

Shape space in terms of Wasserstein geometry and
applications to quantum physics

Dissertation
zur Erlangung des Doktorgrades
“Doctor rerum naturalium”
der Georg-August-Universität Göttingen

im Promotionsstudiengang “Mathematical Sciences”
der Georg-August University School of Science (GAUSS)

vorgelegt von
Bernadette Lessel
aus Koblenz

Göttingen, 2018

Betreuungsausschuss:

Prof. Dr. Thomas Schick, Mathematisches Institut,
Georg-August-Universität Göttingen

Prof. Dr. Dorothea Bahns, Mathematisches Institut,
Georg-August-Universität Göttingen

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Thomas Schick, Mathematisches Institut,
Georg-August-Universität Göttingen

Korreferentin: Prof. Dr. Dorothea Bahns, Mathematisches Institut,
Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Ingo Witt, Mathematisches Institut,
Georg-August-Universität Göttingen

Jun. Prof. Dr. Madeleine Jotz Lean, Mathematisches Institut,
Georg-August-Universität Göttingen

Prof. Dr. Anja Sturm, Institut für Mathematische Stochastik,
Georg-August-Universität Göttingen

Prof. Dr. Karl-Henning Rehren, Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 28.06.2018

For me.

Acknowledgements

Foremost, I want to express my sincerest gratitude and respect to my principal advisor Prof. Dr. Thomas Schick for his guidance throughout the years and his mathematical insights. But also for his patience, benevolence and for giving me the freedom I needed. I am also much obliged to my second advisor Prof. Dr. Dorothea Bahns and to the Mathematical Institute of the Georg-August University Göttingen for accepting me as a student and for having me for this long time.

This work would certainly not have been possible without the help and companionship of my friends, who I love dearly and who am very proud of. I am in particular grateful to Mareike Kremling, Leonhard Horstmeyer and Ryszard Kostecki for inspiring me and for helping me grow as a person. Scientifically, I am additionally thankful to Leonhard Horstmeyer, Lukas Katthän, Markus Penz and Christian Röken for many helpful and interesting discussions.

I also wish to express my gratitude and appreciation to all those past and present mathematicians on whose research my results rely on. Without their work, I would have never been able to pursue this research program of mine.

Financially, this thesis has been made possible by the generous support of the Research training group "Mathematical structures in modern quantum physics" of the Georg-August University Göttingen, by the Perimeter Institute of Theoretical Physics in Waterloo, Canada and by the Andrea von Braun Stiftung. Thank you very much.

Finally, I am supremely grateful for the love and support of my parents, Ursula Stäger-Lessel and Wolfgang Lessel, who I am very proud of.

Contents

1	Introduction	6
1.1	Motivation and overview	6
1.2	Structure of the text	8
2	Wasserstein geometry	10
2.1	Optimal transport	10
2.2	Wasserstein spaces $W_p(X)$	16
2.3	Wasserstein geodesics and displacement interpolation	18
2.4	Dynamical equations on $W_2(M)$	20
2.5	Induced differentiable structure on $W_2(M)$	26
3	Differentiable maps between Wasserstein spaces	33
3.1	Absolutely continuous maps	33
3.2	About the image of dF_μ	37
3.3	Differentiable maps between Wasserstein spaces	42
4	Quantum dynamics on $W_2(M)$	49
4.1	Elements of quantum dynamics	49
4.2	Optimal transport for a solution of the free Schrödinger equation	54
4.3	Madelung equations in terms of Wasserstein geometry	62
5	Philosophical interlude	64
6	Shape space	69
6.1	Shape distances $\mathcal{S}_p(X)$	69
6.2	Existence of a minimizer for the Shape distance	73
6.3	The Shape space on Riemannian manifolds	76
6.4	Fisher information on $\mathcal{S}_p(X)$	78
6.5	Some properties of Shape space	78
6.6	Geodesic structure of Shape space	82
7	Infinitesimal dynamics on Shape space	86
7.1	Isometric actions on $TW(M)$	86
7.2	Tangent space on $\mathcal{S}(M)$	89
7.3	Quantum dynamics on $\mathcal{S}(\mathbb{R}^n)$	94

