TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Otto Calculus or The Weak Riemannian Structure of $(\mathcal{P}_2(M), \mathcal{W}_2)$

Ausgeführt am Institut für
Diskrete Mathematik und Geometrie
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof. Dipl.-Ing. Dr.techn. Monika Ludwig
und
Dipl.-Ing. Dr.techn. Gabriel Maresch

durch
Philipp Kniefacz, BSc

Wien, März 2017

Contents

Acknowledgements i

Introduction iii

I. Preliminaries 1

1. Riemannian Geometry Preliminaries 3

1.1. Differentiable Manifolds and Smooth Maps 3
1.2. The Tangent Bundle and the Cotangent Bundle 4
1.3. Riemannian Metrics 6
1.4. Affine Connections, Parallel Transport and the Levi-Civita Connection . . 6
1.5. L^2_μ-vector fields on M 8

2. Absolutely Continuous Curves, Narrow Topology and Metric Derivatives 9

2.1. Absolutely Continuous Curves and Metric Derivatives 9
2.2. Narrow Topology 10

3. Optimal Transport Preliminaries 11

3.1. Optimal Transport Formulation 11
3.2. Existence of Optimal Transport Maps 14
3.3. The Wasserstein Space and its Topology 14
3.4. Geodesics in $(\mathcal{P}_2(M), \mathcal{W}_2)$ 15

II. First Order Analysis 17

4. The Continuity Equation 19

4.1. The Continuity Equation and First Properties 19
4.2. Vector Fields of Minimal Norm 21

5. The Tangent Space $\mathrm{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)$ **27**

5.1. The Tangent Space $\mathrm{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)$ 27
5.2. Picard-Lindelöf on Manifolds 30

III. Second Order Analysis 33

6. Parallel Transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$ 35
6.1. Parallel Transport without Local Charts in $\mathbb{R}^n$ 35
6.2. Regular Curves 42
6.3. Vector Fields along Regular Curves 45
6.4. Parallel Transport in the Wasserstein Setting 47

7. Covariant Derivative in $(\mathcal{P}_2(M), \mathcal{W}_2)$ 55
7.1. Levi-Civita Connection 55

Bibliography 59

Acknowledgements

An dieser Stelle möchte ich meinen Dank all jenen aussprechen, ohne deren Hilfe und Zutun ich es nicht so weit geschafft hätte.

Zuallererst gilt mein Dank Professor Monika Ludwig. Sie gab mir nicht nur die Möglichkeit diese Arbeit zu verfassen, sondern ermöglichte es mir auch am Institut für diskrete Mathematik und Geometrie wertvolle Erfahrungen in der Lehre zu sammeln. Besonders bedanken möchte ich mich auch bei Dr. Gabriel Maresch. Dafür, dass er meine Aufmerksamkeit und mein Interesse auf den Optimalen Transport gelenkt hat, dafür, dass er mir im Laufe der Verfassung dieser Arbeit immer mit Rat und Tat zur Seite stand und zu guter Letzt dafür, dass er mich auch bei der Gestaltung meiner zukünftigen Laufbahn sehr unterstützt hat.

Tiefste Dankbarkeit empfinde ich für meine Eltern, auf die ich mich all die Studienjahre immer verlassen konnte und die mir in jeder Situation beistanden, sei es emotional oder finanziell. Ich möchte mich auch bei meinem Bruder herzlich bedanken, der immer um mein Wohl besorgt war und auf den ich stets zählen konnte.
Besondere Dankbarkeit schulde ich meiner Freundin Niki. Sie musste alle Gemütszustände, die mir das Verfassen dieser Arbeit – und das Mathematikstudium im Allgemeinen – abverlangt hat, überstehen. Sie schaffte es dabei dennoch, niemals die Geduld mit mir zu verlieren und auch noch so kräftezehrende Stunden in erfreuliche Augenblicke zu verwandeln.

Ich danke herzlichst all meinen Freunden und Studienkollegen, die mich die letzten Jahre begleitet und meinen Alltag mitgeprägt und bereichert haben. Insbesondere möchte ich hierbei Andi, Claudia, Dominik, Isi, Ludwig, Philipp und Vici meinen Dank aussprechen, für all die glücklichen und traurigen, stressvollen und entspannten, demotivierenden und inspirierenden Augenblicke, die wir gemeinsam erlebt haben.

Last but not least, danke ich Denis, Felix, Lena, Stephan und Thomas, meinen Bürokollegen und Freunden aus dem 7. Stock, die mir den Arbeitsalltag erleichtert haben und dank A-capella-Gesang, Inlinehockey, Fußball und 3D Druck stets für unterhaltsame, abwechslungsreiche und durchaus erwünschte Ablenkung gesorgt haben.

Introduction

In the theory of optimal transportation, the set $\mathcal{P}_2(M)$ of Borel probability measures with finite second moment on a smooth and compact Riemannian manifold M, endowed with the quadratic Wasserstein distance $\mathcal{W}_2$ is well known. However, it was only in recent years that one started to investigate and understand the *differential* structure of this so called *Wasserstein space* $(\mathcal{P}_2(M), \mathcal{W}_2)$.
Otto [Ott01] was the first to consider the *continuity equation* as a possibility to endow $(\mathcal{P}_2(M), \mathcal{W}_2)$ with a Riemannian-like structure. He formally defined tangent vectors to a Borel probability measure $\mu \in \mathcal{P}_2(M)$ via the continuity equation

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0$$

by $-\operatorname{div}(v\mu)$, for $v \in L^2_\mu$, and endowed this "tangent bundle" with the L^2-scalar product

$$\langle -\operatorname{div}(v\mu), -\operatorname{div}(w\mu) \rangle := \int \langle v, w \rangle \, d\mu.$$

Other authors ([AGS08], [AG$^+$08], [Lot08]) then improved Otto's results by starting to rigorously construct the so called *weak Riemannian structure* of $(\mathcal{P}_2(M), \mathcal{W}_2)$. The aim of this thesis is to give an overview of the construction of the weak Riemannian structure, by introducing concepts known from Riemannian manifolds such as *tangent space*, *parallel transport* or *Levi-Civita connection* in the Wasserstein setting.

This thesis is divided into three parts. Part I recalls the most important results used in the course of this thesis. Chapter 1 deals with the basics of Riemannian Geometry. Chapter 2 covers the most important results about absolutely continuous curves and the weak convergence of measures. In this chapter we also define an important map associated to a given absolutely continuous curve c, namely the *metric derivative* of c, denoted by $|\dot{c}|$. The last chapter of the introductory part, Chapter 3, recalls the most important results and concepts concerning optimal transport. Therein, we formulate the optimal transport problem, that is finding a minimizer of

$$\gamma \mapsto \int_{X \times Y} c(x, y) d\gamma(x, y),$$

and present results concerning the existence of such minimizers. Especially the notion of a *c-cyclical monotone set* will prove to be very useful. For *optimal transport maps* T, that is when looking for measures of the type $\gamma = (\mathrm{id}, T)_{\#}\mu$ with a measurable function T, we present necessary and sufficient conditions for their existence. In this chapter we also introduce the *Wasserstein space* $\mathcal{P}_2(X)$ with the associated *Wasserstein distance* $\mathcal{W}_2$, which is a metric space whose differential structure we study in this thesis. The remainder of the thesis will be concerned with the construction of a structure resembling the one of a Riemannian manifold on $(\mathcal{P}_2(M), \mathcal{W}_2)$, where M is a smooth and compact Riemannian manifold.

Part II starts with the construction of the so called *weak Riemannian structure* of $(\mathcal{P}_2(M), \mathcal{W}_2)$, by introducing the *continuity equation*

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0$$

and analyzing some of its properties. This equation has to be understood in the distributional sense. We extend the heuristical formalism developed by Otto and rigorously define a tangent bundle. In Chapter 4 we summarize a few results about the continuity equation and prove two important theorems: To a given absolutely continuous curve (μ_t), the first theorem provides a unique vector field v_t, such that

$$\|v_t\|_{\mu_t} \leq \|\tilde{v}_t\|_{\mu_t}$$

for all $\tilde{v}_t$ which also solve the continuity equation with respect to μ_t. This unique vector field is called the *velocity vector field of* μ_t and plays a very important role in the construction of the parallel transport on $(\mathcal{P}_2(M), \mathcal{W}_2)$. The second important result in this chapter is the so called *Benamou-Brenier formula*

$$\mathcal{W}_2(\mu, \nu) = \inf \int \|\tilde{v}_t\|_{\tilde{\mu}_t} \, dt,$$

where the infimum is taken over all solutions $(\tilde{\mu}_t, \tilde{v}_t)$ of the continuity equation satisfying $\tilde{\mu}_0 = \mu$ and $\tilde{\mu}_1 = \nu$.
In Chapter 5 we derive two explanations, why the tangent bundle to a given measure μ should be considered as

$$\operatorname{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right) := \overline{\{\nabla\phi \mid \phi \in C_c^\infty(M)\}}^{L^2_\mu(M)}.$$

By Theorem 4.2.1 it will be concluded that only gradients of test functions (and their limits) have to be considered in the tangent bundle. Moreover, the Benamou-Brenier formula suggests to use the L^2-product as scalar product on the tangent bundle. On the other hand, we want the unique velocity vector field to always lie in the tangent bundle. As we will show, these requirements lead to the same definition of $\operatorname{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)$.

Part III deals with the second order analysis of $(\mathcal{P}_2(M), \mathcal{W}_2)$. In this part we define a parallel transport and furthermore derive a definition for the *Levi-Civita connection.* Chapter 6 deals in great detail with the construction of the parallel transport. Taking a manifold embedded in $\mathbb{R}^n$ into consideration, a construction of the parallel transport in Euclidean space without using a covariant derivative is described. An important observation is the Lipschitz continuity of the angle between two subspaces V_t and V_s.

Equipped with the tools developped in this section, we mimic the proofs of the Euclidean case in the Wasserstein setting. To that aim we first define the so called *regular curves*, which replace the smooth curves known from classical Riemannian geometry.
An important role in the construction of the parallel transport play translation maps τ_s^t. Since two different tangent vectors $v_t \in \mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M))$ and $v_s \in \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M))$ live in different L^2-spaces and are therefore not directly comparable, we first need a method to translate v_t to a vector field in $\mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M))$. The translation maps τ_s^t provide us with such a translation and allow us to define the *total derivative* by

$$\frac{\mathbf{d}}{dt}v_t = \lim_{s \to t} \frac{\tau_s^t(v_s) - v_t}{s-t}.$$

Additionally, by defining *vector fields along curves*, we are able to construct the parallel transport in the Wasserstein setting and prove its uniqueness. Finally, the notion of parallel transport can naturally be defined by

$$P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right) = 0,$$

where P_{μ_t} denotes the projection onto $\mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M))$. As in the Euclidean case, a Lipschitz-type inequality enables us to carry out the whole construction.
Chapter 7 is dedicated to the definition of the Levi-Civita connection on $(\mathcal{P}_2(M), \mathcal{W}_2)$. With the parallel transport maps $\mathcal{T}_s^t$ at hand, the definition of the covariant derivative reads

$$\frac{\mathbf{D}}{dt}u_t := \lim_{h \to 0} \frac{\mathcal{T}_{t+h}^t(u_{t+h}) - u_t}{h}.$$

Eventually, we will show that $\frac{\mathbf{D}}{dt}u_t$ indeed defines the Levi-Civita connection, that is we show its compatibility with the metric and that it satisfies the torsion free identity.

Part I.

Preliminaries

1. Riemannian Geometry Preliminaries

We first start by presenting the fundamentals of Riemannian geometry. Ultimately, the goal of this thesis is to define a structure on a certain set, resembling that of a Riemannian manifold, although it will not be such a manifold. Here we introduce the most important definitions and results of Riemannian geometry. We omit proofs in this section, but the material presented here is treated in any introductory course on smooth manifolds (for example [Lee01]) and Riemannian geometry (for example [dCV92] or [GHL04]).

1.1. Differentiable Manifolds and Smooth Maps

Without diving to deep into the theory of smooth manifolds, here are the most important definitions, that enable us to talk about smooth manifolds and smooth maps on manifolds.

Definition 1.1.1 (Topological Manifold)**.** M is called a *topological manifold of dimension n*, if the following three properties are satisfied:

(i) M is Hausdorff,

(ii) M is second countable,

(iii) M is locally euclidean.

Definition 1.1.2 (Chart, Local Coordinates)**.** A *chart* is a pair (U, ϕ), where $U \subseteq M$ is an open subset and $\phi : U \mapsto \tilde{U}$ is a homeomorphism with $\phi(U) = \tilde{U} \subseteq \mathbb{R}^n$.
The component functions $(x^1, x^2, \dots, x^n)$ of ϕ are called the *local coordinates in U*.

Definition 1.1.3 (Smoothly Compatible Charts, Atlas, Smooth Atlas)**.** If (U, ϕ) and (V, ψ) are two charts in M with $U \cap V \neq \emptyset$, then we call the homeomorphism $\psi \circ \phi^{-1} : \phi(U \cap V) \to \psi(U \cap V)$ a *coordinate change* from ϕ to ψ. Two charts are called *smoothly compatible*, if either $U \cap V = \emptyset$ or $\psi \circ \phi^{-1}$ is a diffeomorphism.
An *atlas* $\mathcal{A}$ is a set of charts (U_i, ϕ_i), such that the U_i cover M. An atlas $\mathcal{A}$ is called a *smooth atlas* if every two charts are smoothly compatible.

Definition 1.1.4 (Maximal Atlas, Smooth Manifold)**.** A smooth atlas $\mathcal{A}$ is called a *maximal atlas* if there is no strictly greater atlas $\tilde{\mathcal{A}}$ containing $\mathcal{A}$. A maximal atlas $\mathcal{A}$ is also called a *smooth structure*.
A *smooth manifold* is a pair $(M, \mathcal{A})$, consisting of a topological manifold M and a smooth structure $\mathcal{A}$.

Definition 1.1.5 (Coordinate Representation, Smooth Maps)**.** For a map $f : M \to N$ between two smooth manifolds M and N, and charts (U, ϕ) in M and (V, ψ) in N with $f(U) \subseteq V$, we call the function $\hat{f} : \phi(U) \to \psi(V)$ defined by $\hat{f} = \psi \circ f \circ \phi^{-1}$ the *coordinate representation* of f for (U, ϕ) and (V, ψ).
A map $f : M \to N$ is *smooth*, if for every $p \in M$ there exist charts (U, ϕ) in M and (V, ψ) in N, such that $f(U) \subseteq V$, $p \in U$ and the coordinate representation of f for (U, ϕ) and (V, ψ) is smooth in the usual euclidean sense.

1.2. The Tangent Bundle and the Cotangent Bundle

Here we shortly review the first order analysis of smooth manifolds.

Definition 1.2.1 (Derivation)**.** Let M be a smooth manifold. We call a linear functional $v : C^\infty(M) \to \mathbb{R}$ a *derivation at* $p \in M$, if for every $f, g \in C^\infty(M)$ the *product rule*

$$v(fg) = f(p)v(g) + g(p)v(f)$$

holds.

We will follow [Lee01] by defining the *tangent space at* p as the set of all derivations.

Definition 1.2.2 (Tangent Space at p)**.** The vector space of all derivations at $p \in M$ is called the *tangent space* to M at p and denoted by T_pM. An element of T_pM is called a *tangent vector* of M at p.

However, there are other possible (and more intuitive and geometric) definitions of the tangent space at p. A very useful one is the characterization of tangent vectors as derivatives along curves.

Definition 1.2.3 (Alternative Definition of T_pM)**.** Let M be a smooth manifold and $p \in M$. Two curves γ_1, $\gamma_2 : I \to M$ from an open interval I containing 0 to M, such that $\gamma_1(0) = \gamma_2(0) = p$, are said to be equivalent, if in any local chart (U, ϕ) it holds

$$(\phi \circ \gamma_1)'(0) = (\phi \circ \gamma_2)'(0).$$

We call an equivalence class of curves for this equivalence relation a *tangent vector* to p on M. If γ is a representative of such an equivalence class, then we denote the corresponding tangent vector (i.e. the equivalence class of γ) by γ'. The *tangent space at* p then is, as before, the set of all tangent vectors at p and denoted by T_pM.

We will not make a strict distinction between these two characterizations of the tangent space.
The collection of all tangent spaces at points $p \in M$ is called the *tangent bundle*.

Definition 1.2.4 (Tangent Bundle)**.** The disjoint union of all tangent spaces T_pM at p for every $p \in M$ is called the *tangent bundle* and denoted by TM.

Theorem 1.2.5. *Let M be a smooth manifold of dimension n. If TM denotes its tangent bundle, then TM is a smooth manifold of dimension $2n$.*

Whereas tangent vectors can be seen as generalizations of derivatives of curves, tangent covectors act as generalizations of gradients of smooth functions.

Definition 1.2.6 (Covector Space)**.** Let M be a smooth manifold and $p \in M$. We call the dual space of T_pM its *covector space* and denote it by

$$T_p^*M := (T_pM)^*.$$

We will call an element of T_p^*M a *tangent covector*.

As with the tangent bundle, the union of all covector spaces deserves its own name.

Definition 1.2.7 (Covector Bundle)**.** The disjoint union of all covector spaces T_p^*M at p for every $p \in M$ is called the *covector bundle* and denoted by T^*M.

The covector bundle carries the structure of a smooth manifold:

Theorem 1.2.8. *Let M be a smooth manifold of dimension n. If T^*M denotes its covector bundle, then T^*M is a smooth manifold of dimension $2n$.*

Definition 1.2.9 (Differential)**.** Let M and N be two smooth manifolds and $F : M \to N$ be a smooth map. Then the map $dF_p : T_pM \to T_{F(p)}N$ defined by

$$dF_p(v)(f) = v(f \circ F)$$

is called the *differential* or *push-forward* of F at p.

Definition 1.2.10 (Vector Field, Covector Field)**.** A *smooth vector field* is a smooth map $V : M \to TM$ according to Definition 1.1.5, usually written as $p \mapsto V_p$, such that for every $p \in M$ it also holds $V_p \in T_pM$.
Analogously, a *smooth covector field* is a smooth map $W : M \to T^*M$, such that for every $p \in M$ it also holds $W_p \in T_p^*M$.

In the following we will denote by $\mathfrak{X}(M)$ the set of all smooth vector fields on M and by $\mathfrak{X}_c(M)$ the set of all smooth vector fields on M with compact support.
One can also interpret smooth vector fields as maps $C^\infty(M) \to C^\infty(M)$ by associating to a vector field $Y \in \mathfrak{X}(M)$ the map

$$\begin{aligned} Y : C^\infty(M) &\to C^\infty(M) \\ f &\mapsto Yf \end{aligned}$$

where Yf is defined by $(Yf)(p) := Y_p(f)$.
Given two smooth vector fields X and Y, the *Lie bracket* provides us with a way to compute a third smooth vector field $[X, Y]$.

Definition 1.2.11 (Lie Bracket)**.** Let X and Y be smooth vector fields on a smooth manifold M. Their *Lie bracket* is another smooth vector field defined by

$$[X, Y]\, f := X(Yf) - Y(Xf) \quad \forall f \in C^\infty(M).$$

1.3. Riemannian Metrics

Now that we have a basic understanding of smooth manifolds, let us introduce a special type of smooth manifold, a so called Riemannian manifold, that exhibits a metric on each tangent space T_pM.

Definition 1.3.1 (Riemannian Metric, Riemannian Manifold)**.** Let M be a smooth manifold. A *Riemannian metric* is a map g, that associates to every $p \in M$ an inner product $g_p(\cdot, \cdot) =: \langle \cdot, \cdot \rangle_p$ on the tangent space T_pM which varies smoothly in p, i.e. for given smooth vector fields $X, Y \in \mathfrak{X}(M)$ the map $p \mapsto g_p(X_p, Y_p)$ is a smooth map.
We call a smooth manifold together with a given Riemannian metric a *Riemannian manifold*.

We will denote the Riemannian metric by $\langle \cdot, \cdot \rangle$. As a first important result we obtain that every smooth manifold admits a Riemannian metric.

Theorem 1.3.2. *Let M be a smooth manifold. Then there exists a Riemannian metric on M.*

A Riemannian metric allows us to define the notion of length of a curve through its corresponding tangent vector and to introduce a metric on M.

