in L^2(TM, \mu)$, however, is equivalent to saying that $\int_M h(v, w)\, d\mu = 0$ for all $w \in L^2(TM, \mu)$ with $\nabla(w\mu) = 0$. This in turn is equivalent to the following, which we will take as the definition of the tangent space.

Definition 10 (***Tangent space*** $T_\mu W(M)$)**.** The *tangent space* $T_\mu W(M)$ at point $\mu \in W(M)$ is defined as

$$T_\mu W(M) := \overline{\{\nabla\varphi \mid \varphi \in \mathcal{C}_c^\infty(M)\}}^{L^2(TM,\mu)} \subset L^2(TM, \mu). \tag{6}$$

We also give the definition of the normal space:

$$\begin{aligned} T_\mu^\perp W(M) &:= \{w \in L^2(TM, \mu) \mid \int h(w, v)\, d\mu = 0,\ \forall v \in T_\mu W(M)\} \\ &= \{w \in L^2(TM, \mu) \mid \nabla(w\mu) = 0\}. \end{aligned}$$

Remark 11. If (μ_t, v_t) is an a.c. couple, then (μ_t, v_t) is a tangent couple if and only if $v_t \in T_{\mu_t} W(M)$ for almost every $t \in (0, 1)$ ([Gig12], Proposition 1.30).

It is not difficult to see that $\dim T_\delta W(M) = \dim M$, for a Dirac measure δ, whereas in most of the cases $\dim T_\mu W(M) = \infty$. In general, it can be shown that as long as μ is supported on an at most countable set, $T_\mu W(M) = L^2(TM, \mu)$ (see [Gig12], Remark 1.33). Morally, the more points are contained in the support of the measure, the bigger gets the dimension. On the other hand, every probability measure can be approximated by a sequence of measures with finite support (see [Vil08] Thm 6.18), so that in each neighborhood of every measure there is an element μ with $\dim T_\mu W(M) < \infty$.

We call the disjoint union of all tangent spaces,

$$TW(M) := \bigsqcup_{\mu \in W(M)} T_\mu W(M) = \bigcup_{\mu \in W(M)} \{(\mu, v) \mid v \in T_\mu W(M)\},$$

the *tangent bundle of* $W(M)$. Since we are not treating $W(M)$ as a traditional manifold with charts, $TW(M)$ cannot be equipped with a traditional tangent bundle topology. Also, due to the denseness of the probability measures with finite support, local triviality cannot be achieved. However, since there is a natural projection map $\pi : TW(M) \to W(M);\ (\mu, v) \mapsto \mu$, we can in principle still talk about *sections* and *bundle maps* on the pointwise level. Whereas the notion of a vector field - in this context it would effectively be a field of (equivalence classes of) vector fields - has not turned out to be useful so far, we will use the concept of a bundle map later. In this spirit, a *bundle map* between tangent bundles of Wasserstein spaces $W(M)$ and $W(N)$ is a fiber preserving map $B : TW(M) \to TW(N)$ in the sense that together with a continuous map $F : W(M) \to W(N)$ the commutativity of the following diagram is satisfied:

$$\begin{tikzcd} TW(M) \arrow[r, "B"] \arrow[d, "\pi^M"'] & TW(N) \arrow[d, "\pi^N"] \\ W(M) \arrow[r, "F"] & W(N) \end{tikzcd}$$

One could ask about the meaningfulness of the condition that F should be continuous since for B the concept of contiuity does not makes sense. It is just that we require the preservation of as much structure as possible. In any case, we are mainly going to use this idea of a bundle map to make clear how we want to see our notion of a differential of a differentiable function $F : W(M) \to W(N)$.

On $W(M)$ one can furthermore define a (formal) Riemannian structure. Intuition comes from the following formula which is due to J.-D. Benamou and Y. Brenier ([BB99]). It shows that the Wasserstein distance W_2, having been defined through the, static, optimal transport problem, can be recovered by a dynamic formula, being reminiscent of the length functional on Riemannian manifolds, defining the Riemannian metric distance.

Theorem 12 (***Benamou-Brenier formula***)**.** *Let* $\mu,\ \nu \in \mathcal{P}_2(M)$*, then*

$$W(\mu,\nu) = \inf_{(\mu_t,v_t)} \int_0^1 \|v_t\|_{L^2(\mu_t)}\, dt, \tag{7}$$

where the infimum is taken among all a.c. couples (μ_t, v_t) *such that* $\mu_0 = \mu$ *and* $\mu_1 = \nu$*.*

This resemblance of formulas thus inspires the following definition.

Definition 13 (***Formal Riemannian tensor on*** $W_2(M)$)**.** The *formal Riemannian metric tensor* H_μ on $W(M)$ at point $\mu \in W(M)$ is defined as

$$\begin{aligned} H_\mu : T_\mu W(M) \times T_\mu W(M) &\to \mathbb{R} \\ (v,w) &\mapsto \int_M h_x(v,w)\, d\mu(x). \end{aligned}$$

Indeed, since $\|v_t\|_{L^2(\mu_t)} = \sqrt{\int_M h(v,v)\, d\mu} = \sqrt{H_\mu(v,v)}$, we now have $W(\mu,\nu) = \inf_{(\mu_t,v_t)} \int_0^1 \sqrt{H_\mu(v_t,v_t)}\, dt$. The tuple $(T_\mu W(M)), H_\mu)$ constitutes a Hilbert Space.

Gigli [Gig08] emphasizes that Definition 10 does not allow for a traditional Riemannian structure on $W_2(M)$ since the natural exponential map $v \mapsto \exp_\mu(v) := (Id + v)_\# \mu$ has injectivity radius 0 for every μ.

3 Differentiable maps between Wasserstein spaces

Since $W_2(M)$ and $W_2(N)$ are not manifolds in a traditional sense, to be able to talk about differentiability of maps $F : W_2(M) \to W_2(N)$ we cannot compose F with charts and apply Euclidean calculus. Recall, therefore, that a map $f : M \to N$ is differentiable if and only if it maps differentiable curves to differentiable curves.

3.1 Absolutely continuous maps

Having only a notion of absolutely continuous curves, which are metrically differentiable almost everywhere and which are at the foundation of the construction of tangent spaces at Wasserstein spaces, we start with the following definition.

Definition 14 (***Absolutely continuous map***)**.** A map $F : W(M) \to W(N)$ is called *absolutely continuous*, or, *a.c.*, if the curve $F(\mu_t) \subset W(N)$ is absolutely continuous up to redefining $t \mapsto \mu_t$ on a zero set, whenever $\mu_t \subset W(M)$ is absolutely continuous.

We want to build our notion of differentiable maps between Wasserstein spaces on this idea of absolutely continuous maps. Before we continue to do so, we first find some conditions under which maps are absolutely continuous. For this, we want to recall the notion of *proper* maps.

Definition 15 (***Proper map***)**.** A continuous map $f : X \to Y$ between a Hausdorff space X and a locally compact Hausdorff space Y is called *proper*, if for all compact subsets $K \subset Y$, the preimage $f^{-1}(K) \subset X$ is compact in X.

In the following we denote the operator norm of a linear map by $\|\cdot\|$.