A	Additional information	99
A.1	Useful concepts	99
A.1.1	Topological spaces	99
A.1.2	Proper maps	101
A.1.3	Disintegration theorem	101
A.2	Remarks on differentiable structures on probability measures	102
A.2.1	Otto's approach	102
A.2.2	John Lott's geometric calculations on Wasserstein space .	103
A.2.3	Geometric tangent space	103

Chapter 1

Introduction

1.1 Motivation and overview

The core idea of this thesis and its main motivation is a philosophical thought concerning the description of nature. This idea, inspired by Mach's principle, says that it should be possible to describe the dynamics of matter without resorting to a notion of space, in particular to a notion of an absolute spatial background. The conception is rather that matter constitutes space, so that dynamics should solely be concerned with the change of the state of matter itself. The expression of the change of matter with respect to a spatial structure is, on the other hand, considered to be an effective description. Since we believe that matter should fundamentally be described by quantum theory, the aim of this text is to support and quantify mathematically this philosophical idea for quantum theory.¹

The mathematics we will use for this inquiry is mainly Wasserstein geometry. This is an infinite dimensional, formal, Riemannian structure on the space of probability measures, originating in the study of optimal transport. The latter is concerned with the question, how to transport mass densities most cost efficiently onto each other. Is the cost taken with respect to the metric distance of the underlying space, the total cost for the transportation defines a metric distance on the space of probability measures on that space, the so called Wasserstein distance. Investigating the geodesic structure of that metric (in the sense of shortest paths between points), yields a rich geometric structure on the space of probability measures, which eventually leads to a meaningful definition of a formal Riemannian structure on it. We say *formal*, because this structure is not defined by local coordinates as we know it from ordinary manifolds, but by carefully defining objects in such a way that corresponding formulas resemble formulas known from finite dimensional Riemannian geometry. The connection to quantum physics is then established by the fact that the squared modulus of a

¹Let us point out that in this text, we only treat non-relativistic, one-particle quantum mechanics.

solution of the Schrödinger equation is considered as a probability distribution. If the initial conditions are appropriately chosen, such a curve of probability measures stemming from a solution of the Schrödinger equation can naturally be regarded as a curve in Wasserstein space. In the course of this text we will furthermore see that such curves behave in particular very nicely and natural with respect to optimal transport.

We will, now, use Wasserstein geometry as a formal language to express our philosophical idea in the following way. Instead of considering the squared modulus of a solution of the Schrödinger equation as an evolution of densities on space, we just want to concern ourselves with how the *shape* of the density changes in time. To do so, we will *mod out* all the places in space the density can be brought to, to be solely left with the relative configuration of the density itself. Mathematically, we achieve this by looking at the quotient space Wasserstein space modulo the isometry group of the underlying metric space. A shape, thus, will be an equivalence class of probability measures, where two measures are equivalent to each other, in case the one is the image measure of the other with respect to an isometry. We will call this quotient the *Shape space*. Physically, this construction is only relevant for the Wasserstein space modeled on \mathbb{R}^3 , however, we will also be interested in investigating the most general case where the Wasserstein space is modeled on a complete metric space.

The Wasserstein distance induces a pseudo-distance on Shape space and we will find natural conditions with which moreover a proper metric distance is obtained, which we will call *Shape distance*. We will investigate this metric structure regarding its topological, metric and geodesic properties. Regarding our physical application, we will in particular find that probability measures corresponding to solutions of the Schrödinger equation can very naturally be regarded as curves in Shape space and we will even find a solution that, after projection, constitutes a geodesic in Shape space. We want to take this result as the main support of our philosophical idea we started with. It shows that once the passage from the wave function to probability measures has been made, no information is lost when considering only the change of the shapes of the densities. Even more, the evolution of the shapes is very natural and economic with respect to the Shape distance.