Definition 1.3.3 (Length of a Curve)**.** Let M be a Riemannian manifold and $\gamma : [a, b] \to M$ a curve. We define the length of γ as

$$L(\gamma) = \int_a^b \sqrt{\langle \gamma', \gamma' \rangle_{\gamma(t)}} dt.$$

Definition 1.3.4 (Metric on M)**.** Let M be a Riemannian manifold. We can define a metric on M by

$$d_M(x, y) := \inf \{ L(\gamma) \mid \gamma : [0, 1] \to M, \gamma(0) = x, \gamma(1) = y \}.$$

The metric defined this way recovers the topology of the manifold.
As a corollary of Theorem 1.3.2 we get, that every smooth manifold is metrizable.

1.4. Affine Connections, Parallel Transport and the Levi-Civita Connection

In this section we are going to recall the definitions and properties of affine connections, in particular the Levi-Civita connection, and of parallel transports.

Definition 1.4.1 (Affine Connection)**.** Given a Riemannian manifold M and a map

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) : (X, Y) \mapsto \nabla_X Y,$$

we say that ∇ is an *affine connection* on M if, for every $X, Y, Z \in \mathfrak{X}(M)$ and any real valued smooth functions f, g on M, it satisfies the following three properties:

(i) $\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z$,

(ii) $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$,

(iii) $\nabla_X(fY) = f\nabla_X Y + X(f)Y$.

Proposition 1.4.2 (Covariant Derivative)**.** *Let M be a smooth manifold, ∇ an affine connection on M and V, W vector fields along a curve $\gamma : I \to M$. There exists a unique vector field $\frac{\boldsymbol{D}V}{dt}$ along γ, such that for any smooth function f on I*

(i) $\frac{\boldsymbol{D}}{dt}(V+W) = \frac{\boldsymbol{D}V}{dt} + \frac{\boldsymbol{D}W}{dt}$,

(ii) $\frac{\boldsymbol{D}}{dt}(fV) = \frac{df}{dt}V + f\frac{\boldsymbol{D}V}{dt}$,

(iii) if $V(t) = Y(\gamma(t))$ for $Y \in \mathfrak{X}(M)$, then $\frac{\boldsymbol{D}V}{dt} = \nabla_{\dot\gamma}Y$.

This unique vector field $\frac{\boldsymbol{D}V}{dt}$ is called the covariant derivative*.*

We have defined a connection as map $\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, however it is equivalently possible to define a connection as map $T_pM \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, since for the computation of $(\nabla_X Y)(p)$ it is not necessary to know X, but only X_p. For $p \in M$ this allows us to define the *gradient of a vector field Y* from T_pM into itself by

$$\nabla Y(p) : v \mapsto (\nabla_v Y)(p).$$

Covariant derivatives allow us to define a notion of parallelism.

Definition 1.4.3 (Parallel Vector Field)**.** Let M be a smooth manifold and ∇ an affine connection on M. We say a vector field V along a curve γ is *parallel*, if $\frac{\mathbf{D}V}{dt} = 0$.

Proposition 1.4.4 (Parallel Transport)**.** *Let M be a smooth manifold with affine connection ∇. Furthermore, let $\gamma : I \to M$ be a curve in M and $V_0 \in T_{\gamma(t_0)}M$ for $t_0 \in I$. There exists a unique parallel vector field V along γ, such that $V(t_0) = V_0$. We call V the* parallel transport of V_0 along γ*.*
The linear map $\mathrm{P}_t : T_{\gamma(t_0)}M \to T_{\gamma(t)}M$ which associates to every $V_0 \in T_{\gamma(t_0)}M$ the tangent vector $V(t) \in T_{\gamma(t)}M$, where V is the unique parallel transport of V_0 along γ, is called the parallel transport map*.*

Usually, Proposition 1.4.4 is shown as soon as one has defined the covariant derivative. However, as we will see in Chapter 6, it is possible to construct a parallel transport in $\mathbb{R}^n$ without using local charts or a covariant derivative. This approach will be a key observation for the construction of parallel transport (and the Levi-Civita connection) in the space $(\mathcal{P}_2(M), \mathcal{W}_2)$.

Next let us recall what we mean by *compatibility with the metric* and *torsion free identity.*

Definition 1.4.5 (Compatibility with the Metric)**.** Let M be a Riemannian manifold, ∇ an affine connection on M and $\langle \cdot, \cdot \rangle$ a Riemannian metric. We say that ∇ is *compatible with the metric*, if

$$\langle X, Y \rangle = const$$

for any curve γ and for any parallel vector fields X, Y along γ, or equivalently, if

$$\frac{d}{dt} \langle X, Y \rangle = \left\langle \frac{\mathbf{D}X}{dt}, Y \right\rangle + \left\langle X, \frac{\mathbf{D}Y}{dt} \right\rangle .$$

Definition 1.4.6 (Torsion Free Identity)**.** Let M be a Riemannian manifold, ∇ an affine connection on M and $\langle \cdot, \cdot \rangle$ a Riemannian metric. We say that ∇ satisfies the *torsion free identity* (or sometimes that ∇ is *symmetric*), if

$$\nabla_X Y - \nabla_Y X = [X, Y]$$

for all $X, Y \in \mathfrak{X}(M)$.

Theorem 1.4.7 (Levi-Civita)**.** *Let M be a Riemannian manifold with Riemannian metric $\langle \cdot, \cdot \rangle$. Then there exists a unique affine connection ∇ on M that is compatible with the metric and that satisfies the torsion free identity. We will call this unique affine connection the* Levi-Civita connection*.*

For a smooth function f on M, the Riemannian metric allows us to define a generalization of a gradient to manifolds. We define the *gradient of f* as the vector field $\nabla f \in \mathfrak{X}(M)$ which satisfies

$$g(\nabla f, X) = Xf$$

for all vector fields $X \in \mathfrak{X}(M)$.

1.5. L^2_μ-vector fields on M

Let μ be a Borel probability measure on M and denote by $\langle \cdot, \cdot \rangle_p$ the scalar product on T_pM (that is $\langle \cdot, \cdot \rangle$ is the Riemannian metric on TM). Let furthermore u and v be two maps from M to TM such that $u_p := u(p) \in T_pM$ and analogously $v_p \in T_pM$ for μ-a.e. $p \in M$. Then we define the L^2_μ-scalar product of u and v by

$$\langle u, v \rangle_\mu := \int_M \langle u, v \rangle_p \, d\mu(p)$$

with the induced L^2_μ-norm

$$\|u\|^2_\mu := \int_M \langle u, u \rangle_p \, d\mu(p).$$

We say that $u \in L^2_\mu$ if $\|u\| < \infty$. Elements of L^2_μ will also be called *vector fields*. We will sometimes write $L^2_\mu(M)$ to explicitly mention the underlying manifold M. For example, $L^2_\mu(\mathbb{R}^n)$ denotes the space of vector fields on $\mathbb{R}^n$, that is maps from $\mathbb{R}^n$ to $T_p\mathbb{R}^n \cong \mathbb{R}^n$ (not to be confused with L^2_μ-functions from $\mathbb{R}^n$ to $\mathbb{R}$).

2. Absolutely Continuous Curves, Narrow Topology and Metric Derivatives

This chapter shortly summarizes the most important results about convergence of measures, absolutely continuous curves and their metric derivatives. All the results stated here (and their proofs) can be found in [AGS08].

2.1. Absolutely Continuous Curves and Metric Derivatives

Let us start with the important notion of an absolutely continuous curve.

Definition 2.1.1 (Absolutely Continuous Curve)**.** Let (X, d) be a complete metric space and consider a curve $c : I \to X$ on an open interval I. We say that c is an *absolutely continuous curve*, if there exists a measurable function $f \in L^1(I)$ such that

$$d(c(s), c(t)) \leq \int_s^t f(r)dr \quad \forall s, t \in I, s < t. \tag{2.1}$$

By $\mathrm{AC}(I, X)$ we denote the set of all absolutely continuous curves from I to X.

Remark 2.1.2. Note that the definition above (and the result of the following Theorem 2.1.3) can be generalized to curves c, such that the right-hand side of (2.1) considers functions $f \in L^p(I)$ for $p \geq 1$. We would then write $c \in \mathrm{AC}^p(I, X)$. However we will not need such curves in the following. ★

The following theorem characterizes a minimal right-hand side of (2.1).

Theorem 2.1.3. *If (X, d) is a complete metric space, $I \subset \mathbb{R}$ an open interval and $c \in \mathrm{AC}(I, X)$ an absolutely continuous curve, then the map $|\dot{c}| : I \to \mathbb{R}$, defined by*

$$|\dot{c}|(t) := \lim_{s \to t} \frac{d(c(s), c(t))}{|s - t|},$$

exists for λ^1-a.e. $t \in I$. Furthermore, the following three properties hold:

(i) $|\dot{c}| \in L^1(I)$,

(ii) $|\dot{c}|$ *qualifies as integrand for the right-hand side of* (2.1), *i.e.*

$$d(c(s), c(t)) \leq \int_I |\dot{c}|(r)dr,$$

(iii) among all possible functions f that qualify as integrands for the right-hand side of (2.1)*, $|\dot{c}|$ is λ^1-a.e. minimal, i.e.*

$$|\dot{c}|(t) \leq f(t) \quad \textit{for } \lambda^1\textit{-a.e. } t \in I.$$

Definition 2.1.4 (Metric Derivative)**.** Given an absolutely continuous curve $c(t)$, then we call the map $|\dot{c}|(t)$ provided by Theorem 2.1.3 its *metric derivative.* We will sometimes write $|\dot{c}_t|$ as a shorthand for $|\dot{c}|(t)$.

Lemma 2.1.5 (Arc-Length Reparametrization)**.** *Given an absolutely continuous curve $c(t)$, then there exists a reparametrization, such that $|\dot{c}| = 1$.*

2.2. Narrow Topology

We denote the set of all Borel probability measures on X by $\mathcal{P}(X)$. Often we will need the convergence of measures in the following sense:

Definition 2.2.1 (Narrow Convergence)**.** Let $(\mu_n) \subset \mathcal{P}(X)$ be a sequence of Borel probability measures. We say (μ_n) converges *narrowly* to a measure μ, symbolically $\mu_n \to \mu$, if

$$\lim_{n\to\infty} \int_X f d\mu_n = \int_X f d\mu$$

for every continuous and bounded function $f \in C_b(X)$.

In the special case $X = \mathbb{R}^n$ it is sufficient to only consider test functions:

Lemma 2.2.2. *If $(\mu_k) \subset \mathcal{P}(\mathbb{R}^n)$ is a sequence of Borel probability measures, then μ_k already converges narrowly to a measure μ if*

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} f d\mu_k = \int_{\mathbb{R}^n} f d\mu \quad \forall f \in C_c^\infty(\mathbb{R}^n).$$

When dealing with lower or upper semicontinuous functions, we have the following Fatou like property:

Lemma 2.2.3. *If $\mu \in \mathcal{P}(X)$ and $(\mu_n) \subset \mathcal{P}(X)$ such that $\mu_n \to \mu$ narrowly, then for every lower semicontinuous function f bounded from below it holds*

$$\liminf_{n\to\infty} \int_X f d\mu_n \geq \int_X f d\mu.$$

Using $-f$ instead of f in the lemma above, we get the analogous inequality for upper semicontinuous functions bounded from above.

3. Optimal Transport Preliminaries

In this chapter we introduce the most important results concerning Optimal Transport. The reader should be familiar with the results presented here, as we will use them throughout the rest of this thesis. Again we will not give any proofs. For a thorough introduction to optimal transportation theory (where all the omitted proofs can be found) the reader is encouraged to advice [Vil08]. A shorter, but still very clear and informative introduction, can be found in [AG13].

3.1. Optimal Transport Formulation

Before we introduce the problem formulation of optimal transport, we have to establish a few definitions and notations.
As before, we denote by $\mathcal{P}(X)$ the set of all Borel probability measures on X. We will normally denote measures by μ, ν and γ. Given two Polish spaces X and Y, a measure $\mu \in \mathcal{P}(X)$ and a measurable function $T : X \to Y$, we define the measure $T_{\#}\mu \in \mathcal{P}(Y)$, the so called *push-forward of μ through T*, by

$$T_{\#}\mu(A) := \mu(T^{-1}(A)) \text{ for every Borel set } A \subset Y.$$

For two measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, a measure $\gamma \in \mathcal{P}(X \times Y)$ is called a *transport plan* or *transference plan* from μ to ν, if its marginals equal μ and ν, in other words if

$$\pi^X{}_{\#}\gamma = \mu, \quad \pi^Y{}_{\#}\gamma = \nu$$

where π^X, π^Y are the projections onto X and Y respectively. We will denote the set of all transport plans from μ to ν with $\mathrm{trp}(\mu, \nu)$. We can now state the *Monge-Kantorovitch formulation* of Optimal Transport:

Problem 3.1.1 (Monge-Kantorovitch minimization problem)**.** *Let $(X, \mu), (Y, \nu)$ be two Polish probability spaces and let $c : X \times Y \mapsto \mathbb{R} \cup \{+\infty\}$ be the so called* cost function. *Find a measure $\gamma \in \mathrm{trp}(\mu, \nu)$, such that*

$$\gamma \mapsto \int_{X \times Y} c(x, y) d\gamma(x, y) \tag{3.1}$$

is minimized.

A transport plan γ which minimizes (3.1) is called an *optimal transport plan.* The *transport cost* of a given transport plan γ is defined by

$$C(\gamma) := \int_{X\times Y} c(x,y)d\gamma(x,y).$$

A transport plan γ with $C(\gamma) < \infty$ is called a *finite transport plan.*
The following classical existence result holds:

Theorem 3.1.2. *Under the assumptions of (3.1), if the cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and if furthermore $c(x,y) \geq a(x) + b(y)$ $\forall x \in X$, $\forall y \in Y$ holds for two upper semicontinuous functions $a : X \to \mathbb{R} \cup \{-\infty\}$ and $b : Y \to \mathbb{R} \cup \{-\infty\}$, such that $a \in L^1(\mu)$ and $b \in L^1(\nu)$, then (3.1) admits a minimizer.*

Definition 3.1.3 (c-cyclical Monotonicity)**.** Let $\Gamma \subset X \times Y$. If for every $n \in \mathbb{N}$ and every family $(x_1, y_1), \dots, (x_n, y_n)$, where $(x_i, y_i) \in \Gamma$, the inequality

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i+1})$$

holds, Γ is said to be *c-cyclically monotone.* If a transport plan γ is concentrated on a c-cyclically monotone set, then γ is itself said to be c-cyclically monotone.

The next three results characterize optimality conditions and a duality formulation to (3.1) using the so called *c-cyclical monotonicity.* To fully understand those results, we must first introduce a few definitions. In the following, X and Y can be arbitrary sets and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is a function.

Definition 3.1.4 (c-Transforms)**.** Let $\psi : X \to \mathbb{R} \cup \{+\infty\}$ be a function which is not identically $+\infty$. Its *c-transform* ψ^c is defined by

$$\psi^c(y) = \inf_{x\in X} \big(\psi(x) + c(x,y)\big), \quad \forall y \in Y$$

Let $\phi : Y \to \mathbb{R} \cup \{-\infty\}$ be a function which is not identically $-\infty$. Its *c-transform* ϕ^c is defined by

$$\phi^c(x) = \sup_{y\in Y} \big(\phi(y) - c(x,y)\big), \quad \forall x \in X$$

The functions ψ and ψ^c (respectively ϕ and ϕ^c) are said to be c-conjugate.

Now we can define c-convexity and c-concavity.

Definition 3.1.5 (c-convexity, c-concavity)**.** Let ψ and ϕ be defined as in 3.1.4. If there exists a function $\zeta_1 : Y \to \mathbb{R} \cup \{\pm\infty\}$ such that $\psi = \zeta_1^c$, then ψ is said to be *c-convex.*
If there exists a function $\zeta_2 : X \to \mathbb{R} \cup \{\pm\infty\}$ such that $\phi = \zeta_2^c$, then ϕ is said to be *c-concave.*

Finally, let us define *c-subdifferentials* and *c-superdifferentials.*

Definition 3.1.6. Let ψ and ϕ be defined as in Definition 3.1.4. The c-cyclically monotone set

$$\partial_c\psi := \{(x,y) \in X \times Y \mid \psi^c(y) - \psi(x) = c(x,y)\}$$

is called the *c-subdifferential of ψ*. The set

$$\partial_c\psi(x) := \{y \in Y | (x,y) \in \partial_c\psi\}$$

is called the *c-subdifferential of ψ at x*.
In the same way, the c-cyclically monotone set

$$\partial_c\phi := \{(x,y) \in X \times Y \mid \phi(y) - \phi^c(x) = c(x,y)\}$$

is called the *c-superdifferential of ϕ*.

We are now able to formulate an important characterization of optimal transport plans.

Theorem 3.1.7 (Fundamental Theorem of Optimal Transport)**.** *Under the assumptions of Theorem 3.1.2, let $\gamma \in \mathrm{trp}$. Then the following three statements are equivalent:*

(i) *γ is optimal,*

(ii) $\operatorname{supp}\gamma$ *is c-cyclically monotone,*

(iii) *there exists a c-concave function ϕ such that* $\max\{\phi, 0\} \in L^1(\mu)$ *and such that* $\operatorname{supp}\gamma \subset \partial_c\psi(x)$.

Above theorem shows, that optimality depends only on the support of a transport plan γ, therefore restrictions of optimal transport plans are again optimal.
One can improve the result given in Theorem 3.1.2 by removing the continuity assumptions on the cost function c. The following theorem and its proof can be found in [BGMS09].

Theorem 3.1.8. *Let (X,μ) and (Y,ν) be two Polish probability spaces and let $c : X \times Y \mapsto \mathbb{R} \cup \{+\infty\}$ be a Borel measurable cost function. The following results hold:*

(i) *Every finite optimal transport plan is c-cyclical monotone.*

(ii) *Every finite c-cyclical monotone transport plan is optimal if there exist a closed set F and a $\mu \otimes \nu$ null set N, such that $\{(x,y) \in X \times Y \mid c(x,y) = \infty\} = F \cup N$.*

The last result in this section is a duality formulation of (3.1).

Theorem 3.1.9 (Kantorovitch Duality)**.** *If (X,μ) and (Y,ν) are two Polish probability spaces, $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ a lower continuous cost function, $a \in L^1(\mu)$ and $b \in L^1(\nu)$ two upper semicontinuous functions, such that $c(x,y) \geq a(x) + b(y)$ for all pairs $(x,y) \in X \times Y$, then the following duality holds:*

$$\begin{aligned}
\min_{\gamma \in \mathrm{trp}(\mu,\nu)} \int_{X\times Y} c(x,y)\,d\gamma(x,y) &= \sup_{\psi \in L^1(\mu)} \Big(\int_Y \psi^c(y)\,d\nu(y) - \int_X \psi(x)\,d\mu(x) \Big) \\
&= \sup_{\phi \in L^1(\nu)} \Big(\int_Y \phi(y)\,d\nu(y) - \int_X \phi^c(x)\,d\mu(x) \Big)
\end{aligned}$$

One can restrict the suprema to c-convex functions ψ and c-concave functions ϕ.

3.2. Existence of Optimal Transport Maps

When looking for optimal transport plans, we are specifically interested in so called *optimal transport maps*. Given two measures μ and ν, a transport map is a μ-measurable function T, such that $T_{\#}\mu = \nu$. If $\gamma := (id, T)_{\#}\mu$ is an optimal transport plan, we call T an optimal transport map. For the special case where $X = Y$ is a Riemannian manifold and the cost function $c(x, y)$ is the squared distance on this manifold, we have a useful characterization by so called *regular* measures.

Definition 3.2.1 (Regular Measure)**.** A measure $\mu \in \mathcal{P}M$ is called *regular*, if for any semiconvex function $\psi : M \to \mathbb{R}$ it vanishes on the set of points of non differentiability of ψ.

Remark 3.2.2. Every measure which is absolutely continuous w.r.t. the volume measure is regular. ★

Theorem 3.2.3 (Brenier-McCann)**.** *Let M be a smooth compact Riemannian manifold without boundary and $\mu \in \mathcal{P}(M)$. Then the following statements are equivalent:*

(i) For every $\nu \in \mathcal{P}(M)$ there exists only one transport plan from μ to ν and this plan is induced by a map T.

(ii) μ is regular

If either (i) or (ii) holds, the optimal map can be written as $x \mapsto \exp_x(\nabla\psi(x))$ for some c-convex function $\psi : M \to \mathbb{R}$.