Theorem 16. *Let* $F : W(M) \to W(N)$ *be given as* $F(\mu) = f_\# \mu$*,* $f : M \to N$ *being smooth and proper and such that* $\sup_{x\in M} \|df_x\| < \infty$*. Then* F *is absolutely continuous and for every tangent couple* (μ_t, v_t)*, the tuple* $(F(\mu_t), dF_{\mu_t}(v_t))$ *is an a.c. couple, where*

$$dF_{\mu_t}(v_t)_y := \int_{f^{-1}(y)} df_x(v_{t,x})\, d\mu_t^y(x) \tag{8}$$

for almost every t *and for* $y \in f(M)$*. Here,* $df_x : T_x M \to T_{f(x)} N$ *denotes the differential of* f *at the point* x*,* $v_{t,x}$ *means the vector field* v_t *at the point* $x \in M$ *and the probability measures* $\mu_t^y(x)$ *are defined through the disintegration theorem,* $d\mu_t(x) = d\mu_t^y(x) df_\# \mu_t(y)$ *(see Appendix A).*[1] *For all* $y \notin f(M)$*, we set* $dF_{\mu_t}(v_t)_y = 0$*.*

[1]Note that what in A appears as lower index y, now appears as upper index y since here were are additionally dealing with the t-dependence of μ_t.

Although $df_x : T_xM \to T_{f(x)}M$ is well defined for every x as a mapping between tangent spaces, it is not well defined as a mapping between vector fields as long as f is not injective. We thus take the mean value over all the vectors $df_x(v_{t,x})$ as the image vector $dF_{\mu_t}(v_t)_y$ of the vector field v_t at point y, where x stands for the elements of the fiber $f^{-1}(y)$. In case f is injective, $dF_\mu(v)$ reduces to $df(v)$ for every μ, which then can be regarded as full-fledged vector field.

Our naming of the vector field along $F(\mu_t)$, $dF_{\mu_t}(v_t)$ is, of course, very suggestive. Indeed, since the map $(v, \mu) \mapsto dF_\mu(v)$ is linear in v, Theorem 16 supports a natural definition for a notion of differentiability for absolutely continuous maps F. However, before we give such a definition, we need to make some further preparatory observations. Let us first continue with proving Theorem 16.

Proof. Let μ_t be an a.c. curve. Using Theorem 9, we want to prove that there exists a family of vector fields $(\tilde{v}_t)_{t\in[0,1]}$ with $\int_0^1 \|\tilde{v}_t\|_{L^2(F(\mu_t))}\, dt < \infty$, such that $(F(\mu_t), \tilde{v}_t)$ is an a.c. couple.

Let $(v_t)_{t\in[0,1]}$ be the tangent vector field of μ_t. For each t for which $v_t \in T_{\mu_t}W(M)$ (i.e. almost everywhere) we define $dF_{\mu_t}(v_t)$ as in equation (8). We will prove that $dF_{\mu_t}(v_t)$ is an example of such vector fields $\tilde{v}_t$ we are looking for.

Let us first see that $\int_0^1 \|dF_{\mu_t}(v_t)\|_{L^2(F(\mu_t))}\, dt < \infty$. Using the triangle inequality for Bochner integrals, Jensen's inequality, the disintegration theorem and Hölder's inequality (in this order), we have:

$$\int_0^1 \|dF_{\mu_t}(v_t)\|_{L^2(F(\mu_t))}\, dt = \int_0^1 \sqrt{\int_N \|dF_{\mu_t}(v_t)\|^2_{T_yN}\, dF(\mu_t)(y)}\, dt$$

$$\begin{aligned}
&= \int_0^1 \sqrt{\int_N \| \int_{f^{-1}(y)} df_x(v_{t,x})\, d\mu_t^y(x)\|^2_{T_yN}\, df_\#\mu_t(y)}\, dt \\
&\leq \int_0^1 \sqrt{\int_N \left(\int_{f^{-1}(y)} \|df_x(v_x)\|_{T_yN}\, d\mu_t^y(x)\right)^2 df_\#\mu_t(y)}\, dt \\
&\leq \int_0^1 \sqrt{\int_N \int_{f^{-1}(y)} \|df_x(v_{t,x})\|^2_{T_yN}\, d\mu_t^y(x)\, df_\#\mu_t(y)}\, dt \\
&= \int_0^1 \sqrt{\int_M \|df_x(v_{t,x})\|^2_{T_{f(x)}M}\, d\mu_t(x)}\, dt \\
&\leq \int_0^1 \sqrt{\int_M \|df_x\|^2 \cdot \|v_{t,x}\|^2_{T_xM}\, d\mu_t(x)}\, dt \\
&\leq \int_0^1 \sqrt{\int_M \|v_{t,x}\|^2_{T_xM}\, d\mu_t \;\cdot\; \operatorname{ess\,sup}^{\mu_t}_{x\in M}\|df_x\|^2}\, dt \\
&= \int_0^1 \sqrt{\|v_t\|^2_{L^2(\mu_t)} \cdot\; \operatorname{ess\,sup}^{\mu_t}_{x\in M}\|df_x\|^2}\, dt \\
&\leq C\int_0^1 \|v_t\|_{L^2(\mu_t)}\, dt \; < \infty.
\end{aligned}$$

With $\operatorname{ess\,sup}^{\mu_t}_{x\in M}$ we mean the essential supremum with respect to the measure μ_t and $C := \operatorname{ess\,sup}^{\mu_t}_{x\in M}\|df_x\|^2$. The last expression is finite, since we know that $\|v_t\|_{L^2(\mu_t)} \leq |\dot{\mu}_t|$ for almost every t and that the metric derivative of an a.c. map is integrable. (The calculation above shows in particular that $dF_{\mu_t}(v_t) \in L^2(\mu_t)$ for almost every t, as we will point out again below.) The disintegration theorem now allows the following calculation, with g being the Riemannian tensor on N and h the one on M, $\varphi \in \mathcal{C}_c^\infty(N \times (0,1))$ and ∇ the gradient with respect to the first coordinate:

$$
\begin{aligned}
& \int_N g_y\left(\nabla\varphi(y,t), dF_{\mu_t}(v_t)_y\right)\, df_\#\mu_t(y) \\
= & \int_N g_y\left(\nabla\varphi(y,t), \int_{f^{-1}(y)} df_x(v_{t,x}) d\mu_t^y(x)\right)\, df_\#\mu_t(y) \\
= & \int_N \int_{f^{-1}(y)} g_y\left(\nabla\varphi(y,t), df_x(v_{t,x})\right)\, d\mu_t^y(x) df_\#\mu_t(y) \\
= & \int_N \int_{f^{-1}(y)} g_{f(x)}\left(\nabla\varphi(f(x),t), df_x(v_{t,x})\right)\, d\mu_t^y(x) df_\#\mu_t(y) \\
= & \int_M g_{f(x)}\left(\nabla\varphi(f(x),t), df_x(v_{t,x})\right) d\mu_t(x) \\
= & \int_M h_x\left(\nabla(\varphi\circ f)(x,t), v_{t,x}\right) d\mu_t(x).
\end{aligned}
$$