To also be able to speak about infinitesimal change of shapes, we will furthermore propose a definition for a tangent space at a point in Shape space. Here, we take on the one hand inspiration from the finite dimensional case and on the other hand resort for this to the already existing definition of the tangent space on the Wasserstein space. We will find that also in our scenario we obtain a meaningful and well-defined space if we mod out those directions in the tangent space on the Wasserstein space that show towards the orbit of the isometry group. We will see that the gradient of the wave function can naturally be regarded as being tangent in Shape space to the curve of shapes originating from that same wave function. In this sense it is thus possible to give an intrinsic formulation of the motion of the probability density of a particle in terms of shapes.

Finally, the considerations regarding the definition of the tangent space on

Shape space lead to the question how a notion of differentiability for maps between Wasserstein spaces could be established. For to show that our definition of the tangent space on Shape space is well-defined, we need to compare the tangent space on Wasserstein space at all the points which are in the same orbit of the isometry group. Since, as mentioned above, the differentiable structure of the Wasserstein space is not induced by local coordinates, a notion of differentiability for maps between Wasserstein spaces is not naturally given. To our knowledge, this problem has not been studied before in the literature. Our proposal for such a notion is highly tailored to the specific differentiable structure on the Wasserstein space and knowingly differs from notions of differentiability in infinite dimensions, as for example treated in [KM97].

1.2 Structure of the text

As mentioned above, the thesis has three main themes: a basic philosophical thought, Wasserstein geometry and quantum physics. These components are mostly treated each in their own section, only to finally be conducted in the last section, 7.3.

We begin our work with an introduction to Wasserstein geometry in Chapter 2, to supply the reader with the relevant mathematical background. Most of the material included here is considered standard among experts, but we also already add some own technical statements which we will need later on.

In Chapter 3 we continue with developing the notion of differentiable maps between Wasserstein spaces that we have mentioned above. We will first need to introduce and study *absolutely continuous maps* between Wasserstein spaces, to be able to make precise what we should demand from a differentiable map. We will see that our definition is able to fulfill what one expects from a differential. In particular, we will show that the identity map is differentiable with the expected differential and so is the constant map and the composition of differentiable maps. Furthermore, the differential is unique up to a *negligible* set. Also, for a more non-trivial example of mappings between Wasserstein spaces we will explicitly calculate the differential.

Chapter 4 begins with a briefing of the quintessence of quantum theory. As our main aim is a philosophical one, we did not just want to start with the Schrödinger equation, but instead want try to provide a feeling for the theory. After this introduction, we examine a special solution of the Schrödinger equation with some tools of optimal transport. We will find that this solution behaves very nicely with respect to optimal transport. Amongst other things, we will see that it is a geodesic in Wasserstein space and that the gradient of the corresponding wave function is always an element of the tangent space along that curve. We close this chapter by mentioning the work done by M.-K. von Renesse in [vR12], which shows the naturalness of the Wasserstein formalism

for the description of quantum dynamics.

We take a break in Chapter 5 to read more about the philosophical idea we have outlined above, to be prepared for the last two chapters.

The 6th chapter properly introduces the notion of shapes we have motivated in the beginning. After introducing the Shape space, the chapter is concerned with investigating appropriate conditions on the underlying metric space and on the action of the isometry group for the Shape distance to be an actual metric distance. We will show that this is the case whenever the underlying space and the action of the isometry group are *proper*. We will find that luckily, on complete, connected Riemannian manifolds this is always the case, whenever the isometry group is equipped with the compact-open topology. Next to showing that the Fisher information, which plays an important role in the exposition of Renesse, is well-defined on Shape space, we compare the topological, metric and geodesic properties of Shape space with the ones on Wasserstein space. For example, we will show that if two metric spaces are isometric, so are the Shape spaces constructed from the respective Wasserstein spaces. This means that Shape space metrically only depends on the underlying metric point space.