3.3. The Wasserstein Space and its Topology

Until now we worked with two (possibly different) underlying spaces X and Y. Now we turn our attention to the case where X equals Y. We are particularly interested in extending the *optimal transport cost* $C(\mu, \nu) = \inf_{\gamma \in \mathrm{trp}(\mu,\nu)} \int c(x, y) d\gamma(x, y)$ to a metric on a suitable defined subspace of $\mathcal{P}(X)$. When defining the cost function c in terms of a metric d, this goal can be easily achieved. This leads us to following definitions:

Definition 3.3.1 (Wasserstein Distance)**.** Let (X, d) be a Polish metric space and fix $p \in [1, \infty)$. For two probability measures μ, $\nu \in \mathcal{P}(X)$, define the *Wasserstein distance* $\mathcal{W}_p(\mu, \nu)$ *of order* p by

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\gamma \in \mathrm{trp}(\mu,\nu)} \int d(x, y)^p d\gamma(x, y) \right)^{1/p}.$$

We finally arrive at the definition of the space, that we are going to study in more detail in this thesis.

Definition 3.3.2 (Wasserstein Space)**.** Let (X, d) be a Polish metric space and fix $p \in [1, \infty)$. We define the *Wasserstein space of order p* by

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) \;\middle|\; \int_X d(x_0, x)^p d\mu(x) < \infty \right\}$$

for some $x_0 \in X$. Note that this definition does not depend on the particular point $x_0 \in X$.

Lemma 3.3.3. *Let (X, d) be a Polish space and fix $p \in [1, \infty)$. It holds that $\mathcal{W}_p$ is a metric.*

Theorem 3.3.4. *Let (X, d) be a Polish space and fix $p \in [1, \infty)$. Consider a sequence of probability measures $(\mu_k) \subset \mathcal{P}_p(X)$ and $\mu \in \mathcal{P}_p(X)$. Then the following equivalence holds:*

$$\begin{aligned} \mathcal{W}_p(\mu_k, \mu) \to 0 \quad \Leftrightarrow \quad & \mu_k \to \mu \quad \textit{narrowly} \\ & \wedge \int d(x_0, x)^p d\mu_k(x) \to \int d(x_0, x)^p d\mu(x) \quad \textit{for some } x_0 \in X. \end{aligned}$$

3.4. Geodesics in $(\mathcal{P}_2(M), \mathcal{W}_2)$

From now on we will only consider the special case $p = 2$, thus we will be concerned with the structure of $(\mathcal{P}_2(M), \mathcal{W}_2)$, where M is a compact, smooth Riemannian manifold without boundary.

Theorem 3.4.1. *Let M be a Riemannian manifold and (μ_t) a curve in $\mathcal{P}_2(M)$. Then the following two statements are equivalent:*

(i) (μ_t) is a geodesic in $(\mathcal{P}_2(M), \mathcal{W}_2)$

(ii) there exists a plan $\gamma \in \mathcal{P}(TM)$, such that the following two equalities hold:

$$\int |v|^2 d\gamma(x, v) = \mathcal{W}_2^2(\mu_0, \mu_1)$$
$$(EXP(t))_{\#}\gamma = \mu_t$$

where $EXP(t) : TM \to M$ is defined by $(x, v) \mapsto \exp_x(tv)$.

Remark 3.4.2. For the case $M = \mathbb{R}^n$, there exists only one constant speed geodesic connecting any two points $x, y \in \mathbb{R}^n$, namely $t \mapsto (1 - t)x + ty$. It follows that (μ_t) is geodesic if and only if there exists an optimal transport plan γ, such that

$$\mu_t = ((1 - t)\pi^1 + t\pi^2)_{\#}\gamma$$

or if γ is induced by a map

$$\mu_t = ((1 - t)\,\mathrm{id} + tT)_{\#}\mu_0.$$

★

Furthermore, we are interested in interpolating Kantorovitch potentials along such geodesics. This is accomplished through the *Hopf-Lax evolution semigroup*.

Definition 3.4.3 (Hopf-Lax Evolution Semigroup)**.**

$$H_t^s(\psi)(x) := \begin{cases} \inf\limits_{y\in X} c^{t,s}(x,y) + \psi(y) & \text{if } t < s \\ \psi(x) & \text{if } t = s \\ \sup\limits_{y\in X} -c^{t,s}(x,y) + \psi(y) & \text{if } t > s \end{cases},$$

where we have used the *rescaled cost functions*

$$c^{t,s}(x,y) = \frac{d^2(x,y)}{s-t}.$$

Theorem 3.4.4. *Let (X,d) be a geodesic Polish space. If the curve (μ_t) is a constant speed geodesic in $(\mathcal{P}_2(X), \mathcal{W}_2)$ and ψ is a $c^{0,1}$-convex Kantorovitch potential for (μ_0, μ_1), then $\psi := H_0^s(\psi)$ is a $c^{t,s}$-convex Kantorovitch potential for (μ_t, μ_s) for every $t < s$.*

Let us conclude this chapter with an existence and uniqueness result for optimal transport maps along constant speed geodesics.

Theorem 3.4.5. *Let $(\mu_t) \subset \mathcal{P}_2 M$ be a constant speed geodesic in $(\mathcal{P}_2(M), \mathcal{W}_2)$, $t \in (0,1)$ and $s \in [0,1]$. Then there exists only one optimal transport plan from μ_t to μ_s and this transport plan is induced by a Lipschitz map T.*

Part II.

First Order Analysis

4. The Continuity Equation

We start our analysis by introducing the continuity equation in $\mathbb{R}^n$ and developing an understanding of its properties. The continuity equation will enable us to define a tangent space to $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ as a subset of $L^2_\mu(\mathbb{R}^n)$.
Throughout this chapter the underlying space will be $\mathbb{R}^n$. To generalize the results to an arbitrary Riemannian manifold M, one can use Nash's embedding theorem. For an example how this can be done, the reader may take a look at [AG13].
We will not prove every statement, the omitted proofs can be found in [AGS08] or [AG13], which are the main references for this part.

4.1. The Continuity Equation and First Properties

Let us first state the continuity equation. Given a Borel family (μ_t) of probability measures, where t is taken from an open interval $I := (0, T)$, and a Borel vector field $v : I \times \mathbb{R}^n \to \mathbb{R}^n$ satisfying

$$\int_0^T \int_{\mathbb{R}^n} |v_t| \, d\mu_t dt < \infty, \tag{4.1}$$

we say the *continuity equation* holds if

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0. \tag{4.2}$$

This equation has to be understood in the sense of distributions, that is

$$\int_0^T \int_{\mathbb{R}^n} \left(\partial_t \phi(t,x) + \langle v_t(x), \nabla_x \phi(t,x)\rangle\right) d\mu_t(x)dt = 0 \quad \forall \phi \in C_c^\infty\left((0,T) \times \mathbb{R}^n\right) \tag{4.3}$$

or equivalently

$$\frac{d}{dt} \int_{\mathbb{R}^n} \zeta(x) d\mu_t(x) = \int_{\mathbb{R}^n} \langle \nabla \zeta(x), v_t(x)\rangle \, d\mu_t(x) \quad \forall \zeta \in C_c^\infty(\mathbb{R}^n) \tag{4.4}$$

In general, a continuity equation of the form

$$\partial_t \rho + \operatorname{div} u = 0$$

describes the transport of some quantity (such as mass, energy or electric charge) that obeys certain conservation laws. Examples for continuity equations are

- Mass conservation in fluid dynamics: here ρ is the density of the fluid, $u = \rho v$ its mass flux, v its velocity field and the equation essentially states that the amount of mass that leaves a system is equal to the amount of mass that enters the system,

- Probability conservation in quantum mechanics: ρ is a probability density function and the equation states that probability behaves similar to a current, therefore u is called *probability current*,
- Charge conservation in electromagnetic theory: here ρ is the charge density while u is the current density.

We will denote solutions of the continuity equation as pairs (μ_t, v_t). Whenever we say that (μ_t, v_t) is a solution of (4.2), we implicitly assume that v_t satisfies condition (4.1). The first important result about the continuity equation is, that given a family of measures μ_t, we can find a continuous representative. This lemma will allow us to only consider time-continuous curves (μ_t).

Lemma 4.1.1. *Let (μ_t) be a Borel family of probability measures and let v_t be a Borel vector field such that the continuity equation* (4.2) *holds. There exists a continuous representative of μ_t, that is a narrowly continuous curve $(\tilde{\mu}_t)$ such that $\mu_t = \tilde{\mu}_t$ for λ^1-a.e. $t \in (0, 1)$.*

The next property allows us to construct new distributional solutions of (4.2) by time-rescaling.

Lemma 4.1.2. *Let (μ_t, v_t) be a solution of the continuity equation* (4.2) *and let $\gamma(t)$ be a strictly increasing map. If $\bar{\mu}_t$ and $\bar{v}_t$ are defined by $\bar{\mu}_t := \mu_{\gamma(t)}$ and $\bar{v}_t := \gamma(t)' v_t$ respectively, then $(\bar{\mu}_t, \bar{v}_t)$ is a further solution of* (4.2)*.*

Finally, the following two theorems in this section provide a very important representation formula for solutions of the continuity equation and an approximation by curves satisfying special regularity assumptions.

Theorem 4.1.3 (Picard-Lindelöf for Solutions of the Continuity Equation)**.** *Let (μ_t, v_t) be a solution of the continuity equation* (4.2) *and assume that v_t furthermore satisfies*

$$\int_0^T (\sup_B |v_t| + \mathrm{L}_{\mathrm{Lip}}(v_t, \mathbb{R}^n))dt < \infty$$

for every compact set $B \subset \mathbb{R}^n$. Then there exists a unique family of maps $\mathrm{T}(t, s, \cdot) : \mathrm{supp}(\mu_t) \to \mathrm{supp}(\mu_s)$, $t, s \in [0, 1]$, such that the curve $s \mapsto \mathrm{T}(t, s, x)$ is absolutely continuous for every $t \in [0, 1]$, $x \in \mathrm{supp}(\mu_t)$ and such that for every $t \in [0, 1]$ and every $x \in \mathrm{supp}(\mu_t)$ it satisfies

(a) $\mathrm{T}(t, t, x) = x$,

(b) $\frac{d}{ds}\mathrm{T}(t, s, x) = v_s(\mathrm{T}(t, s, x)), \quad$ *for a.e.* $s \in [0, 1]$,

(c) $\mathrm{T}(r, s, \mathrm{T}(t, r, x)) = \mathrm{T}(t, s, x), \quad \forall s, r \in [0, 1]$,

(d) $\mathrm{T}(t, s, \cdot)_{\#}\mu_t = \mu_s, \quad \forall s \in [0, 1]$.

Lemma 4.1.4. *Let μ_t be a solution of the continuity equation (4.2) w.r.t. the vector field v_t, such that*

$$\int_0^T \int_{\mathbb{R}^n} |v_t|^2 \, d\mu_t dt < \infty.$$

It holds that there exists a family $(\mu_t^\epsilon, v_t^\epsilon)_{\epsilon>0}$, such that

(i) *μ_t^ϵ is a continuous solution of the continuity equation (4.2) w.r.t. v_t^ϵ,*

(ii) *v_t^ϵ satisfies $\int_0^T (\sup_B |v_t^\epsilon| + \mathrm{L}_{\mathrm{Lip}}(v_t^\epsilon, \mathbb{R}^n))dt < \infty$ for every compact set $B \subset \mathbb{R}^n$, in particular $(\mu_t^\epsilon, v_t^\epsilon)$ satisfies the conditions of Theorem 4.1.3,*

(iii) *$\mathcal{W}_2(\mu_t^\epsilon, \mu_t) \to 0$ for every $t \in (0,1)$ as $\epsilon \to 0$,*

(iv) $\|v_t^\epsilon\|_{\mu_t^\epsilon} \leq \|v_t\|_{\mu_t} \quad \forall t \in (0,T)$,

(v) *$v_t^\epsilon \mu_t^\epsilon \to v_t \mu_t$ narrowly,*

(vi) $\lim_{\epsilon \to 0^+} \|v_t^\epsilon\|_{\mu_t^\epsilon} = \|v_t\|_{\mu_t}$.

4.2. Vector Fields of Minimal Norm

We can now state and proof our main result of this chapter. The following theorem gives us a complete characterization of absolutely continuous curves in $(\mathcal{P}_2(\mathbb{R}^n), \mathcal{W}_2)$ and allows us to define *velocity vector fields* and a *tangent space.*
As we are only interested in $L^2_\mu(\mathbb{R}^n)$ spaces, we will only formulate the theorem in the case $p = 2$, however it is possible to obtain more general results for $L^p_\mu(\mathbb{R}^n)$ spaces. Those more general results can be found in [AGS08], where the proofs stated here have been taken from.

Theorem 4.2.1. *Let $(\mu_t) : I \to \mathcal{P}_2(\mathbb{R}^n)$ be an absolutely continuous curve on an open interval $I := (0,T) \subseteq \mathbb{R}$ and let $|\dot{\mu}| \in L^1(I)$ be its metric derivative. Then there exists a Borel vector field $v : (x,t) \mapsto v_t(x)$ such that the following holds:*

(i) $v_t \in L^2_{\mu_t}(\mathbb{R}^n)$,

(ii) *$\|v_t\|_{\mu_t} \leq |\dot{\mu}_t|$ for λ^1-a.e. $t \in I$,*

(iii) *the continuity equation holds,*

(iv) *for λ^1-a.e. t, v_t belongs to the L^2-closure of the gradients of test functions, i.e.*

$$v_t \in \overline{\{\nabla\phi \mid \phi \in C_c^\infty(\mathbb{R}^n)\}}^{L^2_{\mu_t}(\mathbb{R}^n)}.$$

If, on the other hand, there exists a narrowly continuous curve $(\mu_t) : I \to \mathcal{P}_2(\mathbb{R}^n)$ and a Borel vector field v_t, such that the continuity equation holds in the distributional sense and such that $\|v_t\|_{L^2(\mu_t, \mathbb{R}^n)} \in L^1(I)$, then the curve (μ_t) is absolutely continuous and

$$|\dot{\mu}_t| \leq \|v_t\|_{L^2(\mu_t, \mathbb{R}^n)}$$

for λ^1-a.e. $t \in I$.

Proof. We begin with the first part of the theorem and therefore assume, that (μ_t) is an absolutely continuous curve. According to Lemma 4.1.2 and Lemma 2.1.5 we may assume that the metric derivative $|\dot{\mu}|$ is constant. For any $\phi \in C_c^\infty(\mathbb{R}^n)$ define the map $H_\phi(x,y)$ by

$$H_\phi(x,y) := \begin{cases} |\nabla\phi(x)| & \text{if } x = y \\ \frac{|\phi(x)-\phi(y)|}{|x-y|} & \text{if } x \neq y \end{cases}.$$

Note that $H_\phi(x,y)$ is upper semicontinuous, i.e. $\limsup_{(x,y)\to(x_0,y_0)} H_\phi(x,y) \le H_\phi(x_0,y_0)$. Letting γ_t^{t+h} be any optimal transport plan for μ_t and μ_{t+h}, we have

$$\begin{aligned} \frac{1}{h}\int_{\mathbb{R}^{2n}} |\phi(x)-\phi(y)|\, d\gamma_t^{t+h} &\le \frac{1}{h}\int_{\mathbb{R}^{2n}} |x-y|\, H_\phi(x,y) d\gamma_t^{t+h} \\ &\le \frac{1}{h}\left(\int_{\mathbb{R}^{2n}} |x-y|^2\, d\gamma_t^{t+h}\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^{2n}} H_\phi^2(x,y) d\gamma_t^{t+h}\right)^{\frac{1}{2}} \qquad (4.5) \\ &= \frac{\mathcal{W}_2(\mu_t,\mu_{t+h})}{h}\left(\int_{\mathbb{R}^{2n}} H_\phi^2(x,y) d\gamma_t^{t+h}\right)^{\frac{1}{2}} \end{aligned}$$

Now we see that both marginals of γ_t^{t+h} (i.e. μ_t and μ_{t+h}) converge narrowly to μ_t, which means $\lim_{h\to 0^+}\gamma_t^{t+h}$ is an optimal transport plan from μ_t to μ_t and is therefore concentrated on the diagonal of $\mathbb{R}^n\times\mathbb{R}^n$. Thus $\lim_{h\to 0^+}\gamma_t^{t+h} = (\mathrm{Id},\mathrm{Id})_\#\mu_t$. Taking the upper semicontinuity of H_ϕ and Lemma 2.2.3 into account and using (4.5), we obtain

$$\begin{aligned} \limsup_{h\to 0^+}\frac{1}{h}\int_{\mathbb{R}^{2n}} |\phi(x)-\phi(y)|\, d\gamma_t^{t+h} &\le \limsup_{h\to 0^+}\frac{\mathcal{W}_2(\mu_t,\mu_{t+h})}{h}\left(\int_{\mathbb{R}^{2n}} H_\phi^2(x,y) d\gamma_t^{t+h}\right)^{\frac{1}{2}} \\ &\le |\dot{\mu}_t|\left(\int_{\mathbb{R}^n} H_\phi^2(x,x) d\mu_t\right)^{\frac{1}{2}} \\ &= |\dot{\mu}_t|\, \|\nabla\phi\|_{\mu_t}. \end{aligned}$$

Let $\mu \in \mathcal{P}(I\times\mathbb{R}^n)$ be the measure whose disintegration is $\{\mu_t\}_{t\in I}$, i.e. $\mu(J\times A) = \int_J \mu_t(A)dt$. For any $\phi\in C_c^\infty(I\times\mathbb{R}^n)$ we have

$$\begin{aligned} \int_{I\times\mathbb{R}^n} \frac{d}{dt}\phi(t,x)d\mu &= \int_{I\times\mathbb{R}^n} \lim_{h\to 0^+}\frac{\phi(t+h,x)-\phi(t,x)}{h} d\mu \\ &= \lim_{h\to 0^+}\int_I \frac{1}{h}\left(\int_{\mathbb{R}^n}\phi(t+h,x)d\mu_t - \int_{\mathbb{R}^n}\phi(t,x)d\mu_t\right)dt \\ &= \lim_{h\to 0^+}\int_I \frac{1}{h}\left(\int_{\mathbb{R}^n}\phi(t,x)d\mu_{t-h} - \int_{\mathbb{R}^n}\phi(t,x)d\mu_t\right)dt \\ &= \lim_{h\to 0^+}\int_I \frac{1}{h}\left(-\int_{\mathbb{R}^{2n}}\phi(t,x)-\phi(t,y)d\gamma_t^{t+h}\right)dt. \end{aligned}$$

Combining this result with (4.5) and using Fatou's lemma one obtains

$$\begin{aligned}\left|\int_{I\times\mathbb{R}^n}\frac{d}{dt}\phi(t,x)d\mu\right| &\le \int_I |\dot\mu_t|\,\|\nabla\phi(t,\cdot)\|_{\mu_t}\,dt \\ &\le \left(\int_I |\dot\mu_t|^2\,dt\right)^{\frac{1}{2}}\left(\int_I \|\nabla\phi(t,\cdot)\|_{\mu_t}\,dt\right)^{\frac{1}{2}}.\end{aligned} \tag{4.6}$$

To simplify notation, let us denote the set of gradients of test functions on $I\times\mathbb{R}^n$ by

$$V_I := \{\nabla\phi \mid \phi\in C_c^\infty(I\times\mathbb{R}^n)\}$$

and define a linear functional L on V_I by

$$L(\nabla\phi) := -\int_{I\times\mathbb{R}^n}\frac{d}{dt}\phi(t,x)d\mu.$$

If we consider the L^2_μ-closure $\overline{V_I} := \overline{V_I}^{L^2_\mu(I\times\mathbb{R}^n)}$, then (4.6) tells us, that $L(\nabla\phi)$ can be uniquely extended to a bounded functional on $\overline{V_I}$. Since $\overline{V_I}$ is a Hilbert space, the Riesz representation theorem assures us the existence of an element $v(t,x)\in\overline{V_I}$ such that $L(\nabla\phi)$ can be represented as scalar product with $v(t,x)$. Setting $v_t(x) := v(t,x)$ we get

$$\int_I\int_{\mathbb{R}^n}\frac{d}{dt}\phi(t,x)d\mu_t dt = -L(\nabla\phi) = \int_I\int_{\mathbb{R}^n}\langle\nabla\phi, v_t\rangle\,d\mu_t dt \quad \forall\phi\in C_c^\infty(I\times\mathbb{R}^n)$$

which is exactly the weak formulation of the continuity equation. Since convergence in L^p implies existence of a subsequence which converges pointwise a.e., we find for any sequence $(\phi_n)\subset V$ with

$$\int\int |\nabla\phi_n(t,\cdot)-v(t,\cdot)|^2\,d\mu_t dt = \int |\nabla\phi_n - v|^2\,d\mu \to 0$$

a subsequence (ϕ_{n_k}), such that

$$\int |\nabla\phi_{n_k}(t,\cdot)-v(t,\cdot)|^2\,d\mu_t \to 0$$

for λ^1-a.e. $t\in I$. That is, for λ^1-a.e. t, v_t even satisfies

$$v_t \in \overline{\{\nabla\phi \mid \phi\in C_c^\infty(\mathbb{R}^n)\}}^{L^2_{\mu_t}(\mathbb{R}^n)}.$$