By $(\varphi \circ f)(x,t)$ we mean $(\varphi \circ (f \times id))(x,t)$. For the second equality we used the continuity of the Riemannian tensor at every point $y \in N$. The last step is true because for every vector $X \in T_xM$,

$$
\begin{aligned}
h_x(\nabla(\varphi\circ f)(x), X) &= X(\varphi\circ f)(x) = df(X)(\varphi)(f(x)) \\
&= g_{f(x)}\left(\nabla\varphi(f(x)), d_x f(X)\right).
\end{aligned}
$$

With this, we can now prove our claim that $\frac{d}{dt}F(\mu_t) + \nabla(dF_{\mu_t}(v_t)F(\mu_t)) = 0$ in the weak sense: For every $\varphi \in \mathcal{C}_c^\infty(N \times (0,1))$ it is

$$
\begin{aligned}
& \int_0^1 \int_N \left(\frac{\partial}{\partial t}\varphi\right)(y,t) + g_y\left(\nabla\varphi(y,t), dF_{\mu_t}(v_t)_y\right)\, df_\#\mu_t(y)dt \\
= & \int_0^1 \int_M \left(\frac{\partial}{\partial t}\varphi\right)(f(x),t) + h_x\left(\nabla(\varphi\circ f)(x,t), v_{t,x}\right) d\mu_t(x)dt \\
= & \int_0^1 \int_M \left(\frac{\partial}{\partial t}(\varphi\circ f)\right)(x,t) + h_x\left(\nabla(\varphi\circ f)(x,t), v_{t,x}\right) d\mu_t(x)dt. \\
= & \ 0.
\end{aligned}
$$

Since f is smooth and proper, $\varphi \circ f \in \mathcal{C}_c^\infty(M \times (0,1))$ and we can apply our assumption on (μ_t, v_t) to be an a.c. couple. ☐

3.2 About the image of dF_μ

For Theorem 16 we did not need to test whether $dF_\mu(v) \in T_{F(\mu)}W(N)$ for all $v \in T_\mu W(M)$, since we only needed $(F(\mu_t), dF_{\mu_t}(v_t))$ to be an a.c. couple. But is it still true, given that (μ_t, v_t) is a tangent couple?

To begin with, the proof of Theorem 16 also guarantees that for every $\mu \in W(M)$ and $v \in T_\mu W(M)$, $dF_\mu(v) \in L^2(F(\mu))$. Knowing this, we can consider formula (8) as the prescription for a map between $T_\mu W(M)$ and $L^2(F(\mu))$.

It is also useful to know that this map is always bounded, which we will see in the next proposition. For the rest of this section, let $F : W(M) \to W(N)$ be as in Theorem 16 and $dF_\mu(v)$ as in formula (8).

Proposition 17 (***Boundedness of*** dF)**.** *For each $\mu \in W(M)$, $dF_\mu : T_\mu W(M) \to L^2(F(\mu))$ is bounded with*

$$\|dF_\mu\| \leq \operatorname{ess\,sup}^\mu_{x\in M} \|df_x\|. \tag{9}$$

Here, $\|\cdot\|$ denotes the operator norm of the respective linear map and $\operatorname{ess\,sup}^\mu_{x\in M}$ the essential supremum with respect to μ.

Inequality (9) can be attained by taking similar steps as in the proof of Theorem 16. The right-hand side of equation (9) is finite since we demanded $\sup_{x\in M}\|dg_x\|$ to be finite.

Let us give an example for a function F for which equality is attained for every μ in inequality (9).

Example 18. Let $g : M \to M$ be a Riemannian isometry, i.e. $g^*h = h$, where h is the Riemannian metric tensor on M. Then, for $F = g_\#$ and for all $\mu \in W(M)$, $\|dF_\mu\| = \operatorname{ess\,sup}^\mu_{x\in M}\|dg_x\| = 1$. This is, because on the one hand, for all $x \in M$, $\|dg_x\| = 1$, since dg is an isometry between the tangent spaces T_xM and $T_{g(x)}M$. On the other hand,

$$\|dg_\#\| = \sup_{\|v\|_{T_\mu W(M)}=1} \|dg(v)\|_{T_{g_\#\mu}W(M)} = \sup_{\|v\|_{T_\mu W(M)}=1} \|v\|_{T_\mu W(M)} = 1.$$

To come back to our question, whether $dF_\mu(v)$ is always an element of $T_{F(\mu)}W(M)$, we first want to study the following simple cases.

Lemma 19. *Let $\mu = \delta_x$, for $x \in M$. Then $dF_\mu(v) \in T_{F(\mu)}W(N)$ for all $v \in T_\mu W(M)$.*

Proof. This is true because $F(\delta_x) = \delta_{f(x)}$ and for every $y \in N$, $L^2(\delta_y) \cong \mathbb{R}^n \cong T_{\delta_y}W(N)$, $n = dim\ N$. □

Lemma 20. *Let $g : M \to M$ be a Riemannian isometry, i.e. $g^*h = h$, and $v = \nabla\varphi \in T_\mu W(M)$, $\varphi \in \mathcal{C}_c^\infty(M)$. Then for every $\mu \in W(M)$, $dF_\mu(v) = dg(v) = \nabla(\varphi \circ g^{-1}) \in T_{F(\mu)}W(M)$.*

Proof. For the Riemannian metric h on M and for every vector field X

$$\begin{aligned} h(\nabla(\varphi\circ g^{-1}), X) &= d(\varphi\circ g^{-1})(X) = d\varphi(dg^{-1}(X)) = h(\nabla\varphi, dg^{-1}(X)) \\ &= h(dg(\nabla\varphi), X). \end{aligned}$$

□

Since we know from Proposition 17 that $dg_{\#\mu}$ is bounded and therefore continuous for every $\mu \in W(M)$, we can infer the following more general statement.

Corollary 21. *Let $g : M \to M$ be a Riemannian isometry and $T_\mu W(M) \ni v = \lim_{n\to\infty} \nabla\varphi_n$. Then $dg(v) = \lim_{n\to\infty} \nabla(\varphi_n \circ g^{-1}) \in T_{F(\mu)}W(M)$.*

However, the case in Lemma 19 is extreme and the choice of functions in Lemma 20 specific. We will now see that it can well be that dF_μ does not always hit the tangent space at $F(\mu)$.

Theorem 22. *Let M be a compact manifold without boundary and $f = \mathrm{id}_M : (M, h_1) \to (M, h_2)$ the identity map on M, where $h_2 = \nu^2 h_1$ and $\nu : M \to (0,\infty)$ nonconstant. Then for $F = \mathrm{id}_\# : W(M, h_1) \to W(M, h_2)$ there exists a $\nabla\varphi \in T_\mu W(M, h_1)$ so that $dF_\mu(\nabla\varphi) \notin T_\mu W(M, h_2)$, where $\mu = C \cdot \mu_{h_1}$, μ_{h_1} the volume measure on M with respect to h_1 and $C = 1/\mu_{h_1}(M)$.*

Proof. It is clear that $F = \mathrm{id}_{W(M)}$ and $dF_\mu(v) = v\ \forall v \in T_\mu W(M, h_1)$. However, v is not automatically a member of $T_\mu W(M, h_2)$. We will show that if φ is chosen appropriately, $v = \nabla^{h_1}\varphi$ is not a limit of gradients with respect to h_2.