The last chapter, 7, is subjected to treating infinitesimal change of shapes. We there begin with a section on isometric actions on the tangent bundle of the Wasserstein space and continue with the section where we define the notion of tangent space on the Shape space we have already mentioned above. However, due to technical reasons we constrain ourselves to the Shape spaces on \mathbb{R}^n , nevertheless conjecturing that this definition should be possible also in more general setups. The last section finally conducts all lines of effort to investigate the behavior of curves of probability measures stemming from a solution of the Schrödinger equation in Shape space. It shows that quantum motion in Shape space is natural and happens without loss of information once the transition from the wave function to its square modulus has been made. In particular, we can show for a specific solution that in Shape space it behaves like a geodesic. Also the gradient of the phase of the corresponding wave function can naturally be considered as a tangent vector on Shape space along its associated curve of probability measures.

Chapter 2

Wasserstein geometry

We start the thesis with an introductory chapter on Wasserstein geometry. Most of the material included here is standard within the theory of Wasserstein geometry, however, from Chapter 2.4 on we also add own statements. Despite for Lemma 2.53 and Theorem 2.57 we only give the proofs for statements that are our own. Our main references in this chapter are [AG09], [AGS08], [Gig08], [Gig12], [Vil03] and [Vil08].

Throughout this treatise let $\mathcal{P}(X)$ be the set of probability measures on the topological space X , with respect to the Borel σ -algebra $\mathcal{B}(X)$. A measurable map between two measurable spaces $T : (X, \mathcal{B}(X)) \rightarrow (Y, \mathcal{B}(Y))$ induces a map between the respective spaces of probability measures via the *pushforward* $T_{\#}$ of measures: $T_{\#} : \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$, $\mu \mapsto T_{\#}\mu$, where $T_{\#}\mu(A) := \mu(T^{-1}(A))$, for $A \in \mathcal{B}(Y)$. The *support* of a measure μ is defined by $\text{supp}(\mu) := \{x \in X \mid \text{every open neighbourhood of } x \text{ has positive } \mu\text{-measure}\}$. The Lebesgue measure on \mathbb{R}^n is denoted by λ .

2.1 Optimal transport

Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be probability measures. A natural question is how to *couple* μ and ν , i.e. how to relate them with each other. One possibility is to couple them with the help of a measurable map $T : X \rightarrow Y$, namely such that $T_{\#}\mu = \nu$. However, such a T cannot always be found. This is the case, for example, whenever μ is a Dirac measure and ν is not (maps cannot "split mass"). A further idea is to try to see μ and ν as two sides of the same thing, so to say. This is by looking at the elements of the set

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

the *admissible plans* between μ and ν . Here, $\pi^X : X \times Y \rightarrow X$ is the projection onto the X -component, i.e. $\pi^X(x, y) = x$. Similarly π^Y . $\text{Adm}(\mu, \nu)$ is never empty, since the product measure $\mu \otimes \nu$ is always an element. And in case there

is a map T like above, $\gamma = (Id, T)_{\#}\mu \in Adm(\mu, \nu)$. So any coupling in terms of maps can be seen as a coupling in terms of admissible plans.

Since $Adm(\mu, \nu)$ is not just not empty but in general has more than one element (for example if μ and ν are the sum of n Dirac measures), the question regarding the *best* coupling arises. Of course, a priori it is not clear what "best" actually means. Our perspective is that a coupling should be interpreted as a plan telling how to, instantaneously, rearrange μ such that it yields ν . Or, put differently, as a plan encoding how to *transport*, μ onto ν . In this interpretation we can think of $\gamma(A \times B)$ as being the amount of mass which is transported from A to B , where according to the definition of γ , $\gamma(A \times Y) = \mu(A)$ and $\gamma(X \times B) = \nu(B)$ for $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(Y)$.