To conclude the first part, we still need to show, that $\|v_t\|_{\mu_t}$ is bounded by $|\dot\mu_t|$. We already know the estimate

$$\int_I \|v_t\|_{\mu_t}\,dt \le \int_I |\dot\mu_t|\,dt$$

for the interval I, but the same steps can be repeated with any arbitrary interval $J\subseteq I$. Thus for each J, we get a vector $v_t^J\in\overline{V_J}^{L^2_{\mu_J}(\mathbb{R}^n)}$, where $\mu_J := \mu\big|_{J\times\mathbb{R}^n}$ (remember that restrictions of optimal plans are again optimal plans) and such that

$$\int_J \|v_t\|_{\mu_t}\,dt \le \int_J |\dot\mu_t|\,dt.$$

Furthermore (4.2) holds (for J instead of I), which means that $v_t^J = v_t$ μ-a.e. for every $J \subseteq I$ and this finally proves the first claim.
Now we are going to show the converse claim, starting with a narrowly continuous curve (μ_t) and a vector field v_t such that the continuity equations holds and such that $\|v_t\|_{\mu_t} < \infty$. We want to show that μ_t is absolutely continuous and that $\|v_t\|_{\mu_t}$ is bounded from below by $|\dot{\mu}_t|$. To that aim consider the approximations μ_t^ϵ and v_t^ϵ provided by Lemma 4.1.4. They satisfy all the necessary conditions for Theorem 4.1.3. Hence we obtain measurable functions T_t^ϵ, which are solutions of the ordinary differential equation given by $\dot{T}_t^\epsilon(x) = v_t^\epsilon(T_t^\epsilon(x))$ with initial condition $T_0^\epsilon(x) = x$, such that $\mu_t^\epsilon = (T_t^\epsilon)_\# \mu_0^\epsilon$. Thus we can define a transport plan $\gamma := (T_t^\epsilon, T_s^\epsilon)_\# \mu_0^\epsilon \in \operatorname{trp}(\mu_t^\epsilon, \mu_s^\epsilon)$ which allows us to estimate the squared Wasserstein distance with

$$\begin{aligned}
\mathcal{W}_2^2(\mu_t^\epsilon, \mu_s^\epsilon) &\leq \int_{\mathbb{R}^{2n}} |x-y|^2 \, d\gamma \\
&= \int_{\mathbb{R}^n} |T_t^\epsilon(x) - T_s^\epsilon(x)|^2 \, d\mu_0^\epsilon \\
&= \int_{\mathbb{R}^n} \left(\int_t^s \left| \dot{T}_r^\epsilon \right| dr \right)^2 d\mu_0^\epsilon \\
&\leq (s-t) \int_{\mathbb{R}^n} \int_t^s \left| \dot{T}_r^\epsilon \right|^2 dr d\mu_0^\epsilon \\
&= (s-t) \int_t^s \int_{\mathbb{R}^n} |v_r^\epsilon|^2 \, d\mu_r^\epsilon dr \\
&\leq (s-t) \int_t^s \int_{\mathbb{R}^n} |v_r|^2 \, d\mu_r dr
\end{aligned}$$

for any $0 \leq t \leq s \leq 1$, where the last inequality is due to Lemma 4.1.4 (iv). Using the triangle inequality we get

$$\mathcal{W}_2(\mu_t, \mu_s) \leq \mathcal{W}_2(\mu_t, \mu_t^\epsilon) + \mathcal{W}_2(\mu_t^\epsilon, \mu_s^\epsilon) + \mathcal{W}_2(\mu_s^\epsilon, \mu_s)$$

and with Lemma 4.1.4 (iii) we arrive at

$$\mathcal{W}_2^2(\mu_t, \mu_s) \leq (s-t) \int_t^s \int_{\mathbb{R}^n} |v_r|^2 \, d\mu_r dr. \tag{4.7}$$

The final estimate (4.7) holds for arbitrary s and t, therefore we obtain

$$\frac{\mathcal{W}_2^2(\mu_t, \mu_s)}{(s-t)^2} \leq \frac{1}{(s-t)} \int_t^s \|v_r\|_{\mu_r}^2 \, dr$$

and taking the limit $t \to s$ on both sides we end up with the desired estimate

$$|\dot{\mu}_t|^2 = \lim_{t \to s} \frac{\mathcal{W}_2^2(\mu_t, \mu_s)}{(s-t)^2} \leq \lim_{t \to s} \frac{1}{(s-t)} \int_t^s \|v_r\|_{\mu_r}^2 \, dr = \|v_t\|_{\mu_t}^2 .$$

$\square$

Theorem 4.2.1 can be used to show the famous *Benamou-Brenier formula.*

Theorem 4.2.2 (Benamou-Brenier)**.** *The following formula holds:*

$$\mathcal{W}_2^2(\mu_0,\mu_1) = \inf \int \|\tilde{v}_t\|_{\tilde{\mu}_t}^2 \, dt = \left(\inf \int \|\tilde{v}_t\|_{\tilde{\mu}_t} \, dt \right)^2 ,$$

where the infimum is taken over all couples $(\tilde{\mu}_t, \tilde{v}_t)$ *satisfying the continuity equation* (4.2) *such that* $\tilde{\mu}_0 = \mu_0$ *and* $\tilde{\mu}_1 = \mu_1$*.*

Proof. Recall that for an absolutely continuous curve, its metric derivative satisfies

$$\mathcal{W}_2(\mu_s, \mu_t) \le \int_s^t |\dot{\mu}_t| \, dt$$

(see Theorem 2.1.3). Using the second part of Theorem 4.2.1 we get

$$\mathcal{W}_2(\mu_0,\mu_1) \le \int_0^1 |\dot{\mu}_t| \, dt \le \int_0^1 \|v_t\|_{\mu_t} \, dt \le \left(\int_0^1 \|v_t\|_{\mu_t}^2 \, dt \right)^{\frac{1}{2}} .$$

To see that the lower bound $\mathcal{W}_2(\mu_0,\mu_1)$ really can be attained, consider a constant speed geodesic μ_t and recall, that for a constant speed geodesic

$$\mathcal{W}_2(\mu_s,\mu_t) = |t-s| \, \mathcal{W}_2(\mu_0,\mu_1)$$

holds. Thus, the first part of Theorem 4.2.1 yields

$$\begin{aligned}
\left(\int_0^1 \|v_t\|_{\mu_t}^2 \, dt \right)^{\frac{1}{2}} &= \int_0^1 \|v_t\|_{\mu_t} \, dt \\
&\le \int_0^1 |\dot{\mu}_t| \, dt \\
&= \int_0^1 \lim_{s\to t} \frac{\mathcal{W}_2(\mu_s,\mu_t)}{|t-s|} dt \\
&= \int_0^1 \mathcal{W}_2(\mu_0,\mu_1) dt \\
&= \mathcal{W}_2(\mu_0,\mu_1).
\end{aligned}$$

□

Remark 4.2.3. Recall that given a Riemannian manifold M, the metric d on M is defined via its Riemannian metric g by

$$d(x,y) := \inf \int \sqrt{g(\gamma_t', \gamma_t')} dt$$

where the infimum is taken over all curves $\gamma : [0,1] \to M$, such that $\gamma(0) = x$ and $\gamma(1) = y$. We still did not define a tangent space to a measure μ, but under above consideration, the Benamou-Brenier formula suggests to use the L^2_μ-product as the scalar product on the tangent space. ★

5. The Tangent Space $\mathrm{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)$

From now on M will denote a smooth and compact Riemannian manifold without boundary. We can use Nash's embedding theorem to generalize the results from the last chapter to M (see for example [AG13]).

5.1. The Tangent Space $\mathrm{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)$

Let us recap what we have so far:

(i) The second part of Theorem 4.2.1 gives us for any pair (μ_t, v_t), which solves the continuity equation in the distributional sense, a lower bound on the norm of v_t in terms of the metric derivative of μ_t:

$$\|v_t\|_{\mu_t} \geq |\dot{\mu}_t|$$

(ii) On the other hand, the first part of Theorem 4.2.1 says, that for given curve (μ_t) there exists a vector field v_t with minimal norm, i.e.

$$\|v_t\|_{\mu_t} = |\dot{\mu}_t|$$

(iii) This minimal vector field v_t is unique, because given another minimal vector field $\tilde{v}_t$ one could consider the vector field $w := \frac{1}{2}(v_t + \tilde{v}_t)$. This vector field is also a solution of (4.2). But because $\|v_t\|_{\mu_t} = \|\tilde{v}_t\|_{\mu_t} = |\dot{\mu}_t|$, the strict convexity of the L^2-norm gives

$$\|w\|_{\mu_t} = \frac{1}{2}\|v_t + \tilde{v}_t\|_{\mu_t} < \frac{1}{2}\left(\|v_t\|_{\mu_t} + \|\tilde{v}_t\|_{\mu_t}\right) = |\dot{\mu}_t|\,,$$

which is a contradiction.

Let us summarize this results in the following corollary:

Corollary 5.1.1. *Let $(\mu_t) : I \to \mathcal{P}_2(M)$ be an absolutely continuous curve on an open interval $I \subseteq \mathbb{R}$. If $|\dot{\mu}| \in L^1(I)$ is its metric derivative, then there exists a unique Borel vector field v_t with minimal L^2-norm*

$$\|v_t\|_{\mu_t} = |\dot{\mu}_t|$$

such that the continuity equation (4.2) *holds.*

Our aim is to provide a suitable definition for a tangent space to a Borel measure $\mu \in \mathcal{P}_2(M)$. For given absolutely continuous curve $(\mu_t) \subset \mathcal{P}_2(M)$ it would be very desirable to associate a unique "velocity" to it. But if (μ_t, v_t) is a solution to the continuity equation, than (μ_t, v_t+w_t) also solves the continuity equation for every w_t such that $\operatorname{div}(w_t\mu_t) = 0$. That is, using only the continuity equation as condition for a suitable definition of a tangent space will not provide uniqueness. However, Corollary 5.1.1 provides us with a unique vector field which can be used to characterize the tangent space. Because of this considerations, we make the following definition.

Definition 5.1.2 (Velocity Vector Field)**.** Let (μ_t) be an absolutely continuous curve on an open interval. Then we call the unique Borel vector field v_t provided by Corollary 5.1.1 the *velocity vector field* of μ_t.

Let us recall the next lemma, which will provide us with a first possible definition of a tangent space to μ.

Lemma 5.1.3. *Let $\langle \cdot, \cdot \rangle$ denote a scalar product on a vector space X and let $\|x\| := \sqrt{\langle x, x\rangle}$ be the induced norm. Then it holds*

$$\langle x, y\rangle = 0 \Leftrightarrow \|x\| \leq \|x + ty\| \quad \forall t \in \mathbb{R}.$$

Proof. If $\langle x, y\rangle = 0$, then

$$\|x\|^2 = \|x\|^2 + 2t\langle x, y\rangle \leq \|x\|^2 + 2\langle x, ty\rangle + \|y\|^2 = \|x + ty\|^2 .$$

On the other hand, if $\|x\| \leq \|x + ty\|$ for all $t \in \mathbb{R}$, then

$$2t\langle x, y\rangle + t^2\|y\|^2 \geq 0.$$

Since $t \in \mathbb{R}$ is arbitrary, we can choose it such that

$$t^2\|y\|^2 \geq 2t\,|\langle x, y\rangle| \geq 0$$

and letting $t \to 0$ yields the desired result. □

Using Lemma 5.1.3 we can find an equivalent characterization of the minimality condition provided by Theorem 4.2.1.

Lemma 5.1.4. *Let $\mu \in \mathcal{P}_2(M)$ and $v \in L^2_\mu$, such that (μ, v) is a solution of the continuity equation. Then v is minimal amongst all other solution vectors $\tilde{v} \in L^2_\mu$, i.e.*

$$\|v\|_\mu \leq \|\tilde{v}\|_\mu \quad \forall \tilde{v} \in L^2_\mu \text{ such that } (\mu, \tilde{v}) \text{ solves the continuity equation },$$

if and only if

$$v \in T := \left\{ u \in L^2_\mu \;\middle|\; \langle u, w\rangle_\mu = \int_M \langle u, w\rangle_p \, d\mu(p) = 0 \quad \forall w \in L^2_\mu : \operatorname{div}(w\mu) = 0 \right\}.$$

Proof. The minimality condition $\|v\|_\mu \le \|\tilde{v}\|_\mu$ is the same as asking for

$$\|v\|_\mu \le \|v+w\|_\mu \quad \forall w \in L^2_\mu(M) \text{ such that } \operatorname{div}(w\mu)=0,$$

by simply setting $w = \tilde{v} - v$. Applying Lemma 5.1.3 gives the desired result. $\square$

Lemma 5.1.4 suggests to define the tangent space to μ as

$$T^1_\mu := \left\{ v \in L^2_\mu \;\middle|\; \int \langle v, w\rangle \, d\mu = 0 \quad \forall w \in L^2_\mu : \operatorname{div}(w\mu) = 0 \right\}.$$

On the other hand, as we have seen, vector fields solving the continuity equation in the sense of distributions (4.4) can be viewed as functionals that act only on gradients of test functions $\phi \in C^\infty_c(M)$. This observation together with Remark 4.2.3 suggests to define the tangent space by

$$T^2_\mu := \overline{\{\nabla\phi \mid \phi \in C^\infty_c(M)\}}^{L^2_\mu(M)}.$$

However, as Theorem 5.1.6 will show, both spaces coincide. But at first let us find a characterization for the normal space to T^1_μ.

Proposition 5.1.5. *Let $\mu \in \mathcal{P}_2(M)$. The normal space $(T^1_\mu)^\perp$ to T^1_μ can be written as*

$$(T^1_\mu)^\perp = K := \left\{ w \in L^2_\mu \mid \operatorname{div}(w\mu) = 0 \right\}.$$

Proof. By definition of normal space, we have

$$(T^1_\mu)^\perp = \left\{ w \in L^2_\mu \mid \langle v, w\rangle = 0 \quad \forall v \in T^1_\mu \right\}.$$

Since for every $w \in K$ it holds $\langle v, w\rangle = 0$ for arbitrary $v \in T^1_\mu$ by definition of T^1_μ, it clearly holds $K \subseteq (T^1_\mu)^\perp$.
To show that $(T^1_\mu)^\perp \subseteq K$, we are going to show $K^\perp \subseteq ((T^1_\mu)^\perp)^\perp = T^1_\mu$. To that aim choose arbitrary $v \in K^\perp$, that is v satisfies $\langle v, w\rangle = 0$ for all $w \in K$ which is by definition of K the same as $\langle v, w\rangle = 0$ for all $w \in L^2_\mu$ such that $\operatorname{div}(w\mu) = 0$. This directly implies $v \in T^1_\mu$. $\square$

Now we can show the equality of T^1_μ and T^2_μ.

Theorem 5.1.6. *If $\mu \in L^2_\mu(M)$, then $T^1_\mu = T^2_\mu$.*

Proof. First pick $v \in T^2_\mu$ and denote by (ϕ^n) a sequence in $C^\infty_c(M)$, such that $\nabla\phi^n \to v$ in $L^2_\mu(M)$. Then for every $w \in L^2_\mu(M)$, such that $\operatorname{div}(w\mu) = 0$, it holds

$$\langle v, w\rangle = \lim_{n\to\infty} \langle \nabla\phi^n, w\rangle = 0$$

and therefore $v \in T^1_\mu$.
To prove the converse, we are going to show that $(T^2_\mu)^\perp \subseteq (T^1_\mu)^\perp$. Choose $w \in (T^2_\mu)^\perp$. Then, by definition of $(T^2_\mu)^\perp$, w satisfies $\langle \nabla\phi, w\rangle = 0$ for all $\phi \in C^\infty_c(M)$ which is the same as $\operatorname{div}(w\mu) = 0$. But according to Proposition 5.1.5 this is equivalent to $w \in (T^1_\mu)^\perp$. $\square$

From two different points of view we arrived at the same space, suitable to act as tangent space to given Borel measure μ. Therefore we can define the following.

Definition 5.1.7 (Tangent Space to $\mathcal{P}_2(M)$ at μ)**.** Consider a measure $\mu \in \mathcal{P}_2(M)$. Then the *tangent space at* μ is defined by

$$\begin{aligned}\operatorname{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right) &:= \overline{\left\{\nabla\phi \;\mid\; \phi \in C_c^\infty(M)\right\}}^{L^2_\mu(M)} \\ &= \left\{v \in L^2_\mu(M) \;\middle|\; \int \langle v, w\rangle \, d\mu = 0 \quad \forall w \in L^2_\mu(M) : \operatorname{div}(w\mu) = 0\right\}.\end{aligned}$$

Now, using again Proposition 5.1.5, the definition of the normal space naturally follows.

Definition 5.1.8 (Normal Space to $\mathcal{P}_2(M)$ at μ)**.** Let $\mu \in \mathcal{P}_2(M)$ and $\operatorname{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)$ the tangent space at μ. Then the *normal space at* μ is defined by

$$\begin{aligned}\operatorname{Tan}^\perp_\mu\left(\mathcal{P}_2\left(M\right)\right) &:= \left\{w \in L^2_\mu(M) \;\middle|\; \int \langle v, w\rangle \, d\mu = 0 \quad \forall v \in \operatorname{Tan}_\mu\left(\mathcal{P}_2\left(M\right)\right)\right\} \\ &= \left\{w \in L^2_\mu(M) \;\mid\; \operatorname{div}(w\mu) = 0\right\}.\end{aligned}$$

5.2. Picard-Lindelöf on Manifolds

In this section we will first restate the Picard-Lindelöf Theorem 4.1.3 for the more general case of a Riemannian manifold M. Afterwards we are going to define a notion of approximation of solutions (μ_t, v_t) of the continuity equation by *transport couples* (solutions of the continuity equation with special regularity properties) satisfying the conditions for the Picard-Lindelöf theorem and guaranteeing therefore the existence of the flow maps $T(t, s, \cdot)$.

Let us start with the formal definition of a *transport couple*, that is a pair consisting of a curve (μ_t) together with a suitable vector field (for example - but not necessarily - the tangent vector field) (v_t).

Definition 5.2.1 (Transport Couple)**.** Consider a curve $\xi : t \to (\mu_t, v_t)$, with $\mu_t \in \mathcal{P}_2(M)$ and $v_t \in L^2_{\mu_t}$. We call ξ a *transport couple*, if the following two conditions are satisfied:

(i) $\int_0^1 \|v_t\|_{\mu_t} \, dt < \infty$,

(ii) μ_t satisfies the continuity equation (4.2) w.r.t. v_t.

After having defined transport couples, we need to define a notion of convergence for a sequence of transport couples. Before doing so, let us recall what we mean by convergence of vector fields $v \in L^2_\mu$.

Definition 5.2.2 (Convergence of Vector Fields)**.** Let $\mu \in \mathcal{P}_2(M)$, $v \in L^2_\mu(M)$ and let furthermore $\mu^n \in \mathcal{P}_2(M)$ and $v^n \in L^2_{\mu^n}(M)$, $n \in \mathbb{N}$, such that $\mathcal{W}_2(\mu^n, \mu) \to 0$ for $n \to \infty$. Then we say that v^n converges to v if

(i) $\langle v^n, \xi \rangle_{\mu^n} \to \langle v, \xi \rangle_{\mu^n} \quad \forall \xi \in \mathfrak{X}_c(M)$,

(ii) $\lim_{n\to\infty} \|v^n\|_{\mu^n} = \|v\|_{\mu}$.

Definition 5.2.3 (Convergence of Transport Couples)**.** Let (μ_t, v_t) be a transport couple and let (μ_t^n, v_t^n) be a sequence of transport couples. We say that (μ_t^n, v_t^n) converges to (μ_t, v_t), if the following holds for a.e. $t \in [0,1]$:

(i) $\mathcal{W}_2(\mu_t^n, \mu_t) \to 0$ uniformly for $n \to \infty$,

(ii) v_t^n converges to v_t for a.e. $t \in [0,1]$,

(iii) $\lim_{n\to\infty} \int_0^1 \|v_t^n\|_{\mu_t^n} \, dt = \int_0^1 \|v_t\|_{\mu_t} \, dt$.