For this, recall that on a general Riemannian manifold (M, h), there is a duality between vector fields v and 1-forms $v^\flat$ by the formula $v^\flat_h(\cdot) := h(v, \cdot)$, which maps the vector field $\nabla^h\varphi$ to the 1-form $d\varphi$. This identification gives an isomorphism between $\overline{\{\nabla^h\varphi\}}^{L^2(TM,h,\mu)}$ and $\overline{\{d\phi\}}^{L^2(T^*M,h^*,\mu)}$. Since this isomorphism depends on the chosen metric, it is in general $v^\flat_{h_1} \neq v^\flat_{h_2}$, but rather $v^\flat_{h_2} = \nu^2 v^\flat_{h_2}$, as Lemma 24 below shows. And thus $\nabla^{h_1}\varphi^\flat_{h_2} = \nu^2 d\varphi$.

Now $d(\nu^2 d\varphi) = d(\nu^2) \wedge d\varphi$ which one can easily arrange to be non-zero. From Lemma 23 below we can thus infer that $\nabla^{h_1}\varphi^\flat_{h_2} \notin \overline{\{d\varphi\}}^{L^2(T^*M,h_2^*,\mu_{h_2})}$. As $C\mu_{h_1} = C\nu^n\mu_{h_2}$, with $n = \dim(M)$, the topology on $L^2(T^*M, h_2, C\mu_{h_1})$ and $L^2(T^*M, h_2, \mu_{h_2})$ coincide, so one can conclude that $\nu^2 d\varphi$ is not an element of $T_\mu W(M, h_2)$. □

Lemma 23. *If ω is a smooth 1-form on M with $d\omega \neq 0$ then $\omega \notin \overline{\{d\varphi\}}^{L^2(T^*M,g^*,\mu_h)}$, where μ_h is the volume measure on M with respect to h.*

Proof. Assuming the opposite and using the standard inner products, one gets the following contradiction:

$$0 \neq (d\omega, d\omega) = (\omega, d^*d\omega) = \lim(d\varphi_n, d^*d\omega) = \lim(dd\varphi_n, d\omega) = \lim 0 = 0 \tag{10}$$

□

Lemma 24. *In the situation of Theorem 22 and interpreting dF as a map of L^2-one forms, we have $dF_\mu(\omega) = \nu^2\omega$.*

Proof. Every vector field $v \in TM$ corresponds to the covector field $\omega \in T^*M$ by $\omega(w) = h_1(v, w)$. A change of the Riemannian metric h_1 to $h_2 = \nu^2 h_1$ yields $h_2(v, w) = \nu^2 h_1(v, w) = h_1(v, \nu^2 w) = \omega(\nu^2 w) = \nu^2 \omega(w)$, so with respect to h_2, v corresponds to $\nu^2 \omega$. □

3.3 Differentiable maps between Wasserstein spaces

As we have seen in Subsection 3.2, the conditions of Theorem 16 do not guarantee $dF_\mu(v) \in T_{F(\mu)}W(N)$, even though this property is neccessary for a meaningful definition of the differential of F. To help us here, we use the fact that $L^2(\nu) = T_\nu W(N) \oplus T_\nu^\perp W(N)$ for every $\nu \in W(N)$ and compose dF with a projection onto $T_{F(\mu)}W(N)$, so that at least $P^{F(\mu)} \circ dF_\mu : T_\mu W(M) \to T_{F(M)}W(M)$ is a linear and bounded map between $T_\mu W(M)$ and $T_{F(M)}W(M)$.

Definition 25. We call P^μ the orthogonal linear projection

$$\begin{array}{rcl} P^\mu : L^2(\mu) & \longrightarrow & T_\mu W(M) \\ v & \longmapsto & v^\top, \end{array}$$

where $v = v^\top + v^\perp$, with $v^\top \in T_\mu W(M)$ and $v^\perp \in T_\mu^\perp W(M)$.

Proposition 26. *For every a.c. couple (μ_t, v_t), $(\mu_t, P^{\mu_t}(v_t))$ is a tangent couple.*

Proof. Let (μ_t, v_t) be an a.c. couple, then, for $v_t = v_t^\top + v_t^\perp$ we have

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t^\top \mu_t) = \frac{d}{dt}\mu_t + \nabla \cdot ((v_t^\top + v_t^\perp)\mu_t) = 0.$$

And since $\|P^{\mu_t}(v_t)\|_{L^2(\mu_t)} \leq \|v_t\|_{L^2(\mu_t)}$ we have also $\|P^{\mu_t}(v_t)\|_{L^2(\mu_t)} \in L^1(0,1)$. Thus, $(\mu_t, P^{\mu_t}(v_t))$ is an a.c. couple and with Remark 11 a tangent couple. □

With the observations we have collected so far, we can finally give our definition of a differentiable map between Wasserstein spaces.

Definition 27 (*Differentiable map between Wasserstein spaces*). An absolutely continuous map $F : W(M) \to W(N)$ is called *differentiable* in case for every $\mu \in W(M)$ there exists a bounded linear map $dF_\mu : T_\mu W(M) \to T_{F(\mu)}W(N)$ such that for every tangent couple (μ_t, v_t) the image curve $dF_{\mu_t}(v_t)$ is a tangent vector field of $F(\mu_t)$. In this way a bundle map[2] $dF : TW(M) \to TW(N)$ is defined which we want to call the *differential* of F.

When we say a map $F : W(M) \to W(N)$ is differentiable we automatically mean that it is absolutely continuous in the first place.

Remark 28. The reader might be surprised that we only give a global definition of differentiability, without having started with a pointwise definition. The latter is difficult, if at all possible, since the tangent vector fields v_t are only defined for a.e. $t \in [0,1]$, so a pointwise evaluation of these is not well-defined. The situation would change if one would be able to speak about *continuous* curves of tangent vector fields, but it doesn't seem to be so easy to make this notion precise: For differing t, t' the vector fields v_t and $v_{t'}$ are elements of different tangent spaces, potentially even of different dimension, which is why the usual notion of continuity cannot be trivially applied.

Note again that $dF_{\mu_t}(v_t)$ is only well-defined almost everywhere, since v_t is. But this is not harmful to our definition since in particular also the tangent vectors of $F(\mu_t)$ are only well-defined almost everywhere. But in this same manner, Definition 27 does not guarantee uniqueness of dF in a strict sense. (Here we mean that $dF = \widetilde{dF}$ whenever $dF_\mu(v) = \widetilde{dF}_\mu(v)$ for all $(\mu, v) \in TW(M)$.) But, after all, one can say that dF is unique up to a "negligible" set.

Definition 29 (*Negligible set*). A subset $Z \subset TW(M)$ is called *negligible* whenever for every tangent couple (μ_t, v_t) the set $\{t \in (0,1) \mid (\mu_t, v_t) \in Z\}$ is of Lebesgue measure zero.

This definition respects the $L^1(dt)$-nature of the v_t's in the sense that changing any v_t on a set of measure zero does not change the measure of the set $\{t \in (0,1) \mid (\mu_t, v_t) \in Z\}$.

Proposition 30 (*Uniqueness of the differential*). *The differential dF of a differentiable map $F : W(M) \to W(N)$ is unique up to a redefinition on a negligible set $Z \subset TW(M)$.*

[2] In our sense of the word "bundle".