To make precise what a best element should provide, we assume that we have further data which already relates X and Y with each other. Namely, we assume we have given a measurable function $c : X \times Y \rightarrow \mathbb{R}$. In our interpretation, the number $c(x, y)$ says how much it *costs* to transport one unit of something from $x \in X$ to $y \in Y$. Accordingly, we call c the *cost function*. The least cost for transporting μ to ν is then given by

$$(2.1) \quad C(\mu, \nu) := \inf_{\gamma \in Adm(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y).$$

Thus, a transport plan $\gamma_{opt} \in Adm(\mu, \nu)$ can be considered to be the best plan, or to be *optimal*, in case $C(\mu, \nu) = \int_{X \times Y} c(x, y) d\gamma_{opt}(x, y)$. The plan $\mu \otimes \nu$ can be seen as the most inefficient plan, since mass is brought from each measurable subset of positive measure of X to each measurable subset of positive measure of Y : $\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B)$. In case $\gamma \in Adm(\mu, \nu)$ is induced by a measurable map $T : X \rightarrow Y$, i.e. in case $\gamma = (Id, T)_{\#}\mu$, T is called *transport map* and the respective transportation cost is given by $\int_{X \times Y} c(x, T(x)) d\mu(x)$. The optimization problem

$$\tilde{C}(\mu, \nu) := \inf_T \int_X c(x, T(x)) d\mu(x),$$

where $T : X \rightarrow Y$ is a measurable map such that $T_{\#}\mu = \nu$ is called the *Monge formulation* of Optimal transport ([Mon81]), whereas (2.1) is called the *Kantorovich formulation* ([Kan58]).

Minimizer for (2.1) already exist under mild assumptions on c , as we will see in Theorem 2.5. For this, we need to introduce Polish spaces.

Definition 2.1 (Metric distance). A *metric distance*, or just *metric*, on a space X is a map $d : X \times X \rightarrow \mathbb{R}_{\geq 0}$ which satisfies the three conditions $d(x, y) = 0$ if and only if $x = y$, $d(x, y) = d(y, x)$ and $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

The open balls $B(x, r) := \{y \in X \mid d(x, y) < r\}$ form a base for a topology on X , turning X into a topological space. We call this topology the *topology induced by d* .

Definition 2.2 (Completely metrizable space). A topological space X is called *completely metrizable* if there exists at least one metric d on X which induces the given topology on X and which is such that (X, d) is a complete metric space.

Definition 2.3 (Polish space). A *Polish space* is a separable topological space X which is completely metrizable.

When we say that (X, d) is a Polish space, we mean that X is a Polish space and d is a metric on X that induces a topology which coincides with the topology of X and is such that (X, d) is a complete metric space. Such a metric d is called to *metrize* the Polish topology.

- Remark 2.4.*
- 1) Completeness is indeed a property of the metric, not of the topology. For example, with respect to Euclidean metric, the open unit interval $(0, 1)$ is not complete, but it is homeomorphic to \mathbb{R} via $x \mapsto \tan((x - 1/2)\pi)$, which is complete.
 - 2) Every finite or countably infinite discrete space is Polish and so is \mathbb{R}^n with the Euclidean topology. Every separable Banach space with the norm topology is Polish and every compact metrizable space. Finite and countable products and closed subspaces of Polish spaces are also Polish spaces.
 - 3) It can be shown that every probability measure on a Polish space is a Radon measure.

Theorem 2.5 (Existence of a minimizer). Let X and Y be Polish spaces and $c : X \times Y \rightarrow \mathbb{R}$ be a lower semicontinuous cost function such that $c(x, y) \leq a(x) + b(y) \forall (x, y) \in X \times Y$ for upper semicontinuous functions $a : X \rightarrow \mathbb{R} \cup \{-\infty\}$, $b : Y \rightarrow \mathbb{R} \cup \{-\infty\}$ such that $a \in L^1(\mu)$, $b \in L^1(\nu)$. Then there is an element in $\text{Adm}(\mu, \nu)$ which minimizes the Kantorovich formulation of Optimal transport.

See for example [Vil08] for a proof. The idea there is to show that $\int c d\gamma$ is a lower semicontinuous function on a compact set.

For us, the most important cost functions will be the metrics d which metrize the Polish space under consideration. In this case, of course, $X = Y$. With respect to their induced topology, metrics are continuous and they are bounded from below with $a = b = 0$.