Finally, we will need the Lipschitz constant for Borel vector fields v:

Definition 5.2.4 (Lipschitz Constant for Tangent Vector Fields)**.** Let $\xi \in \mathfrak{X}_c(M)$ be a tangent vector field on M. We will call the constant

$$\mathrm{L}(\xi) := \sup_{x \in M} \|\nabla \xi(x)\|$$

its *Lipschitz constant.*
For $\mu \in \mathcal{P}_2(M)$, $v \in L^2_\mu$ and $S(v) := \{(\xi_n) \subset \mathfrak{X}_c(M) \mid \xi_n \to v\}$, we call the constant

$$\mathrm{L}(v) := \inf_{S(v)} \liminf_{n\to\infty} \mathrm{L}(\xi_n)$$

the *Lipschitz constant of* v.

With the definition given above, a tangent vector $v \in L^2_\mu$ is said to be Lipschitz, if $\mathrm{L}(v) < \infty$. After having established the necessary notation and definitions in the last section, we can now state the Picard-Lindelöf theorem (again) for a manifold M. This is basically the same statement as its euclidean counterpart Theorem 4.1.3.

Theorem 5.2.5 (Picard-Lindelöf on Manifolds)**.** *Let* (μ_t, v_t), $t \in [0,1]$ *be a transport couple and assume that* v_t *is Lipschitz. Then there exists a unique family of maps* $\mathrm{T}(t,s,\cdot) : \mathrm{supp}(\mu_t) \to \mathrm{supp}(\mu_s)$, $t, s \in [0,1]$, *such that the curve* $s \mapsto \mathrm{T}(t,s,x)$ *is absolutely continuous for every* $t \in [0,1]$, $x \in \mathrm{supp}(\mu_t)$ *and satisfying*

(a) $\mathrm{T}(t,t,x) = x$,

(b) $\frac{d}{ds}\mathrm{T}(t,s,x) = v_s(\mathrm{T}(t,s,x)), \quad$ *for a.e.* $s \in [0,1]$,

(c) $\mathrm{T}(r,s,\mathrm{T}(t,r,x)) = \mathrm{T}(t,s,x), \quad \forall s, r \in [0,1]$,

(d) $\mathrm{T}(t,s,\cdot)_{\#}\mu_t = \mu_s, \quad \forall s \in [0,1]$,

for every $t \in [0,1]$ *and every* $x \in \mathrm{supp}(\mu_t)$.

The next approximation result allows us to approximate transport couples with more regular ones. This result will prove useful later when defining parallel transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$. We will not give a proof here, the interested reader may take a look at chapter 8 in [AGS08] and chapter 2 in [Gig12] (there one can also find an even stronger approximation with tangent vectors v_t^n).

Theorem 5.2.6 (Approximation of Transport Couples)**.** *Let (μ_t, v_t) be a transport couple, such that*

$$\int_0^1 \mathrm{L}(v_t)dt < \infty.$$

Then (μ_t, v_t) can be approximated by transport couples (μ_t^n, v_t^n), such that:

(i) *(μ_t^n, v_t^n) converges to (μ_t, v_t) in the sense of Definition 5.2.3,*

(ii) *v_t are defined for every $t \in [0,1]$ on the whole manifold M and $v_t \in \mathfrak{X}_c(M)$,*

(iii) *the flow maps $T(t,s,\cdot)$ provided by Theorem 5.2.5 are defined on the whole manifold M, $(t,s,x) \mapsto T(t,s,x)$ is C^∞ and the equations in Theorem 5.2.5 hold for every choice of $t, s \in [0,1]$ and $x \in M$,*

(iv) *there exists a compact set $K \subset M$, such that $T(t,s,x) = x$ for any $x \notin K$ and for any $t, x \in [0,1]$.*

Notice the importance of the condition $\int_0^1 \mathrm{L}(v_t)dt$ here. In the next chapter we will define so called *regular curves* based on this condition, which will allow us to come up with a definition for a parallel transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$.

Part III.

Second Order Analysis

6. Parallel Transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$

We turn to the second order analysis of $(\mathcal{P}_2(M), \mathcal{W}_2)$. At first, we are going to define a parallel transport for our setting. Afterwards we can define the covariant derivative.

6.1. Parallel Transport without Local Charts in $\mathbb{R}^n$

Parallel transport is usually defined using a connection. However, it is possible to define parallel transport in $\mathbb{R}^n$ without a connection and local charts. This is important, as we don't have this tools at our disposal in $(\mathcal{P}_2(M), \mathcal{W}_2)$. This section shows how to define parallel transport in $\mathbb{R}^n$ in such a way. We can then mimic this approach to define a suitable parallel transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$.

Throughout this section M will be a Riemannian manifold embedded in $\mathbb{R}^n$. We can identify the tangent space T_pM at a point $p \in M$ with a linear subspace $V \subset \mathbb{R}^n$.
Let us first consider a smooth curve $\gamma : [0,1] \to M$ and let us denote by $V_t := T_{\gamma(t)}M$ the tangent space at $\gamma(t)$. Furthermore we will denote by $P_t : \mathbb{R}^n \to V_t$ the orthogonal projections from $\mathbb{R}^n$ to V_t. If $u : [0,1] \to V_t$ is a regular vector field along γ, then the Levi-Civita derivative of u along γ is given by

$$\nabla_{\dot\gamma(t_0)} u(t_0) = P_{t_0}\left(\frac{du}{dt}(t_0)\right)$$

In this setting, the parallel transport of $u(0)$ along γ satisfies

$$P_t\left(\frac{du}{dt}(t)\right) = 0. \tag{6.1}$$

Let us first show uniqueness of the parallel transport. Note that the time derivative of the norm is 0, because

$$\frac{d}{dt}\|u(t)\|^2 = 2\left\langle u(t), \frac{d}{dt}u(t)\right\rangle = 2\left\langle u(t), P_t\left(\frac{d}{dt}u(t)\right)\right\rangle = 0.$$

That means the norm of a parallel transport is constant (w.r.t. the time t). Now assume there exist two parallel transports $u(t)$ and $\tilde{u}(t)$. Since $u - \tilde{u}$ also satisfies (6.1) (that means $u - \tilde{u}$ is also a parallel transport), we have $\|u(t) - \tilde{u}(t)\| = c$ for every t and for a suitable constant c. Setting $t = 0$ we get

$$c = \|u(0) - \tilde{u}(0)\| = \|u_0 - u_0\| = 0,$$

which means $u = \tilde{u}$.

The next step is to show existence of a solution of (6.1). To make use of such a proof in our Wasserstein setting, we may only use tools available also in $(\mathcal{P}_2(M), \mathcal{W}_2)$. Our key observations, which enable us to show the existence, are the following two estimates:

Lemma 6.1.1. *For any $s, t \in [0,1]$ and for arbitrary $v_t \in V_t$ and $v_t^\perp \in V_t^\perp$ the following two estimates hold:*

$$|v_t - P_s(v_t)| \leq C\, |v_t|\, |t - s|\,, \tag{6.2}$$

$$\left| P_s(v_t^\perp) \right| \leq C \left| v_t^\perp \right| |t - s|\,. \tag{6.3}$$

Proof. Without loss of generality assume $s < t$. For arbitrary $v \in \mathbb{R}^n$ define the map

$$\theta_v : [0,1]^2 \to \mathbb{R}^n : (s,t) \mapsto P_s(v) - P_t(v).$$

We begin with the simpler case, when there exists an open subset $U \subseteq M$ and an embedding $\phi : \tilde{U} \subseteq \mathbb{R}^k \to \phi(\tilde{U}) = U$, such that both $\gamma(t)$ and $\gamma(s)$ lie in U. Then we can express the tangent vectors at $\gamma(s)$ and $\gamma(t)$ as derivatives of ϕ through

$$\begin{aligned} v_s &= \sum v_s^i \frac{d\phi}{dx^i}(\gamma(s)), \quad v_s \in V_s, \\ v_t &= \sum v_t^i \frac{d\phi}{dx^i}(\gamma(t)), \quad v_t \in V_t. \end{aligned}$$

Using the basis vectors $f_i(s) := \frac{d\phi}{dx^i}(\gamma(s))$ we can express a projection onto V_s through

$$P_s(v) = \sum \frac{\langle v, f_i(s) \rangle}{|f_i(s)|^2} f_i(s)$$

and analogously for $P_t(v)$. Now $\theta_v(s,t)$ can be written as

$$\theta_v(s,t) = \sum \frac{\langle v, f_i(s) \rangle}{\langle f_i(s), f_i(s) \rangle} f_i(s) - \sum \frac{\langle v, f_i(t) \rangle}{\langle f_i(t), f_i(t) \rangle} f_i(t)$$

and lies therefore in $C^\infty\left([0,1]^2\right)$ (notice that every f_i is a composition of smooth functions and therefore smooth). As smooth function on a compact domain it is in particular Lipschitz-continuous. Furthermore, because $\theta_v(s,t)$ is linear in v, the Lipschitz constant is of the form $L = C\,|v|$, where C depends on γ but not on v. Choosing $v = v_t \in V_t$ we get

$$\begin{aligned} |v_t - P_s(v_t)| &= |(P_t(v_t) - P_t(v_t)) - (P_s(v_t) - P_t(v_t))| \\ &= |\theta_{v_t}(t,t) - \theta_{v_t}(s,t)| \\ &\leq C\,|v_t|\,|s - t|\,. \end{aligned}$$

Now consider the general case when there is no open subset U and embedding ϕ, such that U is a neighbourhood for $\gamma(s)$ and $\gamma(t)$. Because $\gamma([0,1])$, as smooth image of a compact set, is itself a compact set, we can find a finite number of times t_i, $i = 1, \dots, N$, open subsets $U_i \subseteq M$ and embeddings ϕ_i, $i = 0, \dots, N$, such that $t < t_1 < \dots < t_N < s$, $\gamma(t_i) \in U_{i-1} \cap U_i$ and such that $\gamma(s) \in U_0$ and $\gamma(t) \in U_N$. Using the triangle inequality we get

$$\begin{aligned} |v_t - P_s(v_t)| &= |P_t(v_t) - P_s(v_t)| \\ &\leq |P_t(v_t) - P_{t_1}(v_t)| + |P_{t_1}(v_t) - P_s(v_t)| \\ &\leq C\,|v_t|\,|t - t_1| + |P_{t_1}(v_t) - P_s(v_t)| \end{aligned} \tag{6.4}$$

where the last inequality is due to the first part ($\gamma(s)$ and $\gamma(t_1)$ both lie in U_0). In the same manner we iteratively proceed with the second summand on the righthand side of (6.4). We finally end up with

$$|v_t - P_s(v_t)| \leq C\,|v_t| \left(|t - t_1| + \sum_{i=1}^{N-1} |t_i - t_{i+1}| + |t_{i+1} - s| \right) = C\,|v_t|\,|s - t|\,.$$

As this inequality holds for every $v_t \in V_t$ with $|v_t| = 1$, we even get

$$\sup_{\substack{v_t \in V_t,\\ \|v_t\| = 1}} |v_t - P_s(v_t)| \leq C\,|s - t|\,. \tag{6.5}$$

To show the second estimate (6.3), we observe that for $\tilde{v}_t^\perp \in V_t^\perp$ with $\left|\tilde{v}_t^\perp\right| = 1$ we have

$$\begin{aligned} \left|P_s(\tilde{v}_t^\perp)\right| \leq \sup_{\substack{v_t^\perp \in V_t^\perp,\\ \left\|v_t^\perp\right\| = 1}} \left|v_t^\perp - P_s^\perp(v_t^\perp)\right| \quad &= \left\| P_s\big|_{V_t^\perp} \right\| \\ = \sup_{\substack{v_s \in V_s, v_t^\perp \in V_t^\perp,\\ \|v_s\| = \left\|v_t^\perp\right\| = 1}} \left\langle v_t^\perp, v_s \right\rangle \quad &= \sup_{\substack{v_s \in V_s, v_t^\perp \in V_t^\perp,\\ \|v_s\| = \left\|v_t^\perp\right\| = 1}} \left\langle v_s, v_t^\perp \right\rangle \\ = \left\| P_s^\perp\big|_{V_t} \right\| \quad &= \sup_{\substack{v_t \in V_t,\\ \|v_t\| = 1}} |v_t - P_s(v_t)| \\ \overset{(6.5)}{\leq} C\,|s - t|\,. \end{aligned}$$

For general $v_t^\perp \in V_t^\perp$ we therefore conclude

$$\left|P_s(v_t^\perp)\right| \leq C \left|v_t^\perp\right| |s - t|\,.$$

□

Lemma 6.1.1 is the most important observation in this section. It enables us to define the parallel transport without using tools relying on a Riemannian setting. Our aim for the following sections is to find an analogous result in the Wasserstein setting. After establishing this lemma (or its Wasserstein analogon), the remainder is mainly iterative application of the two estimates (6.2) and (6.3).

Remark 6.1.2. Lemma 6.1.1 has a very intuitive interpretation. We denote with $\theta(V_t, V_s) \in [0, \pi/2]$ the *angle between the subspaces* V_t and V_s, where

$$\sin\theta(V_t, V_s) := \sup_{\substack{v_t \in V_t, \\ \|v_t\|=1}} |v_t - P_s(v_t)| .$$

The above lemma gives us a Lipschitz condition for the map $(t, s) \mapsto \theta(V_t, V_s)$. It essentially tells us, that the angle between V_t and V_s varies smoothly in time. ★

Let us now return to our initial question: for an initial point u_0, how can we construct a vector field $u(t)$, such that $u(0) = u_0$ and (6.1) holds? The key idea is the following: for $\gamma(0) = u_0 \in V_0$ define the curve $u(t) := P_t(u_0)$, then (6.1) can be written as

$$\begin{aligned} P_0\left(\dot{u}(0)\right) =& P_0\left(\frac{d}{dt}P_t(u_0)\right) = P_0\left(\lim_{t\to 0}\frac{P_t(u_0) - P_0(u_0)}{t}\right) \\ =& \lim_{t\to 0}\frac{1}{t}P_0\left(P_t(u_0) - P_0(u_0)\right) = \lim_{t\to 0}\frac{1}{t}P_0\left(P_t(u_0) - u_0\right) \end{aligned} \tag{6.6}$$

As $P_t(u_0) - u_0 \in V_t^{\perp}$, we can use the estimates (6.2) and (6.3) to get

$$|P_0(P_t(u_0) - u_0)| \le C\,|P_t(u_0) - u_0|\,|t| \le C^2\,|u_0|\,|t|^2$$

and therefore (6.6) simplifies to

$$P_0\left(\frac{d}{dt}P_t(u_0)\right) = 0,$$

which already resembles (6.1) (for $t = 0$).
To finally compute the parallel transport, we have to consider limits on the set of partitions of $[0, 1]$ (or more generally partitions of $[s, t]\,,\ s, t \in [0, 1]$).

Definition 6.1.3 (Partition of $[s, t]$ and Set of Partitions)**.** A set $\mathcal{P} = \{t_0, t_1, \dots, t_N\}$ is called a partition of $[s, t]$ if $s = t_0 < t_1 < \cdots < t_N = t$. We will denote the set of all partitions on $[s, t]$ by $\mathfrak{P}_{s,t}$.
We say a partition $\mathcal{Q}$ is a *refinement* of $\mathcal{P}$, symbolically $\mathcal{Q} \ge \mathcal{P}$, if $\mathcal{P} \subseteq \mathcal{Q}$.

Note that $(\mathfrak{P}_{s,t}, \le)$ is a direct set. Next we recall what a limit over $\mathfrak{P}_{s,t}$ is.

Definition 6.1.4 (Limit over $\mathfrak{P}_{s,t}$)**.** Let (X, d) be a complete metric space and let $f : \mathfrak{P}_{s,t} \to X$ be a function. We say $x \in X$ is a limit of f over $\mathfrak{P}_{s,t}$, if

$$\forall \epsilon > 0 \quad \exists \mathcal{P} \in \mathfrak{P}_{s,t} : d(x, f(\mathcal{Q})) < \epsilon \quad \forall \mathcal{Q} \ge \mathcal{P}$$

We will denote this limit by $x = \lim_{\mathcal{P}\in\mathfrak{P}_{s,t}} f(\mathcal{P})$.

Remark 6.1.5. If the limit $\lim_{\mathcal{P}\in\mathfrak{P}_{s,t}} f(\mathcal{P})$ exists, it is unique. Furthermore, as (X, d) is complete, we could have equivalently required

$$\forall \epsilon > 0 \quad \exists \mathcal{P} \in \mathfrak{P}_{s,t} : d(f(\mathcal{P}), f(\mathcal{Q})) < \epsilon \quad \forall \mathcal{Q} \ge \mathcal{P}$$

in the definition of the limit. ★

In the following we will consecutively apply orthogonal projections to a starting point u_0. To simplify notation, we will therefore denote with $P_{s_k,s_{k-1},\ldots,s_1}(u_0)$ the function

$$P_{s_k,s_{k-1},\ldots,s_1}(u_0) := P_{s_k}(P_{s_{k-1}}(\ldots P_{s_1}(u_0)\ldots))$$

for arbitrary given times $s_1, s_2, \ldots, s_k$, not necessarily ordered.
For a partition $\mathcal{P} = \{t_0 < t_1 < \cdots < t_N\} \in \mathfrak{P}_{s,t}$ and $u_0 \in V_0$ we will denote by $\mathcal{P}(u_0)$ the function

$$\mathcal{P}(u_0) := P_{t_N}(P_{t_{N-1}}(\ldots P_{t_1}(u_0)\ldots)) = P_{t_N,t_{N-1},\ldots,t_1}(u_0) \in V_{t_N}.$$

We now show, that the limit $\lim_{\mathcal{P}\in\mathfrak{P}_{s,t}} \mathcal{P}(u_0)$ exists for every $u_0 \in V_0$. This limit will provide us with a suitable definition for the parallel transport in $\mathbb{R}^n$. The proof is mainly based on the following lemma.

Lemma 6.1.6. *Let $0 \le t_0 \le t_1 \le \ldots \le t_N \le 1$. Then for any $u \in V_{t_0}$ it holds*

$$\left|P_{t_N}(u) - P_{t_N,t_{N-1},\ldots,t_1}(u)\right| \le C^2 \,|u|\, |t_0 - t_N|^2 .$$

Proof. We prove the lemma by induction. We start with $N = 2$. For any $u \in V_{t_0}$ we have

$$\begin{aligned}
|P_{t_2}(u) - P_{t_2,t_1}(u)| &= |P_{t_2}(u - P_{t_1}(u))| \\
&\overset{(6.3)}{\le} C\,|u - P_{t_1}(u)|\,|t_2 - t_1| \\
&\overset{(6.2)}{\le} C^2\,|u|\,|t_2 - t_1|\,|t_1 - t_0| \\
&\le C^2\,|u|\,|t_2 - t_0|^2 .
\end{aligned} \tag{6.7}$$

Now assume we have shown the thesis already for $N-1$. Remember that we can estimate the length of a projection by

$$|P_t(u)| \le |u| . \tag{6.8}$$

Applying (6.7) and using the induction hypothesis we get

$$\begin{aligned}
\left|P_{t_N}(u) - P_{t_N,t_{N-1},\ldots,t_1}(u)\right| &\le \left|P_{t_N}(u) - P_{t_N,t_{N-1}}(u)\right| + \\
&\quad + \left|P_{t_N,t_{N-1}}(u) - P_{t_N,t_{N-1},\ldots,t_1}(u)\right| \\
&\overset{(6.8)}{\le} \left|P_{t_N}(u) - P_{t_N,t_{N-1}}(u)\right| + \\
&\quad + \left|P_{t_{N-1}}(u) - P_{t_{N-1},\ldots,t_1}(u)\right| \\
&\overset{(6.7)}{\le} C^2\,|u|\,|t_N - t_{N-1}|\,|t_{N-1} - t_0| + \\
&\quad + \left|P_{t_{N-1}}(u) - P_{t_{N-1},\ldots,t_1}(u)\right| \\
&\le C^2\,|u|\,|t_N - t_{N-1}|\,|t_{N-1} - t_0| + C^2\,|u|\,|t_{N-1} - t_0|^2 \\
&\le C^2\,|u|\,|t_0 - t_N|^2
\end{aligned}$$

which proves the lemma. □

We can now prove the existence of a limit of $\mathcal{P}(u)$ in $\mathfrak{P}_{s,t}$.