Proof. Let dF and $\widetilde{dF}$ be two pointwise linear bundle maps, dF being the differential of an a.c. map F. It is to show that dF and $\widetilde{dF}$ are both a differential of F if and only if $\{(\mu, v) \in TW(M) \mid dF_\mu(v) \neq \widetilde{dF}_\mu(v)\}$ is negligible.
Let dF and $\widetilde{dF}$ be different only on a negligible set. In this case, for each tangent couple (μ_t, v_t) the image velocities $\widetilde{dF}_{\mu_t}(v_t)$ are different from the ones of $dF_{\mu_t}(v_t)$ only on a null set and thus still equal the tangent vector fields along $F(\mu_t)$ almost everywhere. Let on the other hand dF and $\widetilde{dF}$ both fulfill the conditions of Definition 27. By definition, for each tangent couple (μ_t, v_t) both $dF_{\mu_t}(v_t)$ and $\widetilde{dF}_{\mu_t}(v_t)$ are equal almost everywhere to the tangent vectors along $F(\mu_t)$. Thus, for every tangent couple (μ_t, v_t), $\{t \in (0,1) \mid dF_{\mu_t}(v_t) \neq \widetilde{dF}_{\mu_t}(v_t)\}$ has Lebesgue measure zero. □

Let us now analyse some properties of negligible sets.

Proposition 31. *1.) $T_\mu(W(M)) \setminus \{0\}$ is negligible, for every $\mu \in W(M)$. But $T_\mu(W(M))$ isn't.*

2.) The countable union of negligible sets is negligible.

3.) Every subset of a negligible set is negligible.

4.) The following is an equivalence relation on the set of mappings between tangent bundles on Wasserstein spaces:
$F \sim G :\Leftrightarrow \{(\mu, v) \in TW(M) \mid F(\mu, v) \neq G(\mu, v)\}$ *is negligible.*

Remark 32. Let dF be a differential of a map $F : W(M) \to W(N)$. Then there are members of its equivalence class $[dF]$ which are not a differential of F since not every member has to be pointwise linear and bounded. Restricting, however, the equivalence relation onto the subset of pointwise linear and bounded maps between tangent bundles of Wasserstein spaces solves this issue. In this case $[dF]$ contains precisely all the possible differentials of F. Whenever we refer to a representative of dF, we mean an element of the latter equivalence class.

Proof. 1.) Let (μ_t, v_t) be a tangent couple, v_t a fixed representative of $v_t \in L^1(dt)$ and $T_\mu := \{t \in (0,1) \mid \mu_t = \mu,\ v_t \in T_\mu W(M)\}$ for some $\mu \in W(M)$. Let us further assume that $v_t \neq 0$ for every $t \in T_\mu$ which in particular means that $|\dot{\mu}_t| \neq 0$ for every $t \in T_\mu$. From this we can also infer that for no $t_0 \in T_\mu$ there exists a neighborhood on which μ_t is constant. Let $a \in T_\mu$ be a point which is not isolated. This means that in every neighborhood of a is another point of T_μ. The consequence of this would be that the metric derivative would not exist at that point which we excluded in the definition of T_μ. So T_μ must consist of only isolated points and thus must be countable. Choosing another representative of $v_t \in L^1(\mu)$ only changes the amount of t's in T_μ by a null set.
$T_\mu(W(M))$ is not negligible since $\mu_t = \mu$ is absolutely continuous with metric derivative 0.

2.) This follows from the fact that any countable union of sets of measure zero again is of measure zero.

3.) Let N be a subset of a negligible set and (μ_t, v_t) an a.c. curve with a fixed representative v_t. The amount of times where $(\mu_t, v_t) \in N$ can only be a subset of a set of zero measure. Since the Lebesgue measure is a complete measure this subset itself is measurable and in particular of measure zero.

4.) This follows from 1.) and 2.)

□

The following corollary finally recovers the properties expected of a differential.

Corollary 33. *1.) In case $F = f_\#$ and f is as in Theorem 16, F is differentiable with $dF_\mu = P^{F(\mu)} \circ \widehat{dF}_\mu$, where $P^{F(\mu)}$ is the orthogonal projection onto $T_{F(\mu)}N$ from Proposition 26 and*

$$\widehat{dF}_\mu(v)_y := \int_{f^{-1}(y)} df(v_x) d\mu^y(x),$$

as in formula (8)*. In case f is a Riemannian isometry, the additional projection P is not necessary, as we have seen in Corollary 21. Then, $dF_\mu = df$ for all $\mu \in W(M)$.*

2.) In particular, the identity mapping $F(\mu) = \mu$ is differentiable with $dF_\mu(v) = v$ up to a negligible map.

3.) Let $F : W(M) \to W(N)$ and $G : W(N) \to W(O)$ be two differentiable maps. Then also $G \circ F : W(M) \to W(O)$ is differentiable with $d(G \circ F)_\mu(v) = \left(dG_{F(\mu)} \circ dF_\mu\right)(v)$ up to a negligible set.

4.) Whenever F is differentiable, bijective with differentiable inverse F^{-1}, then dF is also invertible with inverse $d(F^{-1})$, up to a negligible set.

Proof. 1.) This follows from Theorem 16 and Proposition 26.

2.) This is immediate.

3.) First we observe that the composition of two absolutely continuous maps between Wasserstein spaces is again absolutely continuous. Also, the composition of two bounded linear maps is again a bounded linear map. To show differentiability, we will check that $dG_{F(\mu)} \circ dF_\mu : T_\mu W(M) \to T_{(G\circ F)(\mu)}W(O)$ is such that for every tangent couple (μ_t, v_t), also $((G \circ F)(\mu_t), (dG_{F(\mu)} \circ dF_\mu)(v_t))$ is a tangent couple. So let (μ_t, v_t) be a tangent couple. Since F is differentiable, we know that $(F(\mu_t), dF_{\mu_t}(v_t))$ is a tangent couple. Similarly, also $\big(G(F(\mu_t)), dG_{F(\mu_t)}(dF_{\mu_t}(v_t))\big)$ is a tangent couple. Since $G(F(\mu_t)) = (G \circ F)(\mu_t)$ and $dG_{F(\mu_t)}(dF_{\mu_t}(v_t)) = (dG_{F(\mu_t)} \circ dF_{\mu_t})(v_t)$, we have proven the claim.

4.) This is an immediate consequence of 2.) and 3.).

□

Remark 34. Let us again emphasize that this type of differentiability is highly tailored to the structure given by optimal transport. It knowingly does not fit into the framework of, e.g., [KM97]. Nevertheless, let us mention that also in this reference, the notion of differentiable maps between infinite dimensional manifolds is established via the property that differentiable curves should be mapped to differentiable curves.

3.4 Pullbacks and formal Riemannian isometries

As an application of the previous section, we propose a definition for the pullback of the formal Riemannian tensor on $W_2(M)$ and furthermore a definition for formal Riemannian isometries. As the formal Riemannian metric was defined by comparison of formulae to actual Riemannian structures (see Definition 13), the performance of pullbacks now gives rise to definitions of further possible formal Riemannian metrics on W_2-spaces, in cases where dF_μ is injective for every μ, i.e. in case F can be considered to be an immersion.