There is a very important theorem expressing when a plan $\gamma \in \mathcal{P}(X \times Y)$ is optimal for its marginals. To be able to formulate it, we need to introduce some further notions. Again we will not detail the argumentation.

Definition 2.6 (c -cyclical monotone set). A set $Z \subset X \times Y$ is called *c -cyclically monotone* if for each $N \in \mathbb{N}$ and each subset $\{(x_i, y_i)\}_{1 \leq i \leq N} \subset Z$ of Z containing N elements, it is

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

for every permutation of the set $\{1, \dots, N\}$.

Definition 2.7 (*c_+ -concavity*). The c_+ -transform of a function $\psi : Y \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is the function ψ^{c_+} defined by

$$\begin{aligned} \psi^{c_+} : X &\longrightarrow \mathbb{R} \cup \{-\infty\} \\ x &\longmapsto \inf_{y \in Y} c(x, y) - \psi(y). \end{aligned}$$

A function $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ is called c -concave if it is the c_+ -transform of another function $\psi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$, i.e. if $\varphi = \psi^{c_+}$.

Definition 2.8 (*c -superdifferential*). For a c -concave function $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$, the c -superdifferential $\partial^{c_+}\varphi \subset X \times Y$ is defined by

$$\partial^{c_+}\varphi := \{(x, y) \in X \times Y \mid \varphi(x) + \varphi^{c_+}(y) = c(x, y)\}.$$

The c -superdifferential at $x \in X$ is the set $\partial^{c_+}\varphi(x) := \{y \in Y \mid (x, y) \in \partial^{c_+}\varphi\}$.

The following characterization will be important for us in Section 4.2.

Proposition 2.9. *Let $X = Y = \mathbb{R}^n$ and $c(x, y) = \|x - y\|^2/2$. A function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{-\infty\}$ is c -concave if and only if the map $\bar{\varphi}(x) := \|x\|^2/2 - \varphi(x)$ is convex and lower semicontinuous. In this case, $y \in \partial^{c_+}\varphi(x)$ if and only if $y \in \partial^-\bar{\varphi}(x)$, where ∂^- denotes the usual subdifferential from convex calculus.*

Now we cite from [AG09] the so called *Fundamental theorem of Optimal transport*.

Theorem 2.10 (*Fundamental theorem of Optimal transport*). *Let the cost function $c : X \times Y \rightarrow \mathbb{R}$ be continuous and bounded from below. Assume further that the measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are such that*

$$(2.2) \quad c(x, y) \leq a(x) + b(y),$$

for some functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$. For $\gamma \in \text{Adm}(\mu, \nu)$ the following three statements are equivalent:

- 1) *The plan γ is optimal.*
- 2) *The set $\text{supp}(\gamma)$ is c -cyclically monotone.*
- 3) *There exists a c -concave function φ such that $\max\{\varphi, 0\} \in L^1(\mu)$ and $\text{supp}(\gamma) \subset \partial^{c_+}\varphi$.*

Consequently, optimality depends only on the support of the plan γ , not on the distribution of mass. If γ is optimal for its marginals and $\tilde{\gamma} \in \mathcal{P}(X \times Y)$ is such that $\text{supp}(\tilde{\gamma}) \subset \text{supp}(\gamma)$, then $\tilde{\gamma}$ is optimal, too, for its marginals.

Remark 2.11. Let $T : X \rightarrow Y$ be a map with $T(x) \in \partial^{c_+}\varphi(x)$ for a c -concave function φ , for all $x \in X$. Then for every $\mu \in \mathcal{P}_2(X)$ such that condition (2.2) is satisfied for $\nu = T_{\#}\mu$, the map T is optimal between μ and $T_{\#}\mu$.