Lemma 6.1.7 (Existence of Limit of $\mathcal{P}(u)$)**.** *Let $[s,t] \subseteq [0,1]$ and $u_0 \in V_s$. Then the limit $\lim_{\mathcal{P}\in\mathfrak{P}_{s,t}} \mathcal{P}(u_0)$ exists.*

Proof. According to Remark 6.1.5 we want to show

$$\forall \epsilon > 0 \quad \exists \mathcal{P} \in \mathfrak{P}_{s,t} : |\mathcal{P}(u_0) - \mathcal{Q}(u_0)| < C_{u_0}\epsilon \quad \forall \mathcal{Q} \geq \mathcal{P}$$

with a constant C_{u_0} depending only on u_0. We are going to construct $\mathcal{P}$ and show that it satisfies (6.1).
For given $\epsilon > 0$ choose $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$ and define $\mathcal{P}$ by

$$\mathcal{P} := \{t_0, t_1, t_2, \dots, t_N\}, \quad \text{with } t_i = s + \frac{i(t-s)}{N}, \quad \forall i = 0, \dots, N.$$

Let $\mathcal{Q} \geq \mathcal{P}$ be a refinement of $\mathcal{P}$ with $\mathcal{Q} := \{s_0 < s_1 < \cdots < s_K\}$ and let $l \leq K$ be the index, such that $s_l = t_1$ (such an index exists because $\mathcal{P} \subseteq \mathcal{Q}$). Now define two partitions of $[t_1, t_N]$ through

$$\begin{aligned}
\mathcal{P}_1 &:= \mathcal{P} \cap [t_1, t_N] = \{t_1, t_2, \dots, t_N\}, \\
\mathcal{Q}_1 &:= \mathcal{Q} \cap [t_1, t_N] = \{s_l, s_{l+1}, \dots, s_K\}.
\end{aligned}$$

Furthermore, if we set

$$\begin{aligned}
v &:= P_{t_1}(u_0) \in V_{t_1}, \\
w &:= P_{s_l, s_{l-1}, \dots, s_1}(u_0) \in V_{s_l},
\end{aligned}$$

we have

$$\begin{aligned}
\mathcal{P}(u_0) &= P_{t_N, t_{N-1}, \dots, t_1}(u_0) = P_{t_N, t_{N-1}, \dots, t_2}\left(P_{t_1}(u_0)\right) = \mathcal{P}_1(v), \\
\mathcal{Q}(u_0) &= P_{s_K, s_{K-1}, \dots, s_1}(u_0) = P_{s_K, s_{K-1}, \dots, s_{l+1}}\left(P_{s_l, s_{l-1}, \dots, s_1}(u_0)\right) = \mathcal{Q}_1(w).
\end{aligned}$$

We therefore get

$$\begin{aligned}
|\mathcal{P}(u_0) - \mathcal{Q}(u_0)| &= |\mathcal{P}_1(v) - \mathcal{Q}_1(v) + \mathcal{Q}_1(v) - \mathcal{Q}_1(w)| \\
&\leq |\mathcal{P}_1(v) - \mathcal{Q}_1(v)| + |\mathcal{Q}_1(v) - \mathcal{Q}_1(w)| \\
&\leq |\mathcal{P}_1(v) - \mathcal{Q}_1(v)| + |v - w| \\
&= |\mathcal{P}_1(v) - \mathcal{Q}_1(v)| + \left|P_{s_l}(u_0) - P_{s_l, s_{l-1}, \dots, s_1}(u_0)\right| \\
&\leq |\mathcal{P}_1(v) - \mathcal{Q}_1(v)| + C^2 |u_0| |t_0 - t_1|^2, \qquad (6.9)
\end{aligned}$$

where the last step is due to Lemma 6.1.6. Setting $[s', t'] := [t_1, t]$ we can repeat the steps above for the first summand on the righthand side of (6.9). Proceeding inductively we finally get

$$
\begin{aligned}
|\mathcal{P}(u_0) - \mathcal{Q}(u_0)| &\le C^2 |u_0| \sum_{i=0}^{N-1} |t_i - t_{i+1}|^2 \\
&\le C^2 |u_0| \sum_{i=0}^{N-1} \frac{(t-s)^2}{N^2} \\
&= C^2 |u_0| \frac{(t-s)^2}{N} \\
&< C^2 |u_0| \epsilon,
\end{aligned}
$$

which is what we wanted to show. □

We are finally able to come up with a definition for the parallel transport in $\mathbb{R}^n$, using only tools also available in $(\mathcal{P}_2(M), \mathcal{W}_2)$:

Definition 6.1.8. Let $[s,t] \subseteq [0,1]$. For an initial tangent vector $u_s \in V_s$ we denote by $T_s^t(u_s)$ the limit

$$T_s^t(u_s) := \lim_{\mathcal{P} \in \mathfrak{P}_{s,t}} \mathcal{P}(u_s).$$

Lemma 6.1.9. *Let $0 \le t_1 \le t_2 \le t_3 \le 1$. Then $T_{t_2}^{t_3} \circ T_{t_1}^{t_2} = T_{t_1}^{t_3}$.*

Proof. Note that projections are continuous and that taking the limit of all partitions of $[t_1, t_3]$ is the same as taking the limit of all partitions of $[t_1, t_3]$ containing the point t_2. So we have

$$
\begin{aligned}
\lim_{\mathcal{P} \in \mathfrak{P}_{t_2,t_3}} \mathcal{P}\Big(\lim_{\mathcal{Q} \in \mathfrak{P}_{t_1,t_2}} \mathcal{Q}(u)\Big) &= \lim_{\mathcal{P} \in \mathfrak{P}_{t_2,t_3}} \lim_{\mathcal{Q} \in \mathfrak{P}_{t_1,t_2}} \mathcal{P}(\mathcal{Q}(u)) \\
&= \lim_{\substack{\mathcal{P} \in \mathfrak{P}_{t_1,t_3}, \\ t_2 \in \mathcal{P}}} \mathcal{P}(u) \\
&= \lim_{\mathcal{P} \in \mathfrak{P}_{t_1,t_3}} \mathcal{P}(u).
\end{aligned}
$$

□

Although we write only $T_s^t(u)$ without mentioning the smooth curve $\gamma(t)$, remember that T_s^t depends on γ through the projections P_t, which map onto $V_t = T_{\gamma(t)}M$. This means using a different curve $\tilde{\gamma}(t)$ will result in a different map $\tilde{T}_s^t$.
Now we can state the main result of this section:

Theorem 6.1.10 (Parallel Transport in $\mathbb{R}^n$)**.** *Let $[0,t] \subseteq [0,1]$ and $u_0 \in V_0$. Then the curve $t \mapsto T_0^t(u_0)$ is the parallel transport of u_0 along γ.*

Proof. First note that because $T_0^t(u_0)$ is a limit of projections, we have $\left|T_0^t(u_0)\right| \leq |u_0|$. Furthermore according to Lemma 6.1.6 we have

$$\begin{aligned}\left|P_t\left(T_0^{t-h}(u_0)\right) - T_0^t(u_0)\right| &= \left|P_t\left(T_0^{t-h}(u_0)\right) - T_{t-h}^t\left(T_0^{t-h}(u_0)\right)\right| \\ &= \left|P_t\left(u_{t-h}\right) - T_{t-h}^t\left(u_{t-h}\right)\right| \\ &\leq C^2\left|u_{t-h}\right| h^2 \\ &\leq C^2\left|u_0\right| h^2,\end{aligned}$$

where $u_{t-h} := T_0^{t-h}(u_0)$. Using this we get

$$\begin{aligned}P_t\left(\dot{T}_0^t(u_0)\right) &= P_t\left(\lim_{h\to 0}\frac{T_0^t(u_0) - T_0^{t-h}(u_0)}{h}\right) \\ &= \lim_{h\to 0}\frac{1}{h}P_t\left(T_0^t(u_0) - T_0^{t-h}(u_0)\right) \\ &= \lim_{h\to 0}\frac{1}{h}\left(T_0^t(u_0) - P_t\left(T_0^{t-h}(u_0)\right)\right) \\ &= 0,\end{aligned}$$

which proves our claim according to (6.1). □

6.2. Regular Curves

Our goal for the next three sections is to immitate all the steps of the last section to get to the same result as in Theorem 6.1.10 for our space of interest $(\mathcal{P}_2(M), \mathcal{W}_2)$ with a Riemannian manifold M. But before defining a parallel transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$, we have to come up with an equivalent of a smooth curve. We will call such curves *regular curves*. These curves will satisfy a similar lipschitz condition as those established in (6.2) and (6.3), which was a key condition for the construction of parallel transport in $\mathbb{R}^n$. We start this section by defining regular curves and investigate some of their properties.

Definition 6.2.1 (Regular Curve)**.** An absolutely continuous curve (μ_t) with velocity vector field (v_t), $t \in [0,1]$ is called a *regular curve* if it satisfies

(i) $\int \|v_t\|_{\mu_t}^2\, dt < \infty$,

(ii) $\int_0^1 \mathrm{L}(v_t)dt < \infty$.

Given a regual curve (μ_t) and its tangent velocity vector (v_t), the assumptions of Theorems 5.2.5 and 5.2.6 are met. In such a case, we will call the maps $\mathrm{T}(t,s,\cdot)$ the *flow maps* of μ_t. These flow maps allow us to define a translation along (μ_t).

Definition 6.2.2 (Translation Maps $(\tau_x)_s^t$ and τ_s^t)**.** Let (μ_t) be a regular curve, $\mathrm{T}(t,s,\cdot)$ its flow map and fix $t, s \in [0,1]$ and $x \in \operatorname{supp}(\mu_t)$. Denote with $\mathrm{P}_{s,t}$ the parallel transport in M along the absolutely continuous curve $r \mapsto \mathrm{T}(t,r,x)$ from $r = s$ to $r = t$. Then define the maps $(\tau_x)_s^t$ through

$$\begin{aligned}(\tau_x)_s^t : T_{\mathrm{T}(t,s,x)}M &\to T_xM \\ v &\mapsto \mathrm{P}_{s,t}(v).\end{aligned}$$

Let furthermore $u \in L^2_{\mu_s}$. The translation maps $\tau_s^t(u) \in L^2_{\mu_t}$ are then defined by

$$\tau_s^t(u)(x) = (\tau_x)_s^t\left(u\left(\mathrm{T}(t,s,x)\right)\right).$$

Thanks to the group properties of $\mathrm{T}(t,s,x)$ and of the parallel transport $\mathrm{P}_{s,t}(x)$, the translation maps τ_s^t satisfy the group property themselves. Another nice property, which follows directly from the definition of the translation maps, is that the translations τ_s^t are isometries.

Lemma 6.2.3. *Let $r, s, t \in [0,1]$. Then the following three statements hold:*

(i) the translation maps are linear,

(ii) the translation maps satisfy the group property

$$\tau_s^t \circ \tau_r^s = \tau_r^t,$$

(iii) τ_s^t maps $L^2_{\mu_s}$ isometrically to $L^2_{\mu_t}$.

Proof. The linearity follows directly from the linearity of the parallel transport $P_{s,t}(v)$. To proof (ii) fix an $u \in L^2_{\mu_r}$ and $x \in M$. First of all, since $\mathrm{T}(t,s,\mathrm{T}(s,r,x)) = \mathrm{T}(t,r,x)$, the composition $(\tau_s^t \circ \tau_r^s)(u)(x)$ is well defined. To show the claim, it is only a matter of expanding all definitions:

$$\begin{aligned}(\tau_s^t \circ \tau_r^s)(u)(x) &= (\tau_x)_s^t\left(\tau_r^s(u)\left(\mathrm{T}(t,s,x)\right)\right) \\ &= (\tau_x)_s^t\left((\tau_{\mathrm{T}(t,s,x)})_r^s\left(u\left(\mathrm{T}(s,r,\mathrm{T}(t,s,x))\right)\right)\right) \\ &= (\tau_x)_s^t\left((\tau_{\mathrm{T}(t,s,x)})_r^s\left(u\left(\mathrm{T}(t,r,x)\right)\right)\right) \\ &= (\tau_x)_s^t\left(\mathrm{P}_{r,s}\left(u\left(\mathrm{T}(t,r,x)\right)\right)\right) \\ &= \left(\mathrm{P}_{s,t} \circ \mathrm{P}_{r,s}\right)\left(u\left(\mathrm{T}(t,r,x)\right)\right) \\ &= \mathrm{P}_{r,t}\left(u\left(\mathrm{T}(t,r,x)\right)\right) \\ &= (\tau_x)_r^t\left(u\left(\mathrm{T}(t,r,x)\right)\right) \\ &= \tau_r^t(u)(x)\end{aligned}$$

To show the second part remember that parallel transports are norm-preserving. We therefore have

$$\begin{aligned}
\left\|\tau_s^t(v) - \tau_s^t(w)\right\|_{\mu_t} &= \int \left(((\tau_x)_s^t \circ v) - ((\tau_x)_s^t \circ w)\right)^2 \circ \mathrm{T}\,(t,s,\cdot)\, d\mu_t \\
&= \int \left(((\tau_x)_s^t \circ v) - ((\tau_x)_s^t \circ w)\right)^2 d\mu_s \\
&= \int ((\mathrm{P}_{s,t} \circ v) - (\mathrm{P}_{s,t} \circ w))^2 d\mu_s \\
&= \left\|\mathrm{P}_{s,t} \circ (v - w)\right\|_{\mu_s} \\
&= \left\|v - w\right\|_{\mu_s}.
\end{aligned}$$

□

These translation maps allow us to define a Lipschitz-like constant for the flow maps.

Definition 6.2.4 (Constant $\mathrm{L}(\mathrm{T}\,(t,s,\cdot))$)**.** Let (μ_t) be a regular curve, (v_t) its velocity vector field, $\mathrm{T}\,(t,s,\cdot)$ its flow maps and fix $s,t \in [0,1]$. We define the constant $\mathrm{L}(\mathrm{T}\,(t,s,\cdot))$ by

$$\mathrm{L}(\mathrm{T}\,(t,s,\cdot)) := \inf \liminf_{n\to\infty} \sup_{x\in\mathrm{supp}(\mu_t^n)} \left\|\nabla(\mathrm{T}^n\,(t,s,\cdot))(x) - (\tau_x)_s^t\right\|_{\mathrm{op}}.$$

Without proof we state the following proposition. The interested reader can find the proof in [Gig12].

Proposition 6.2.5 (Bounds for Lipschitz Constants)**.** *Let (μ_t) be a regular curve, (v_t) its velocity vector field and $\mathrm{T}\,(t,s,x)$ its flow maps. Then the following bounds for the Lipschitz constants hold:*

$$\begin{aligned}
\mathrm{L}_{\mathrm{Lip}}(\mathrm{T}\,(t,s,\cdot)) &\le e^{\left|\int_t^s \mathrm{L}(v_r)dr\right|}, \\
\mathrm{L}(\mathrm{T}\,(t,s,\cdot)) &\le e^{\left|\int_t^s \mathrm{L}(v_r)dr\right|} - 1.
\end{aligned}$$

We can now show the following important result, which enables us to define the parallel transport in the Wasserstein setting as we did in the special case of a manifold embedded in $\mathbb{R}^n$.

Proposition 6.2.6 (Bounds for Projections)**.** *Let (μ_t) be a regular curve, (v_t) its velocity vector field, $\mathrm{T}\,(t,s,x)$ its flow maps and $u \in \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M))$. Then the following bound for the error of the projections holds:*

$$\left\|\tau_s^t(u) - P_{\mu_t}(\tau_s^t(u))\right\|_{\mu_t} \le \mathrm{L}(\mathrm{T}\,(t,s,\cdot))\left\|u\right\|_{\mu_s}.$$

Proof. First of all, in order to show the statement, it is sufficient to show

$$\left\|\tau_s^t(\nabla\phi) - P_{\mu_t}(\tau_s^t(\nabla\phi))\right\|_{\mu_t} \le \mathrm{L}(\mathrm{T}\,(t,s,\cdot))\left\|\nabla\phi\right\|_{\mu_s} \quad \forall\phi \in C_c^\infty,$$

that is we search for a function $\psi \in C_c^\infty(M)$, such that the error $\left\|\tau_s^t(\nabla\phi) - \nabla\psi\right\|_{\mu_t}$ can be controlled by $\mathrm{L}(\mathrm{T}\,(t,s,\cdot))\,\|\nabla\phi\|_{\mu_s}$.
Remember that a projection leads to a best approximation and therefore

$$\left\|\tau_s^t(\nabla\phi) - P_{\mu_t}(\tau_s^t(\nabla\phi))\right\|_{\mu_t} \leq \left\|\tau_s^t(\nabla\phi) - \nabla\psi\right\|_{\mu_t} \quad \forall\nabla\psi \in \mathrm{Tan}_{\mu_t}\left(\mathcal{P}_2\left(M\right)\right).$$

Let us start with velocity vectors v_t that satisfy the regularity assumptions (iii) and (iv) in Theorem 5.2.5. Assuming that those assumptions are satisfied, we can set $\psi := \phi \circ \mathrm{T}\,(t,s,\cdot)$. Because ϕ and $\mathrm{T}\,(t,s,\cdot)$ belong to C^∞, so does ψ. And since according to assumption (iv) in Theorem 5.2.6 $\mathrm{T}\,(t,s,\cdot)$ differs from the identity only on a compact set, we even have $\psi \in C_c^\infty$. We therefore obtain

$$\begin{aligned}
\left\|\tau_s^t(\nabla\phi) - P_{\mu_t}(\tau_s^t(\nabla\phi))\right\|_{\mu_t} &\leq \left\|\tau_s^t(\nabla\phi) - \nabla\psi\right\|_{\mu_t} \\
&= \left\|\tau_s^t(\nabla\phi) - \Big((\nabla\phi\circ\mathrm{T}\,(t,s,\cdot))^\top \cdot \nabla\,\mathrm{T}\,(t,s,\cdot)\Big)^\top\right\|_{\mu_t} \\
&= \sqrt{\int \left|(\tau_x)_s^t(\nabla\phi\circ\mathrm{T}\,(s,t,x)) - \nabla\,\mathrm{T}\,(t,s,x)^\top \cdot \nabla\phi\circ\mathrm{T}\,(t,s,x)\right|^2 d\mu_t} \\
&= \sqrt{\int \left|\Big((\tau_x)_s^t - \nabla\,\mathrm{T}\,(t,s,x)^\top\Big)\left(\nabla\phi\circ\mathrm{T}\,(t,s,x)\right)\right|^2 d\mu_t} \\
&\leq \sqrt{\int \left\|(\tau_x)_s^t - \nabla\,\mathrm{T}\,(t,s,\cdot)^\top\right\|_{\mathrm{op}}^2 \left|\nabla\phi\circ\mathrm{T}\,(t,s,\cdot)\right|^2 d\mu_t} \\
&\leq \mathrm{L}(\mathrm{T}\,(t,s,\cdot))\,\|\nabla\phi\circ\mathrm{T}\,(t,s,\cdot)\|_{\mu_t} \\
&= \mathrm{L}(\mathrm{T}\,(t,s,\cdot))\,\|\nabla\phi\|_{\mu_s}
\end{aligned}$$

which is what we wanted to show.
For the general case (where (μ_t, v_t) does not satisfy the regularity assumptions from above) one can approximate the transport couple (μ_t, v_t) using Theorem 5.2.6 with transport couples satisfying those regularity assumptions. We won't dig deeper into this subject, the interested reader can find the proof in [Gig12]. □

6.3. Vector Fields along Regular Curves

Definition 6.3.1 (Vector Fields along a Curve)**.** Let (μ_t) be a curve in $\mathcal{P}_2(M)$ and let $u : [0,1] \to \mathfrak{X}(M)$ be a measurable map such that $u_t \in L^2_{\mu_t}$ for any t. Then we will call such a map a *vector field along* (μ_t) and denote it by (u_t).

We now turn our attention to the regularity of vector fields. If we only consider regular curves (μ_t), then we can translate u_t, which is defined in $L^2_{\mu_t}$, to $\tau_t^s(u_t) \in L^2_{\mu_s}$.

Definition 6.3.2 (Regularity of Vector Fields)**.** Let (μ_t) be a regular curve and (u_t) be a vector field along (μ_t). We say that (u_t) is absolutely continuous if the map $t \mapsto \tau_t^s(u_t)$ is absolutely continuous for any $s \in [0,1]$. In the same way we say (u_t) is C^n (or C^∞) if the map $t \mapsto \tau_t^s(u_t)$ is C^n (or C^∞ respectively) for any $s \in [0,1]$.