Definition 35 (*Pullback of the formal Riemannian tensor*). Let $F : W(N) \to W(M)$ be differentiable, dF be a fixed differential of F, $\mu \in W(N)$ and $H_{F(\mu)}$ the formal Riemannian metric tensor on $W(M)$ at point $F(\mu) \in W(M)$. Then, for $v, w \in T_\mu W(M)$, the *pullback* $(F^*H)_\mu$ of $H_{F(\mu)}$ is defined as

$$(F^*H)_\mu(v, w) := H_{F(\mu)}(dF_\mu(v), dF_\mu(w)).$$

Unfortunately, this definition depends on the choice of the differential of F, which is, as we have seen, only unique up to a negligible set.

Definition 36 (*Formal Riemannian isometry*). Analogously to the finite dimensional case, we call a bijective differentiable map $F : W(M) \to W(M)$ with differentiable inverse a *formal Riemannian isometry*, in case there is a representative of dF such that for all $\mu \in W(M)$ $(F^*H)_\mu(v, w) = H_\mu(v, w)$ for all $(v, w) \in T_\mu W(M) \times T_\mu W(M)$.

It is straightforward to see that F is a formal Riemannian isometry iff there is a representative of dF such that for every $\mu \in W(M)$ $dF_\mu : T_\mu W(M) \to T_{F(\mu)}W(M)$ is a metric isometry with respect to the metrics induced by the L^2-norms.

Important formal Riemannian isometries are generated by the isometry group of the underlying metric space. By means of the pushforward, $ISO(M)$ acts isometrically also on $\mathcal{P}_p$ and the map

$$\begin{array}{rcl} G \times TW(M) & \to & TW(M) \\ (g, (\mu, v)) & \mapsto & (g_\#\mu, dg(v)) \end{array}$$

defines an induced action of every subgroup G of $ISO(M)$ on the tangent bundle of $W(M)$, where we regard dg as a differential of $g_\#$. It is quick to check that for $g \in ISO(M)$, $g_\# : W(M) \to W(M)$ is a formal Riemannian isometry.

Lemma 37. *Let $g \in ISO(M)$, then $T_{g_\#\mu}W(M) = dg\,(T_\mu W(M))$ for all $\mu \in W(M)$. Here, we again regard dg as a, fixed, differential of $g_\#$.*

Proposition 38. *Every formal Riemannian isometry is an isometry in the metric sense of its Wasserstein space.*

Proof. Let F be a formal Riemannian isometry. Since by definition F is bijective with differentiable inverse, every a.c. couple (μ_t, v_t) can be represented as the image of another a.c. couple $(\tilde{\mu}_t, \tilde{v}_t)$. Just choose $\tilde{\mu}_t := F^{-1}(\mu_t)$ and $\tilde{v}_t := dF^{-1}(v_t)$. Then, $\mu_t = F(\tilde{\mu}_t)$ and, using Corollary 33, $v_t = dF(\tilde{v}_t)$ almost everywhere. Conversely, every image of an a.c. couple, in the above sense, is an a.c. couple. Let dF be a suitable representative. For $\mu, \nu \in W(M)$ and μ_t a.c. connecting them, we then have according to 12:

$$\begin{aligned} W(F(\mu), F(\nu)) &= \inf_{(F(\mu_t), dF(v_t))} \int_0^1 \sqrt{H_{F(\mu_t)}(dF(v_t), dF(v_t))}\, dt \\ &= \inf_{(\mu_t, v_t)} \int_0^1 \sqrt{H_{\mu_t}(v_t, v_t)}\, dt \; = \; W(\mu, \nu). \end{aligned}$$

□

It would be interesting to find out whether the converse implication of Proposition 38 is true as well, as it is the case for finite dimensional Riemannian manifolds.

3.5 Convex mixing of maps

In the examples, we so far have only been concerned with maps $F : W(M) \to W(N)$ which are induced by maps $f : M \to N$. Now one could wonder how a map F which is not of this type could look like and what its differentiability properties are. As a first hint, we recall that whenever there is an $f : M \to N$ such that $F = f_\#$, then for $x \in M$ it is $F(\delta_x) = \delta_{f(x)}$. Based on this, we can construct the following examples.

Example 39.

- If $F(\mu) = \mu_0$ is a constant map such that $\mu_0 \neq \delta_{y_0},\ y_0 \in N$, then there exists no map $f : M \to N$ such that $F = f_\#$. In case $F(\mu) = \delta_{y_0}$, it is $F = f_\#$ with $f(x) = y_0\ \forall x \in M$.
- Let $F_i : W(M) \to W(N)$, $i = 1, 2$, such that they do not coincide on $\{\delta_x \mid x \in M\}$. The mixing of measures $F := (1 - \lambda)F_1 + \lambda F_2$ for $0 < \lambda < 1$, then, cannot be a pushforward of measures.

Remark 40. Another way to think about this issue is the following: Every map $F : W(M) \to W(N)$ has a decomposition into a map $\tilde{F} : W(M) \to \mathcal{P}(M \times N)$ with $\pi^1_\# \tilde{F}(\mu) = \mu$ and the map $\pi^2_\# : \mathcal{P}(M \times N) \to W(N)$, i.e. $F = \pi^2_\# \circ \tilde{F}$. Certainly, $\tilde{F}$ is not unique, but one can always choose $\tilde{F}(\mu) = \mu \otimes F(\mu)$. Thus, F is a pushforward with respect to a map f if and only if there exists a map $\tilde{F}$ in such a way that $\tilde{F}(\mu) = (Id, f)_\# \mu$. According to [AG13], Lemma 1.20 this is equivalent to saying that for every μ there exists a $\tilde{F}(\mu)$-measurable set $\Gamma \subset M \times N$ on which $\tilde{F}(\mu)$ is concentrated such that for μ-a.e. x there exists only one $y = f(x) \in M$ with $(x, y) \in \Gamma$. And in this case, $\tilde{F}(\mu) = (Id, f)_\# \mu$.

It is easy to see that any constant map $F : W(M) \to W(N),\ \mu \mapsto \mu_0$, is differentiable with $dF = 0$ up to a negligible set. In the following we will investigate whether maps of the form $F = (1 - \lambda)\ F_1 + \lambda\ F_2$ are also differentiable. Let us start with asserting that the convex mixing of of a.c. maps is a.c..

Proposition 41. *Let $F_i : W(M) \to W(N)$, $i = 1, 2$, be arbitrary a.c. maps. Then, for $0 \leq \lambda \leq 1$, also $F := (1 - \lambda)\ F_1 + \lambda\ F_2$ is a.c.*

For the proof of Proposition 41 we will use that already the convex mixing of of a.c. curves is a.c.