Remark 2.12. The notions c -cyclical monotonicity, c_+ -concavity and c -superdifferential generalize notions known from convex analysis: For $X = Y = \mathbb{R}^n$ and $c(x, y) = \langle x, y \rangle$ the Euclidean scalar product, a set is c -cyclical monotone if and only if it is cyclically monotone. A function is c -convex if and only if it is convex and lower semicontinuous and the c -subdifferential is the known subdifferential.

Next, one can be curious about when an optimal plan γ is actually induced by a map, i.e. when $\gamma = (Id, T)_{\#}\mu$ with μ being such that $\pi_{\#}^1\gamma = \mu$. One can show ([AG09]) that γ is induced by a map if and only if there exists a γ -measurable set $\Gamma \subset X \times Y$ on which γ is concentrated, such that for μ -a.e. x there exists only one $y = T(x) \in Y$ such that $(x, y) \in \Gamma$. In this case, γ is induced by the map T . Since we know from Theorem 2.10 that for optimal γ $supp(\gamma)$ is a subset of the c -superdifferential of a c -concave function φ , it is necessary to understand in which cases the c -superdifferential is single valued. As in [AG09], we will give an answer to this for the cases $X = Y = \mathbb{R}^n$, $c(x, y) = \|x - y\|^2/2$ and $X = Y = M$, $c = d^2/2$, where M is a connected, complete smooth Riemannian manifold and d the corresponding Riemannian metric distance (see Definition 2.16). In both cases, the characterization of the situation in which γ is induced by a map holds for so called *regular measures*, which we want to introduce first.

Definition 2.13 (c -c hypersurface). A subset $A \subset \mathbb{R}^n$ is called a *convex-concave hypersurface* (c -c hypersurface), whenever there exists convex functions $f, g : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ such that

$$A = \{(y, t) \in \mathbb{R}^{n-1} \times \mathbb{R} \mid t = f(y) - g(y)\}.$$

Definition 2.14. (Regular measure) A measure $\mu \in \mathcal{P}(\mathbb{R}^n)$ is called *regular*, in case $\mu(A) = 0$ for every c -c hypersurface $A \subset \mathbb{R}^n$.

Measures which are absolutely continuous with respect to the Lebesgue measure are, for example, regular.

The following theorem is due to Yann Brenier ([Bre87], [Bre91]).

Theorem 2.15 (Existence of optimal maps). *In case $\mu \in \mathcal{P}(\mathbb{R}^n)$ is such that $\int |x|^2 d\mu(x) < \infty$ and the cost function c is $c(x, y) = \|x - y\|^2/2$, the next two statements are equivalent:*

- 1) *For every $\nu \in \mathcal{P}(\mathbb{R}^n)$ with $\int |x|^2 d\nu(x) < \infty$, the optimal plan γ between μ and ν is unique and induced by a map T , i.e. it is $\gamma = (Id, T)_{\#}\mu$.*
- 2) *μ is regular.*

If either 1) or 2) hold, the optimal map T is the gradient of a convex function.

In fact, the convex function whose gradient is optimal is the c_+ -transform of the c -concave function φ for which $supp(\gamma) \subset \partial^{c_+}\varphi$.

As already announced above, a similar statement is true for Riemannian manifolds equipped with the Riemannian distance.

Definition 2.16 (Riemannian metric distance). Let (M, h) be a connected Riemannian manifold. The following formula determines a metric distance on M and is called *Riemannian (metric) distance* or *geodesic distance*:

$$d(x, y) := \inf_{\gamma} \int_0^1 h(\dot{\gamma}, \dot{\gamma}) dt,$$

for $x, y \in M$, where the infimum is taken over all differentiable curves γ for which $\gamma(0) = x$ and $\gamma(1) = y$.

Remark 2.17. The topology induced by d coincides with the original topology on M . By the Hopf-Rinow theorem, (M, d) is complete as a metric space if and only if M is geodesically complete, i.e. if for all $x \in M$ every geodesic $\gamma(t)$ starting at x is defined for all $t \in \mathbb{R}$ (compare also Theorem 6.34).