We are now able to define the total derivative along a regular curve.

Definition 6.3.3 (Total Derivative of an Absolutely Continuous Vector Field along a Regular Curve)**.** Let (μ_t) be a regular curve and (u_t) an absolutely continuous vector field along (μ_t). Then we define the *total derivative* of (u_t) as

$$\frac{\mathbf{d}}{dt}u_t = \lim_{s\to t}\frac{\tau_s^t(u_s)-u_t}{s-t}.$$

The above limit is intended to be in $L^2_{\mu_t}$.

The next proposition discusses three important properties: linearity, representation as translation of the time-derivative and the Leibniz rule.

Proposition 6.3.4. *Let (μ_t) be an absolutely curve and (u_t), $(\tilde{u}_t)$ two absolutely continuous vector fields along (μ_t). Then the following three properties hold:*

(i) the total derivative is linear: $\frac{\mathbf{d}}{dt}(u_t+\tilde{u}_t) = \frac{\mathbf{d}}{dt}u_t + \frac{\mathbf{d}}{dt}\tilde{u}_t$.

(ii) The total derivative can be represented by

$$\frac{\mathbf{d}}{dt}u_t = \tau_s^t\left(\frac{d}{dt}(\tau_t^s(u_t))\right) \quad \text{for a.e. } t\in[0,1]\,,\ \forall s\in[0,1]\,.$$

This in particular means, the total derivative of an absolutely continuous vector field is itself an L^1 vector field.

(iii) The total derivative satisfies the Leibniz rule:

$$\frac{d}{dt}\langle u_t,\tilde{u}_t\rangle_{\mu_t} = \left\langle \frac{\mathbf{d}}{dt}u_t,\tilde{u}_t\right\rangle_{\mu_t} + \left\langle u_t,\frac{\mathbf{d}}{dt}\tilde{u}_t\right\rangle_{\mu_t}.$$

Proof. The linearity of the total derivative follows immediately from the linearity of the translation maps τ_s^t.
The second property is due to the group property of the translations:

$$\tau_s^t\left(\frac{d}{dt}(\tau_t^s(u_t))\right) = \tau_s^t\left(\lim_{h\to 0}\frac{\tau_{t+h}^s(u_{t+h})-\tau_t^s(u_t)}{h}\right) = \lim_{h\to 0}\frac{\tau_{t+h}^t(u_{t+h})-u_t}{h} = \frac{\mathbf{d}}{dt}u_t.$$

Finally the Leibniz rule can be shown by

$$\begin{aligned}
\frac{d}{dt}\langle u_t,\tilde{u}_t\rangle_{\mu_t} &= \frac{d}{dt}\langle \tau_t^0(u_t),\tau_t^0(\tilde{u}_t)\rangle_{\mu_0}\\
&= \left\langle \frac{d}{dt}\tau_t^0(u_t),\tau_t^0(\tilde{u}_t)\right\rangle_{\mu_0} + \left\langle \tau_t^0(u_t),\frac{d}{dt}\tau_t^0(\tilde{u}_t)\right\rangle_{\mu_0}\\
&= \left\langle \frac{d}{dt}u_t,\tilde{u}_t\right\rangle_{\mu_0} + \left\langle u_t,\frac{d}{dt}\tilde{u}_t\right\rangle_{\mu_0}.
\end{aligned}$$

□

6.4. Parallel Transport in the Wasserstein Setting

We directly start with the definition of the parallel transport in the Wasserstein setting and show that this definition indeed satisfies the conditions for a parallel transport.

Definition 6.4.1 (Parallel Transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$)**.** Let (μ_t) be a regular curve, (v_t) its velocity vector field and (u_t) an absolutely continuous tangent vector field. Then we say (u_t) is a parallel transport if

$$P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right) = 0$$

for almost every $t \in [0,1]$.

First let us discuss uniqueness of parallel transports. As in the Riemannian case this can easily be shown by first showing that the norm of a parallel transport is constant. This follows from Proposition 6.3.4:

$$\frac{d}{dt}\left\|u_t\right\|_{\mu_t}^2 = 2\left\langle u_t, \frac{\mathbf{d}}{dt}u_t\right\rangle = 2\left\langle u_t, P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right)\right\rangle = 0.$$

Considering now two different parallel transports u_t and $\tilde{u}_t$, we have due to the linearity of the total derivative that $u_t - \tilde{u}_t$ is a parallel transport too, and therefore $\left\|u_t - \tilde{u}_t\right\|_{\mu_t} = c$ for a suitable constant c. Choosing $t = 0$ we get

$$c = \left\|u_0 - \tilde{u}_0\right\|_{\mu_0} = 0$$

which shows the uniqueness.
To show existence we proceed as in the Riemannian case. We start by stating the analogous result to Lemma 6.1.1. Using Proposition 6.2.5 and Proposition 6.2.6 the analogous estimates to (6.2) and (6.3) read

$$\left\|P_s^t(w)\right\|_{\mu_t} \leq C\left|\int_t^s \mathrm{L}(v_r)dr\right| \left\|w\right\|_{\mu_s} \quad t,s \in [0,1]\,, \quad w \in \mathrm{Tan}_{\mu_s}^{\perp}\left(\mathcal{P}_2\left(M\right)\right) \tag{6.10}$$

$$\left\|\tau_t^s(u) - P_t^s(u)\right\|_{\mu_s} \leq C\left|\int_t^s \mathrm{L}(v_r)dr\right| \left\|u\right\|_{\mu_t} \quad t,s \in [0,1]\,, \quad u \in \mathrm{Tan}_{\mu_t}\left(\mathcal{P}_2\left(M\right)\right). \tag{6.11}$$

Before we proceed let us simplify our notation a little bit. In the Riemannian case we worked with projections P_s, which could be applied to a tangent vector $v_t \in V_t$ to map it into V_s. In the Wasserstein setting however we cannot simply apply a projection to a tangent vector $v_t \in \mathrm{Tan}_{\mu_t}\left(\mathcal{P}_2\left(M\right)\right)$ to get a tangent vector in $\mathrm{Tan}_{\mu_s}\left(\mathcal{P}_2\left(M\right)\right)$, because v_t and v_s lie in different L^2 spaces. We have to first perform a translation to the correct $L^2_{\mu_s}$ space where the projection then is defined. Let us therefore denote such a translation and projection from $\mathrm{Tan}_{\mu_t}\left(\mathcal{P}_2\left(M\right)\right)$ to $\mathrm{Tan}_{\mu_s}\left(\mathcal{P}_2\left(M\right)\right)$ with

$$P_t^s(u) := P_{\mu_s}\left(\tau_t^s(u)\right).$$

Again we will apply consecutively such operations, therefore let us introduce a simpler notation. For arbitrary numbers $0 \leq s_1 \leq s_2 \leq \cdots \leq s_k \leq 1$ we denote by

$$P_{s_k,s_{k-1},\ldots,s_1}(u) := P^{s_k}_{s_{k-1}}(P^{s_{k-1}}_{s_{k-2}}(\ldots P^{s_2}_{s_1}(u)\ldots))$$

the map which repeatedly applies such translations and projections.
As in the case of $\mathbb{R}^n$ we will denote partitions of $[0,1]$ (or more generally of $[s,t]$) with $\mathcal{P}$ or $\mathcal{Q}$ and the set of all partitions will be denoted by $\mathfrak{P}$. Furthermore for $\mathcal{P} = \{t_0 < t_1 < \cdots < t_N\}$ and $u \in \mathrm{Tan}_{\mu_{t_0}}(\mathcal{P}_2(M))$ we denote by $\mathcal{P}(u)$ the map

$$\mathcal{P}(u) := P_{t_N,\ldots,t_0}(u).$$

In the proof of existence of a parallel transport we will need the following result:

Lemma 6.4.2. *Let $\mathfrak{P}$ be the set of all partitions of $[0,1]$ and for each partition $\mathcal{P} \in \mathfrak{P}$ denote with $t_0, \ldots, t_n$ its partitions points. Let furthermore (v_t) be the velocity vector field of a regular curve. Then*

$$\lim_{\mathcal{P}\in\mathfrak{P}} \sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{L}(v_r)dr \right)^2 = 0.$$

Proof.

$$\begin{aligned}\sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{L}(v_r)dr \right)^2 &\leq \max_{i=0,\ldots,n-1} \left\{ \int_{t_i}^{t_{i+1}} \mathrm{L}(v_r)dr \right\} \sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{L}(v_r)dr \right) \\ &\leq \max_{i=0,\ldots,n-1} \left\{ \int_{t_i}^{t_{i+1}} \mathrm{L}(v_r)dr \right\} \int_0^1 \mathrm{L}(v_r)dr\end{aligned}$$

which tends to 0 for $n \to \infty$. □

Now that we have established almost the same notation as in the Riemannian setting, the proofs can be easily adopted to our needs. Let us first state and prove the analogous of Lemma 6.1.6.

Lemma 6.4.3. *Let $0 \leq t_0 \leq t_1 \leq \ldots \leq t_N \leq 1$ be given numbers. Then for any $u \in \mathrm{Tan}_{\mu_{t_0}}(\mathcal{P}_2(M))$ it holds*

$$\left\| P^{t_N}_{t_0}(u) - P_{t_N,t_{N-1},\ldots,t_0}(u) \right\|_{\mu_{t_N}} \leq C^2 \|u\|_{\mu_{t_0}} \left(\int_{t_0}^{t_N} \mathrm{L}(v_r)dr \right)^2$$

Proof. The proof is almost the same as its Riemannian analogon. Again we proof the lemma with induction. Let us start with $N = 2$. First we see that because of the group property of the translations τ^s_t we have

$$\begin{aligned}P^{t_2}_{t_0}(u) - P^{t_2}_{t_1}\left(P^{t_1}_{t_0}(u)\right) &= P_{\mu_{t_2}}\left(\tau^{t_2}_{t_0}(u)\right) - P_{\mu_{t_2}}\left(\tau^{t_2}_{t_1}\left(P_{\mu_{t_1}}\left(\tau^{t_1}_{t_0}(u)\right)\right)\right) \\ &= P_{\mu_{t_2}}\left(\tau^{t_2}_{t_0}(u) - \tau^{t_2}_{t_1}\left(P^{t_1}_{t_0}\right)\right) \\ &= P_{\mu_s 2}\left(\tau^{t_2}_{t_1}\left(\tau^{t_1}_{t_0}(u) - P^{t_1}_{t_0}(u)\right)\right) \\ &= P^{t_2}_{t_1}\left(\tau^{t_1}_{t_0}(u) - P^{t_1}_{t_0}(u)\right).\end{aligned}$$

We have that $\tau_{t_0}^{t_1}(u)-P_{t_0}^{t_1}(u)\in \mathrm{Tan}^{\perp}_{\mu_{t_1}}(\mathcal{P}_2(M))$ and applying (6.10) and (6.11) we finally get

$$\begin{aligned}
\left\|P_{t_0}^{t_2}(u)-P_{t_2,t_1,t_0}(u)\right\|_{\mu_{t_2}} &= \left\|P_{t_0}^{t_2}(u)-P_{t_1}^{t_2}\left(P_{t_0}^{t_1}(u)\right)\right\|_{\mu_{t_2}}\\
&= \left\|P_{t_1}^{t_2}\left(\tau_{t_0}^{t_1}(u)-P_{t_0}^{t_1}(u)\right)\right\|_{\mu_{t_2}}\\
&\le C^2\left|\int_{t_1}^{t_2}\mathrm{L}(v_r)dr\right|\left\|\tau_{t_0}^{t_1}(u)-P_{t_0}^{t_1}(u)\right\|_{\mu_{t_1}}\\
&\le C^2\left|\int_{t_1}^{t_2}\mathrm{L}(v_r)dr\right|\left|\int_{t_0}^{t_1}\mathrm{L}(v_r)dr\right|\|u\|_{\mu_{t_0}} \qquad (6.12)\\
&\le C^2\left|\int_{t_0}^{t_2}\mathrm{L}(v_r)dr\right|^2\|u\|_{\mu_{t_0}}.
\end{aligned}$$

Now let us consider a general $N>2$ and assume we know that the thesis holds for $N-1$. Since due to Lemma 6.2.3 the translation maps τ_s^t are isometries, $\left\|P_s^t(u)\right\|_{\mu_t}$ can be estimated by

$$\left\|P_s^t(u)\right\|_{\mu_t}=\left\|P_{\mu_t}(\tau_s^t(u))\right\|_{\mu_t}\le\left\|\tau_s^t(u)\right\|_{\mu_t}=\|u\|_{\mu_s}. \qquad (6.13)$$

Then using (6.12) we have

$$\begin{aligned}
\left\|P_{t_0}^{t_N}(u)-P_{t_N,t_{N-1},\dots,t_0}(u)\right\|_{\mu_{t_N}} \le\ & \left\|P_{t_0}^{t_N}(u)-P_{t_N,t_{N-1},t_0}(u)\right\|_{\mu_{t_N}}\\
&+\left\|P_{t_N,t_{N-1},t_0}(u)-P_{t_N,t_{N-1},\dots,t_0}(u)\right\|_{\mu_{t_N}}\\
\overset{(6.13)}{\le}\ & \left\|P_{t_0}^{t_N}(u)-P_{t_N,t_{N-1},t_0}(u)\right\|_{\mu_{t_N}}\\
&+\left\|P_{t_{N-1},t_0}(u)-P_{t_{N-1},\dots,t_0}(u)\right\|_{\mu_{t_{N-1}}}\\
\le\ & \left\|P_{t_0}^{t_N}(u)-P_{t_N,t_{N-1},t_0}(u)\right\|_{\mu_{t_N}}+\\
&+C^2\left|\int_{t_0}^{t_{N-1}}\mathrm{L}(v_r)dr\right|^2\|u\|_{\mu_{t_0}}\\
\overset{(6.12)}{\le}\ & C^2\left|\int_{t_{N-1}}^{t_N}\mathrm{L}(v_r)dr\right|\left|\int_{t_0}^{t_{N-1}}\mathrm{L}(v_r)dr\right|\|u\|_{\mu_{t_0}}\\
&+C^2\left|\int_{t_0}^{t_{N-1}}\mathrm{L}(v_r)dr\right|^2\|u\|_{\mu_{t_0}}\\
\le\ & C^2\|u\|_{\mu_{t_0}}\left(\int_{t_0}^{t_N}\mathrm{L}(v_r)dr\right)^2.
\end{aligned}$$

□

To proof that $\mathcal{P}(u)$ has a limit in the set $\mathfrak{P}$ is now a consequence of the preceding lemma.

Theorem 6.4.4. *Let (μ_t) be a regular curve and consider a tangent vector $u_0 \in \operatorname{Tan}_{\mu_0}(\mathcal{P}_2(M))$. Then the limit $\lim_{\mathcal{P}\in\mathfrak{P}_{s,t}} \mathcal{P}(u_0)$ exists.*

Proof. Without loss of generality we may assume that $[s,t] = [0,1]$. Again, the proof is almost the same as in the euclidean case. The main idea is to repeatedly apply Lemma 6.4.3, such that, for given ϵ, there exists a partition $\mathcal{P}$ with

$$\|\mathcal{P}(u_0) - \mathcal{Q}(u_0)\|_{\mu_s} \leq C^2 \|u_0\|_{\mu_t} \sum_{i=0}^{N-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{L}(v_r) dr \right)^2 < \epsilon \tag{6.14}$$

for every partition $\mathcal{Q} \geq \mathcal{P}$. Fix therefore an arbitrary $\epsilon > 0$. According to Lemma 6.4.2 we can find a partition $\mathcal{P} = \{0 = t_0 < t_1 < \cdots < t_N = 1\}$ such that

$$C^2 \|u_0\|_{\mu_t} \sum_{i=0}^{N-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{L}(v_r) dr \right)^2 < \epsilon.$$

It remains to show the first inequality in (6.14) for every $\mathcal{Q} \geq \mathcal{P}$. Consider therefore a refinement $\mathcal{Q} = \{0 = s_0 < s_1 < \cdots < s_K = 1\}$ of $\mathcal{P}$ and let $l \leq K$ be the index, such that $s_l = t_1$ (such an index exists because $\mathcal{P} \subseteq \mathcal{Q}$). As in the Riemannian case, define two partitions of $[t_1, 1]$ and two new tangent vectors through

$$\begin{aligned}
\mathcal{P}_1 &:= \mathcal{P} \cap [t_1, 1] = \{t_1 < t_2 < \cdots < t_N = 1\}, \\
\mathcal{Q}_1 &:= \mathcal{Q} \cap [t_1, 1] = \{t_1 = s_l < s_{l+1} < \cdots < s_K = t_N = 1\}, \\
v &:= P_{t_0}^{t_1}(u_0) \in \operatorname{Tan}_{\mu_{t_1}}(\mathcal{P}_2(M)), \\
w &:= P_{s_l, s_{l-1}, \ldots, s_0}(u_0) \in \operatorname{Tan}_{\mu_{t_1}}(\mathcal{P}_2(M))
\end{aligned}$$

so that

$$\begin{aligned}
\mathcal{P}(u_0) &= P_{t_N,\ldots,t_0}(u_0) = P_{t_N,\ldots,t_1}\left(P_{t_0}^{t_1}(u_0)\right) = \mathcal{P}_1(v), \\
\mathcal{Q}(u_0) &= P_{s_K,\ldots,s_0}(u_0) = P_{s_K,\ldots,s_l}\left(P_{s_l,\ldots,s_0}(u_0)\right) = \mathcal{Q}_1(w).
\end{aligned}$$

We have

$$\begin{aligned}
\|\mathcal{P}(u_0) - \mathcal{Q}(u_0)\|_{\mu_1} &\leq \|\mathcal{P}_1(v) - \mathcal{Q}_1(v)\|_{\mu_1} + \|\mathcal{Q}_1(v) - \mathcal{Q}_1(w)\|_{\mu_1} \\
&\leq \|\mathcal{P}_1(v) - \mathcal{Q}_1(v)\|_{\mu_1} + \|v - w\|_{\mu_1} \\
&\leq \|\mathcal{P}_1(v) - \mathcal{Q}_1(v)\|_{\mu_t} + C^2 \|u_0\|_{\mu_t} \left(\int_{t_0}^{t_1} \mathrm{L}(v_r) dr \right)^2
\end{aligned} \tag{6.15}$$

where the last step is due to Lemma 6.4.3. We can repeat the steps above with the first summand on the right-hand side of (6.15), by defining $\mathcal{P}_2 = \mathcal{P}_1 \cap [t_2, 1]$ and $\mathcal{Q}_2 = \mathcal{Q}_1 \cap [t_2, 1]$ which gives us an estimate for $\|\mathcal{P}_1(v) - \mathcal{Q}_1(v)\|_{\mu_t}$. Repeatedly applying Lemma 6.4.3 to all subintervals $[t_i, 1]$ by constructing partitions $\mathcal{P}_m = \mathcal{P}_{m-1} \cap [t_m, 1]$ and $\mathcal{Q}_m = \mathcal{Q}_{m-1} \cap [t_m, 1]$ we finally end up with

$$\|\mathcal{P}(u_0) - \mathcal{Q}(u_0)\|_{\mu_s} \leq C^2 \|u_0\|_{\mu_t} \sum_{i=0}^{N-1} \left(\int_{t_i}^{t_{i+1}} \mathrm{L}(v_r) dr \right)^2 < \epsilon$$

which shows the existence of a limit. $\square$

Now, as in the euclidean case, we naturally get a definition for our parallel transportation map.

Definition 6.4.5 (Limit Process Map $\mathcal{T}_s^t(u_s)$)**.** Let $[s,t] \subseteq [0,1]$ and let (μ_t) be a regular curve. For an initial tangent vector $u_s \in \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M))$ we define the vector $\mathcal{T}_s^t(u_s)$ as the vector obtained by the limit process above, namely

$$\begin{aligned} \mathcal{T}_s^t : \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M)) &\to \mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M)) \\ u_s &\mapsto \lim_{\mathcal{P} \in \mathfrak{P}_{s,t}} \mathcal{P}(u_s). \end{aligned}$$

For $s > t$ the same definition holds, but instead of (μ_t) we consider the curve (μ_{1-t}).

Before showing that this map indeed gives us a parallel transport, we show that it satisfies the desired group property.