Lemma 42. *Let μ^1_t and μ^2_t be a.c. curves. Then also the convex mixing $\mu_t := (1 - \lambda)\mu^1_t + \lambda \mu^2_t$ with $0 \leq \lambda \leq 1$ is an a.c. curve.*

Proof. Since the μ^i_t are a.c. curves, for every $s \leq t \in (0, 1)$ there is a $g_i \in L^1(0, 1)$ such that

$$W\left(\mu^i_s, \mu^i_t\right) \leq \int_s^t g_i(\tau)\, d\tau.$$

Now let $\gamma_i \in Adm(\mu^i_s, \mu^i_t)$. Then $(1 - \lambda)\gamma_1 + \lambda\gamma_2 \in Adm\,(\mu_s, \mu_t)$. This is because for every measurable set A and π^i the projection onto the i-th component,

$$\begin{aligned} \pi^1_\# \left((1 - \lambda)\gamma_1 + \lambda\gamma_2\right)(A) &= \left((1 - \lambda)\gamma_1 + \lambda\gamma_2\right)\left((\pi^1)^{-1}(A)\right) \\ &= (1 - \lambda)\gamma_1((\pi^1)^{-1}(A)) + \lambda\gamma_2((\pi^1)^{-1}(A)) \\ &= \left((1 - \lambda)\mu^1_s + \lambda\mu^2_s\right)(A) \; = \; \mu_s(A). \end{aligned}$$

Similarly for π^2. Then for $\widetilde{Adm}\,(\mu_s,\mu_t) := \{(1-\lambda)\gamma_1 + \lambda\gamma_2 \mid \gamma_i \in Adm(\mu_s^i,\mu_t^i)\} \subset Adm\,(\mu_s,\mu_t)$ we have

$$\begin{aligned}
W(\mu_s,\mu_t)^2 &= W\left((1-\lambda)\mu_s^1 + \lambda\mu_s^2, (1-\lambda)\mu_t^1 + \lambda\mu_t^2\right)^2 \\
&\leq \inf_{\pi\in\widetilde{Adm}(\mu_s,\mu_t)} \int d^2(x,y)\, d\pi(x,y) \\
&= (1-\lambda) \inf_{\gamma_1\in Adm(\mu_s^1,\mu_t^1)} \int d^2(x,y)\, d\gamma_1 + \lambda \inf_{\gamma_2\in Adm(\mu_s^2,\mu_t^2)} \int d^2(x,y)\, d\gamma_2 \\
&= (1-\lambda)\, W(\mu_s^1,\mu_t^1))^2 + \lambda\, W(\mu_s^2,\mu_t^2)^2
\end{aligned}$$

This means that

$$\begin{aligned}
W(\mu_s,\mu_t) &= \sqrt{(1-\lambda)\, W(\mu_s^1,\mu_t^1))^2 + \lambda\, W(\mu_s^2,\mu_t^2)^2} \\
&\leq \sqrt{(1-\lambda)}\, W(\mu_s^1,\mu_t^1) + \sqrt{\lambda}\, W(\mu_s^2,\mu_t^2) \\
&\leq \sqrt{(1-\lambda)} \int_s^t g_1(\tau)\, d\tau + \sqrt{\lambda} \int_s^t g_2(\tau)\, d\tau \\
&= \int_s^t (\sqrt{(1-\lambda)}\, g_1 + \sqrt{\lambda}\, g_2)\, d\tau.
\end{aligned}$$

□

Before continuing with the proof of Proposition 41 we give this immediate corollary from the proof of Lemma 42.

Corollary 43. *Let (X,d) be a metric space and $\mu_{11},\mu_{12},\mu_{21},\mu_{22}$ four probability measures on X. Then,*

$$W_p\left((1-\lambda)\mu_{11} + \lambda\mu_{12}, (1-\lambda)\mu_{21} + \lambda\mu_{22}\right) \leq \sqrt[p]{(1-\lambda)} W_p(\mu_{11},\mu_{21}) + \sqrt[p]{\lambda} W_p(\mu_{12},\mu_{22}).$$

Proof of Proposition 41. Let μ_t be an a.c. curve. Then by definition $F_i(\mu_t)$, $i=1,2$, are a.c. curves. From Lema 42 we now know that also $F(\mu_t)$ is an a.c. curve. □

Theorem 44. *Let $F_i : W(M) \to W(N)$, $i = 1,2$, be two differentiable maps. Then $F = (1-\lambda)\, F_1 + \lambda\, F_2$ is differentiable.*

Since we have already seen that with the conditions of Theorem 44 F is a.c., as both F_i are a.c., we know that F maps a.c. curves to a.c. curves. We know further that along each of these a.c. image curves there has to be a tangent vector field. To find the tangent map, mapping curves of tangent vector fields along a.c. curves to the corresponding curves of tangent vector fields along the image a.c. curves, i.e. to prove the theorem, we first give a formula for a canonical image tangent vector field.

Lemma 45. *Let $F_i : W(M) \to W(N)$, $i = 1,2$, be two differentiable maps. For an a.c. curve γ_t in $W(M)$, we define the a.c. curves $\mu_t := F_1(\gamma_t)$, $\nu_t := F_2(\gamma_t)$ and $\alpha_t := \lambda\mu_t + (1-\lambda)\nu_t$ in $W(N)$. With the Lebesgue decompositiom theorem, the measures μ_t and ν_t give rise to unique measures $\tau_t^\mu, \tau_t^\nu, \beta_t$ and Radon-Nykodym derivatives ρ_t such that*

1. *For each t the measures τ_t^μ, τ_t^ν and β_t are mutually singular: there exist Borel subsets A_t, B_t, C_t that are pairwise disjoint with union N such that B_t and C_t are nullsets for τ_t^μ, A_t and C_t are nullsets for τ_t^ν and A_t, B_t are nullsets for β_t.*
2. *$\mu_t = \tau_t^\mu + \beta_t$*
3. *$\nu_t = \tau_t^\nu + \rho_t\beta_t$*
4. *ρ_t is zero only on a nullset of C_t.*

If furthermore v_t is a tangent vector field for μ_t and w_t is an accompanying vector field for ν_t, we can give the formula for a canonical accompanying vector field $u_t \in L^2(N,\alpha_t)$ for α_t as

$$u_t(x) := \begin{cases} v_t(x); & x \in A_t \\ w_t(x); & x \in B_t \\ \frac{\lambda v_t(x) + \rho_t(1-\lambda) w_t(x)}{\lambda + (1-\lambda)\rho_t}; & x \in C_t. \end{cases}$$

Proof. Since $\frac{d}{dt}\alpha_t$ is linear in α_t, the continuity equation for (α_t, u_t) is satisfied if and only if

$$\int_0^T \int_N h(\nabla\phi(x,t), u_t(x))d\alpha_t\, dt = \int_0^T (\int_N h(\nabla\phi(x,t), v_t(x))\lambda d\mu_t + h(\nabla\phi(x,t), w_t(x))(1-\lambda)d\nu_t)\, dt \quad (11)$$

for all $\varphi \in C_c^\infty((0,T)\times N)$ and $u_t \in L^2(N, \alpha_t)$.

Let us first check that $u_t \in L^2(TN, \alpha_t)$. Since $N = A_t \dot\cup B_t \dot\cup C_t$, the condition can be checked separately on A_t, B_t and C_t. First,

$$\int_{A_t} |u_t(x)|^2 d\alpha_t = \int_{A_t} |v_t(x)|^2 \lambda d\mu_t < \infty,$$

and similarly for B_t. To check the situation on C_t, we start with

$$\begin{aligned}\int_{C_t} |u_t(x)|^2 d\alpha_t &= \int_{C_t} \frac{|\lambda v_t(x) + (1-\lambda)\rho_t w_t(x)|^2}{(\lambda + (1-\lambda)\rho_t)^2}(\lambda d\beta_t + (1-\lambda)\rho_t d\beta_t) \quad (12)\\ &\leq 2\int_{C_t}\left(\frac{\lambda}{\lambda+(1-\lambda)\rho_t}|v_t(x)|^2\lambda + \frac{(1-\lambda)\rho_t}{\lambda+(1-\lambda)\rho_t}|w_t(x)|^2(1-\lambda)\rho_t\right) d\beta_t. \quad (13)\end{aligned}$$

Now it holds that $\frac{\lambda}{\lambda+(1-\lambda)\rho_t} \leq 1$ and $\int_{C_t} |v_t(x)|^2 d\beta_t < \infty$ (as one summand in the L^2-norm of v_t with respect to μ_t). Similarly for the second summand, so we see that the whole expression in Equation (13) is finite.

Let us now check Equation (11). This can be done separately for (almost all) $t \in [0,T]$ and again separately for the integrals over A_t, B_t, C_t. On A_t, Equation (11) holds because here $u_t = v_t$ and $\alpha_t = \lambda\mu_t = \lambda\tau_t^\mu$, whereas $\nu_t(A_t) = 0$. A similar argument works on B_t. On C_t, formally,

$$u_t d\alpha_t = \frac{\lambda v_t + (1-\lambda)\rho_t w_t}{\lambda + (1-\lambda)\rho_t} d(\lambda\beta_t + (1-\lambda)\rho_t\beta_t) = (\lambda v_t + (1-\lambda)\rho_t w_t)\, d\beta_t = v_t\lambda d\mu_t + w_t(1-\lambda)d\nu_t.$$

□

Proof of Theorem 44. First, we need to check that u_t is indeed an accompanying vector field for α_t, i.e. that $\|u_t\|_{L^2(\alpha_t)} \in L^1(0,1)$, so that its projection onto the tangent spaces is indeed a tangent vector field along α_t.

Since $N = A_t \dot\cup B_t \dot\cup C_t$,

$$\begin{aligned}\|u_t\|_{L^2(\alpha_t)} &= \|\, u_t|_{A_t} + u_t|_{B_t} + u_t|_{C_t} \|_{L^2(\alpha_t)} \leq \|\, u_t|_{A_t} \|_{L^2(\alpha_t)} + \|\, u_t|_{B_t} \|_{L^2(\alpha_t)} + \|\, u_t|_{C_t} \|_{L^2(\alpha_t)}\\ &\leq \sqrt{\lambda}\, \|v_t\|_{L^2(\mu_t)} + \sqrt{(1-\lambda)}\, \|w_t\|_{L^2(\nu_t)} + \|\, u_t|_{C_t} \|_{L^2(\alpha_t)}. \quad (14)\end{aligned}$$

We know of the first two summands in Equation (14) that their $L^1(0,1)$-norm is finite, as we demanded v_t and w_t to be accompanying vector fields. It thus suffices to show the finiteness of the $L^1(0,1)$-norm of the last summand. Here, we find with $\bar\rho_{t,\lambda} := \frac{1}{\lambda+(1-\lambda)\rho_t}$,

$$\|\, u_t|_{C_t} \|_{L^2(\alpha_t)} = \|\, (\lambda v_t + (1-\lambda)\rho_t w_t)|_{C_t} \|_{L^2(\bar\rho_{t,\lambda} d\beta_t)} \leq \|\lambda\, v_t|_{C_t} \|_{L^2(\bar\rho_{t,\lambda} d\beta_t)} + \|(1-\lambda)\rho_t\, w_t|_{C_t} \|_{L^2(\bar\rho_{t,\lambda} d\beta_t)}.$$

We have encountered both of those last summands in the proof Lemma 45 and analogously to there (where we have concluded the finiteness of the L^2-norm), we can now conclude the finiteness of the $L^1(0,1)$-norm of these summands and thus the claim that $\|u_t\|_{L^2(\alpha_t)} \in L^1(0,1)$.

Finally, observe that the construction of u_t from (v_t, w_t) is a linear and bounded map $A_\lambda \colon L^2(M,\mu_t) \oplus L^2(M,\nu_t) \to L^2(M,\alpha_t)$, as the formula in the proof of the L^2-property of u_t shows. Composition of A_λ with $dF \oplus dG$ and the projection to the tangent space then defines the derivative of $\lambda F + (1-\lambda)G$ and shows that this convex combination is differentiable. □

A Disintegration theorem

To be able to prove Theorem 16, we rely on the following statement (see [AGS08]).

Theorem 46. *Let X and Y be Radon spaces. Furthermore let $\mu \in \mathcal{P}(X)$ and $f : X \to Y$ be a measurable map. Then there exists a $f_\#\mu$-almost everywhere uniquely determined family of probability measures $\{\mu_y\}_{y\in Y}$ on X such that*

- *for every measurable set $A \subset X$ the map $y \mapsto \mu_y(A)$ is measurable,*

- $\mu_y(X \setminus f^{-1}(y)) = 0$ *for* $f_\#\mu$*-almost every* $y \in Y$,
- *for every measurable function* $g : X \to [0, \infty]$ *it is*

$$\int_X g(x)\, d\mu(x) = \int_Y \int_{f^{-1}(y)} g(x)\, d\mu_y(x) df_\#\mu(y).$$

This means in particular that any $\mu \in \mathcal{P}(X \times Y)$ whose first marginal ν is given can be represented in this disintegrated way.

On the other hand, whenever there is a measurable (in the sense of the first item above) family $\mu_x \in \mathcal{P}(Y)$ given, for any $\nu \in \mathcal{P}(X)$ the following formula defines a unique measure $\mu \in \mathcal{P}(X \times Y)$:

$$\mu(f) = \int_X \left(\int_Y f(x, y)\, d\mu_x(y) \right) d\nu(x),$$

with $f : X \times Y \to \mathbb{R}$ being a nonnegative measurable function. In this sense, disintegration can be seen as an opposite procedure to the construction of a product measure.

References

[AG13] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows*. Birkhäuser, 2008.

[BB99] Jean-David Benamou and Yann Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. *Contemporary Mathematics*, 226:1–12, 1999.

[Gig08] Nicola Gigli. *On the geometry of the space of measures in* $\mathbb{R}^d$ *endowed with the quadratic optimal transportation distance*. PhD thesis, Scuola Normale Superiore, Pisa, 2008.

[Gig12] Nicola Gigli. *Second Order Analysis on* $(P_2(M), W_2)$. American Mathematical Society, 2012.

[KM97] Andreas Kriegl and Peter Michor. *The Convenient Setting of Global Analysis*. American Mathematical Society, sep 1997.

[Lot07] John Lott. Some geometric calculations on Wasserstein space. *Communications in Mathematical Physics*, 277(2):423–437, nov 2007.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, jan 2001.

[Vil03] Cédric Villani. *Topics in Optimal Transportation (Graduate Studies in Mathematics, Vol. 58)*. American Mathematical Society, 2003.

[Vil08] Cédric Villani. *Optimal Transport: Old and New (Grundlehren der mathematischen Wissenschaften)*. Springer, 2008.