Proposition 6.4.6. *Let (μ_t) be a regular curve and let $\mathcal{T}_s^t$ be the limit process map as above. Then $\mathcal{T}_s^t$ satisfies the group property*

$$\mathcal{T}_s^t \circ \mathcal{T}_r^s = \mathcal{T}_r^t \quad \forall r,s,t \in [0,1].$$

Proof. We split the proof into two parts. First we only consider the easier case, where $r \leq s \leq t$. Then the proof is the same as in the Riemannian case, as the limit over all partitions $\mathcal{P}$ coincides with the limit over all partitions with a fixed partition point s. Now we turn to the general case. It is sufficient to show that $\mathcal{T}_s^t = (\mathcal{T}_t^s)^{-1}$, all possible orderings of r,s,t can then easily be derived. To that aim we will show that

$$\lim_{\mathcal{P} \in \mathfrak{P}} \|u - \mathcal{Q}_{\mathcal{P}}(\mathcal{P}(u))\|_{\mu_s} = 0 \quad \forall u \in \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M)) \tag{6.16}$$

where $\mathcal{Q}_{\mathcal{P}}$ is defined by

$$\begin{aligned} \mathcal{Q}_{\mathcal{P}} : \mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M)) &\to \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M)) \\ u &\mapsto P_{0,t_1,\dots,t_{n-1},1}(u) = P_{t_1}^0(P_{t_2}^{t_1}(\dots P_1^{t_{n-1}}(u)\dots)) \end{aligned}$$

for the partition $\mathcal{P} = \{0 < t_1 < \cdots < t_{n-1} < 1\}$. Since for any $u \in \mathrm{Tan}_{\mu_{t_i}}(\mathcal{P}_2(M))$ it holds

$$P_{t_{i+1}}^{t_i}\left(\tau_{t_i}^{t_{i+1}}(u)\right) = P_{\mu_{t_i}}\left(\tau_{t_{i+1}}^{t_i}\left(\tau_{t_i}^{t_{i+1}}(u)\right)\right) = P_{\mu_{t_i}}(u) = u$$

and since $P_{t_i}^{t_{i+1}}(u) - \tau_{t_i}^{t_{i+1}}(u) \in \mathrm{Tan}_{\mu_{t_{i+1}}}^{\perp}(\mathcal{P}_2(M))$, we can apply the lipschitz inequalities (6.10) and (6.11) to get

$$\begin{aligned} \left\|P_{t_{i+1}}^{t_i}(P_{t_i}^{t_{i+1}}(u)) - u\right\|_{\mu_{t_i}} &= \left\|P_{t_{i+1}}^{t_i}\left(P_{t_i}^{t_{i+1}}(u) - \tau_{t_i}^{t_{i+1}}(u)\right)\right\|_{\mu_{t_i}} \\ &\leq C\left\|P_{t_i}^{t_{i+1}}(u) - \tau_{t_i}^{t_{i+1}}(u)\right\|_{\mu_{t_{i+1}}} \left|\int_{t_i}^{t_{i+1}} L(v_r)dr\right| \\ &\leq C^2 \|u\|_{\mu_{t_i}} \left(\int_{t_i}^{t_{i+1}} L(v_r)dr\right)^2. \end{aligned}$$

Applying this estimate to (6.16) for arbitrary $u \in \mathrm{Tan}_{\mu_{t_0}}(\mathcal{P}_2(M))$ yields

$$\begin{aligned}
\|u - \mathcal{Q}_{\mathcal{P}}(\mathcal{P}(u))\|_{\mu_{t_0}} &\leq \left\|u - P_{t_1}^0\left(P_0^{t_1}(u)\right)\right\|_{\mu_{t_0}} + \left\|P_{t_1}^0\left(P_0^{t_1}(u)\right) - \mathcal{Q}_{\mathcal{P}}(\mathcal{P}(u))\right\|_{\mu_{t_0}} \\
&\leq C^2 \|u\|_{\mu_0} \left(\int_0^{t_1} L(v_r)dr\right)^2 + \\
&\quad + \left\|P_{t_1}^0\left(P_0^{t_1}(u)\right) - \mathcal{Q}_{\mathcal{P}}(\mathcal{P}(u))\right\|_{\mu_{t_0}} \\
&\leq C^2 \|u\|_{\mu_0} \left(\int_0^{t_1} L(v_r)dr\right)^2 + \\
&\quad + \left\|P_{t_1}^0\left(P_0^{t_1}(u) - P_{t_2}^{t_1}\left(\dots\left(P_1^{t_{n-1}}(\mathcal{P}(u))\right)\dots\right)\right)\right\|_{\mu_{t_0}} \\
&\leq C^2 \|u\|_{\mu_0} \left(\int_0^{t_1} L(v_r)dr\right)^2 + \\
&\quad + \left\|P_0^{t_1}(u) - P_{t_2}^{t_1}\left(\dots\left(P_1^{t_{n-1}}(\mathcal{P}(u))\right)\dots\right)\right\|_{\mu_{t_1}} \\
&= C^2 \|u\|_{\mu_0} \left(\int_0^{t_1} L(v_r)dr\right)^2 + \\
&\quad + \left\|\tilde{u} - P_{t_2}^{t_1}\left(\dots\left(P_1^{t_{n-1}}(\mathcal{P}'(\tilde{u}))\right)\dots\right)\right\|_{\mu_{t_1}}
\end{aligned}$$

where $\tilde{u} = P_0^{t_1}(u)$ and $\mathcal{P}'$ is a partition of $[t_1, 1]$ such that $\mathcal{P}'(\tilde{u}) = \mathcal{P}(u)$, i.e. $\mathcal{P}' = \{t_1 < t_2 < \cdots < t_{n-1} < 1\}$. Since we can estimate the norm of $\tilde{u}$ by $\|\tilde{u}\|_{\mu_{t_1}} \leq \|u\|_{\mu_0}$, we can iterate the above procedure and finally arrive at

$$\|u - \mathcal{Q}_{\mathcal{P}}(\mathcal{P}(u))\|_{\mu_{t_0}} \leq C^2 \|u\|_{\mu_0} \sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} L(v_r)dr\right)^2.$$

Then Lemma 6.4.2 proves our claim. □

Finally we are able to show that our construction indeed gives us the parallel transport in $(\mathcal{P}_2(M), \mathcal{W}_2)$.

Theorem 6.4.7. *Let (μ_t) be a regular curve, $u \in \mathrm{Tan}_{\mu_0}(\mathcal{P}_2(M))$ and let $\mathcal{T}_s^t$ be the limit process map. Then the vector field*

$$u_t := \mathcal{T}_0^t(u)$$

is the parallel transport of u along (μ_t).

Proof. The first thing we are going to show is the absolute continuity of $t \mapsto \mathcal{T}_0^t(u)$. Consider therefore any interval $[s,t] \subseteq [0,1]$. Applying Lemma 6.4.3 (with $t_0 = s, t_N = t$ and $t_i < t_{i+1}$ for arbitrary points t_i) and passing to the limit over all partitions $\{s < t_1 < \cdots < t_{N-1} < t\} \in \mathfrak{P}_{s,t}$, we get

$$\left\|P_s^t(u) - \mathcal{T}_s^t(u)\right\|_{\mu_t} \leq C^2 \|u\|_{\mu_s} \left(\int_s^t L(v_r)dr\right)^2 \tag{6.17}$$

for any $u \in \mathrm{Tan}_{\mu_s}(\mathcal{P}_2(M))$.
Now fix $s_1 < s_2 < t$. Combining above result with the lipschitz estimate (6.11), with the group property of $\mathcal{T}_s^t$ (Proposition 6.4.6) and with Lemma 6.2.3 (ii) and (iii) yields

$$\begin{aligned}
\left\|\tau_{s_1}^t\left(\mathcal{T}_0^{s_1}(u)\right) - \tau_{s_2}^t\left(\mathcal{T}_0^{s_2}(u)\right)\right\|_{\mu_t} &= \left\|\tau_{s_1}^{s_2}\left(\mathcal{T}_0^{s_1}(u)\right) - \mathcal{T}_0^{s_2}(u)\right\|_{\mu_{s_2}} \\
&= \left\|\tau_{s_1}^{s_2}(\tilde{u}) - \mathcal{T}_{s_1}^{s_2}(\tilde{u})\right\|_{\mu_{s_2}} \\
&\leq \left\|\tau_{s_1}^{s_2}(\tilde{u}) - P_{s_1}^{s_2}(\tilde{u})\right\|_{\mu_{s_2}} + \left\|P_{s_1}^{s_2}(\tilde{u}) - \mathcal{T}_{s_1}^{s_2}(\tilde{u})\right\|_{\mu_{s_2}} \\
&\leq C\left(1 + C\int_0^1 L(v_r)dr\right)\|\tilde{u}\|_{\mu_{s_2}} \int_{s_1}^{s_2} L(v_r)dr
\end{aligned}$$

which shows the absolute continuity.
Now let us show that $t \mapsto \mathcal{T}_0^t(u)$ defines the parallel transport. We want to show that

$$\left\|P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right)\right\|_{\mu_t} = 0$$

or equivalently

$$\lim_{s\to t}\left\|\frac{P_{\mu_t}\left(\tau_s^t(u_s) - u_t\right)}{s-t}\right\|_{\mu_t} = 0.$$

To that aim observe that

$$\begin{aligned}
\left\|P_{\mu_t}\left(\tau_s^t(u_s) - u_t\right)\right\|_{\mu_t} &\leq \left\|P_{\mu_t}\left(\tau_s^t(u_s) - u_t\right) - P_{\mu_t}\left(\tau_s^t\left(P_t^s(u)\right) - u_t\right)\right\|_{\mu_t} + \\
&\quad + \left\|P_{\mu_t}\left(\tau_s^t\left(P_t^s(u)\right) - u_t\right)\right\|_{\mu_t} \\
&\leq \left\|\tau_s^t\left(u_s - P_t^s(u_t)\right)\right\|_{\mu_t} + \left\|P_{\mu_t}\left(\tau_s^t\left(P_t^s(u)\right) - u_t\right)\right\|_{\mu_t} \\
&= \left\|u_s - P_t^s(u_t)\right\|_{\mu_s} + \left\|P_{\mu_t}\left(\tau_s^t\left(P_t^s(u)\right) - u_t\right)\right\|_{\mu_t}.
\end{aligned}$$

According to (6.17), $\|u_s - P_t^s(u_t)\|_{\mu_s}$ is already a $o(s-t)$. To conclude the proof we need to show that $\left\|P_{\mu_t}\left(\tau_s^t\left(P_t^s(u)\right) - u_t\right)\right\|_{\mu_t}$ is $o(s-t)$ too. But since $P_t^s(u_t) - \tau_t^s(u_t) \in \mathrm{Tan}_{\mu_s}^{\perp}(\mathcal{P}_2(M))$, this follows by applying the lipschitz estimates (6.10) and (6.11):

$$\begin{aligned}
\left\|P_{\mu_t}\left(\tau_s^t\left(P_t^s(u)\right) - u_t\right)\right\|_{\mu_t} &= \left\|P_s^t\left(P_t^s(u_t) - \tau_t^s(u_t)\right)\right\|_{\mu_t} \\
&\leq C\left\|P_t^s(u_t) - \tau_t^s(u_t)\right\|_{\mu_s}\int_s^t L(v_r)dr \\
&\leq C^2\|u_t\|_{\mu_t}\left(\int_s^t L(v_r)dr\right)^2.
\end{aligned}$$

□

7. Covariant Derivative in $(\mathcal{P}_2(M), \mathcal{W}_2)$

In this chapter we are going to define the Levi-Civita connection on $(\mathcal{P}_2(M), \mathcal{W}_2)$ through the parallel transport map $\mathcal{T}_s^t$ developed in the last chapter.

7.1. Levi-Civita Connection

Let us directly start with the definition of the covariant derivative.

Definition 7.1.1 (Covariant Derivative)**.** Let (μ_t) be a regular curve and $\mathcal{T}_s^t$ the parallel transport maps along (μ_t). If furthermore (u_t) is an absolutely continuous tangent vector field along (μ_t), i.e. $u_t \in \mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M))\ \forall t \in [0,1]$, then the *covariant derivative* $\frac{\mathbf{D}}{dt}u_t$ of u_t along (μ_t) is defined by

$$\frac{\mathbf{D}}{dt}u_t := \lim_{h\to 0} \frac{\mathcal{T}_{t+h}^t(u_{t+h}) - u_t}{h},$$

where the limit is intended in $\mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M))$.

Remark 7.1.2. Because

$$P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right) = \lim_{h\to 0} \frac{P_{t+h}^t(u_{t+h}) - u_t}{h}$$

we can use the estimate (6.17) to obtain

$$\begin{aligned}
\left\| \frac{\mathbf{D}}{dt}u_t - P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right)\right\|_{\mu_t} &= \lim_{h\to 0}\frac{1}{h}\left\|\mathcal{T}_{t+h}^t(u_{t+h}) - P_{t+h}^t(u_{t+h})\right\|_{\mu_t} \\
&\le \lim_{h\to 0}\frac{1}{h}C^2\left\|u_{t+h}\right\|_{\mu_{t+h}}\left(\int_{t+h}^t L(v_r)dr\right)^2 \\
&= 0.
\end{aligned}$$

Therefore, we could equivalently define the covariant derivative by

$$\frac{\mathbf{D}}{dt}u_t := P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right).$$

From this characterization it immediately follows, that $\frac{\mathbf{D}}{dt}u_t$ is an L^1 vector field, since

$$\left\|\frac{\mathbf{D}}{dt}u_t\right\|_{\mu_t} \le \left\|\frac{\mathbf{d}}{dt}u_t\right\|_{\mu_t}.$$

★

The next step is to show that this covariant derivative is indeed the Levi-Civita connection on $(\mathcal{P}_2(M), \mathcal{W}_2)$, i.e. we need to show its compatibility with the metric and that it satisfies the torsion free identity.

Lemma 7.1.3 (Compatibility with the Metric)**.** *Given two absolutely continuous vector fields $u_t, \tilde{u}_t \in \mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M))$, the covariant derivative $\frac{\boldsymbol{D}}{dt}$ satisfies*

$$\frac{d}{dt}\langle u_t, \tilde{u}_t\rangle_{\mu_t} = \left\langle \frac{\boldsymbol{D}}{dt}u_t, \tilde{u}_t \right\rangle_{\mu_t} + \left\langle u_t, \frac{\boldsymbol{D}}{dt}\tilde{u}_t \right\rangle_{\mu_t}.$$

Proof. Using the Leibniz rule from Proposition 6.3.4 and Remark 7.1.2 we obtain

$$\begin{aligned}\frac{d}{dt}\langle u_t, \tilde{u}_t\rangle_{\mu_t} &= \left\langle \frac{\mathbf{d}}{dt}u_t, \tilde{u}_t \right\rangle_{\mu_t} + \left\langle u_t, \frac{\mathbf{d}}{dt}\tilde{u}_t \right\rangle_{\mu_t} \\ &= \left\langle P_{\mu_t}\left(\frac{\mathbf{d}}{dt}u_t\right), \tilde{u}_t \right\rangle_{\mu_t} + \left\langle u_t, P_{\mu_t}\left(\frac{\mathbf{d}}{dt}\tilde{u}_t\right) \right\rangle_{\mu_t} \\ &= \left\langle \frac{\mathbf{D}}{dt}u_t, \tilde{u}_t \right\rangle_{\mu_t} + \left\langle u_t, \frac{\mathbf{D}}{dt}\tilde{u}_t \right\rangle_{\mu_t}.\end{aligned}$$

□

To state and prove the torsion free identity, we need a little bit more notation. Consider therefore a fixed starting point $\mu \in \mathcal{P}_2(M)$ and two regular curves μ_t^1, μ_t^2 starting in μ, that is $\mu_0^1 = \mu_0^2 = \mu$, with its corresponding velocity vector fields v_t^1 and v_t^2. Note that those velocity vector fields can be chosen continuous according to Lemma 4.1.1. Furthermore, consider two C^1 vector fields, u_t^1 along μ_t^1 and u_t^2 along μ_t^2, which satisfy $u_0^2 = v_0^1$ and $u_0^1 = v_0^2$. Now it makes sense to consider the derivatives $\frac{\mathbf{D}}{dt}u_t^1$ of u_t^1 along μ_t^1 and $\frac{\mathbf{D}}{dt}u_t^2$ of u_t^2 along μ_t^2 at $t = 0$. Let us introduce a new notation for those covariant derivatives:

$$\nabla_{u_0^2}u_t^1 := \frac{\mathbf{D}}{dt}u_t^1\Big|_{t=0}$$

and analogously

$$\nabla_{u_0^1}u_t^2 := \frac{\mathbf{D}}{dt}u_t^2\Big|_{t=0}.$$

Lemma 7.1.4 (Torsion Free Identity)**.** *Given two absolutely continuous vector fields $u_t^1, u_t^2 \in \mathrm{Tan}_{\mu_t}(\mathcal{P}_2(M))$ such as above, the* torsion free identity *holds, i.e.*

$$\left[u_0^1, u_0^2\right] = \nabla_{u_0^1}u_t^2 - \nabla_{u_0^2}u_t^1.$$

Proof. First consider the functional $F_\phi : \mu \mapsto \int \phi d\mu$ for arbitrary but fixed $\phi \in C_c^\infty(M)$. Due to the continuity equation (4.2) we obtain the derivative of F_ϕ along u_t^2 as $\left\langle \nabla\phi, u_t^2\right\rangle_{\mu_t}$. Combining this with the Leibniz rule from Proposition 6.3.4 we get

$$\begin{aligned}u_0^1(u_0^2(F_\phi(\mu))) &= \frac{d}{dt}\left\langle \nabla\phi, u_t^2\right\rangle_{\mu_t}\Big|_{t=0} \\ &= \left\langle \nabla_{u_0^1}\nabla\phi, u_0^2\right\rangle_{\mu} + \left\langle \nabla\phi, \nabla_{u_0^1}u_t^2\right\rangle_{\mu}.\end{aligned}$$

Analogously, we can compute $u_0^2(u_0^1(F_\phi(\mu)))$. Taking the difference we arrive at

$$\begin{aligned}\left[u_0^1, u_0^2\right](F_\phi(\mu)) &= u_0^1(u_0^2(F_\phi(\mu))) - u_0^2(u_0^1(F_\phi(\mu))) \\ &= \left\langle \nabla\phi, \nabla_{u_0^1} u_t^2 - \nabla_{u_0^2} u_t^1 \right\rangle_\mu + \left\langle \nabla_{u_0^1}\nabla\phi, u_0^2 \right\rangle_\mu - \left\langle \nabla_{u_0^2}\nabla\phi, u_0^1 \right\rangle_\mu \\ &= \left\langle \nabla\phi, \nabla_{u_0^1} u_t^2 - \nabla_{u_0^2} u_t^1 \right\rangle_\mu,\end{aligned}$$

where the last equality is due to

$$\left\langle \nabla_{u_0^1}\nabla\phi, u_0^2 \right\rangle_\mu = \left\langle \nabla_{u_0^2}\nabla\phi, u_0^1 \right\rangle_\mu.$$

Since the set $\{\nabla\phi \mid \phi \in C_c^\infty(M)\}$ is by definition dense in $\mathrm{Tan}_\mu(\mathcal{P}_2(M))$, we finally get

$$\left[u_0^1, u_0^2\right](\mu) = \nabla_{u_0^1} u_t^2 - \nabla_{u_0^2} u_t^1$$

which is exactly the torsion free identity. □

Bibliography

[AG$^+$08] Luigi Ambrosio, Nicola Gigli, et al. Construction of the parallel transport in the wasserstein space. *Methods Appl. Anal*, 15(1):1–30, 2008.

[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 in metric spaces and in the space of probability measures.* Birkhäuser, 2008.

[BGMS09] Mathias Beiglböck, Martin Goldstern, Gabriel Maresch, and Walter Schachermayer. Optimal and better transport plans. *Journal of Functional Analysis*, 256(6):1907–1927, 2009.

[dCV92] Manfredo Perdigao do Carmo Valero. *Riemannian geometry.* Birkhäuser, 1992.

[GHL04] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry.* Springer, 2004.

[Gig12] Nicola Gigli. *Second order analysis on* $(\mathcal{P}_2(M), \mathcal{W}_2)$, volume 216. American mathematical society, 2012.

[Lee01] John M Lee. *Introduction to Smooth manifolds.* Springer Verlag, New York, 2001.

[Lot08] John Lott. Some geometric calculations on wasserstein space. *Communications in Mathematical Physics*, 277(2):423–437, 2008.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26:101–174, 2001.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.