Definition 2.18 (Exponential map). Let (M, h) be a Riemannian manifold. Let $T_x M$ be the tangent space at $x \in M$ and let $U \subset T_x M$ be a neighborhood of the origin $0 \in T_x M$ such that the unique geodesic $\gamma_v(t)$ with starting point x , i.e. $\gamma_v(0) = x$, and initial velocity $v \in U$, i.e. $\dot{\gamma}(0) = v$, is well-defined at $t = 1$. Then the *exponential map* \exp_x is defined in the following way.

$$\begin{aligned} \exp_x : U &\longrightarrow M \\ v &\longmapsto \gamma_v(1). \end{aligned}$$

Remark 2.19. According to the Hopf-Rinow theorem, \exp_x can be defined on all of $T_x M$ if and only if (M, d) is a complete metric space. With the inverse function theorem, it is always possible to find a neighborhood of the origin on which the exponential map is a diffeomorphism on its image. The radius of the largest ball around $0 \in T_x M$ such that \exp is a diffeomorphism is called *injectivity radius* at point x . The infimum of the injectivity radii at all points is called the injectivity radius of (M, h) .

There is a generalization of regular measures on differentiable manifolds M , which we will need:

Definition 2.20 (Regular measure on M). A measure $\mu \in \mathcal{P}(M)$ is called *regular*, if it assigns no mass to the set of non-differentiability of any semiconvex function.

Again, in particular measures which are absolutely continuous with respect to the volume measure are regular.

We can now cite a variant of McCann's theorem.

Theorem 2.21 (Existence of optimal maps on manifolds). Let $\mu \in \mathcal{P}(M)$ be a probability measure on a Riemannian manifold M which is smooth, compact and without boundary. Let further $c = d^2/2$ be the cost function, d the Riemannian metric distance. Then the following two statements are equivalent:

- 1) The optimal transport plan between μ and any other measure $\nu \in \mathcal{P}(M)$ is unique and induced by a map T .

2) μ is regular.

In these cases, the optimal map T is of the form $T(x) = \exp_x(-\nabla\varphi(x))$, where $\varphi : M \rightarrow \mathbb{R}$ is a c -concave function.

Here again, the c -concave function φ is the one on whose c -superdifferential the optimal plan γ is concentrated on.

There are similar versions of Theorem 2.21 which do not require the compactness of the manifold, nor compactness of both the measures μ and ν (for example [Vil08], Theorem 10.41). However, in these cases, it seems either necessary to demand that M has nonnegative sectional curvature or that the gradient of φ has to be replaced by the so called *approximate gradient*.

Optimal maps can also be found for more general cost functions (see [Vil08]). In each of these cases, the c -concave function φ can be determined by the so called *partial differential equation of optimal transport* ([Vil08], Chapter 12). In case of the quadratic cost function on \mathbb{R}^n , $c(x, y) = \|x - y\|^2$, this is a Monge-Ampère equation:

$$\det \nabla^2 \varphi(x) = \frac{f(x)}{g(\nabla \varphi(x))},$$

where $\mu = f(x)d\lambda$, $\nu = g(y)d\lambda$ and $\nabla^2 \varphi$ the Hessian of φ .

2.2 Wasserstein spaces $W_p(X)$

From now on, we denote the set of probability measures which have finite p -th moment by $\mathcal{P}_p(X)$:

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

where $p \in [1, \infty)$. It is sufficient to demand the finiteness of the integral only for one $x_0 \in X$. Together with the triangle- and the Minkowski inequality it follows that finiteness holds for every $x_0 \in X$.

Given a cost function c , to every two probability measures μ and ν one can assign to them the number $C(\mu, \nu)$, see formula (2.1), being the optimal total cost for transporting one measure onto the other. One might want to think that at least for $X = Y$ this number encodes some kind of distance between μ and ν . Unfortunately, in general $C(\cdot, \cdot)$ does not satisfy the metric distance axioms. But in case the cost function is a power p of d , C is indeed a metric, if restricted to $\mathcal{P}_p(X)$.

Theorem 2.22. *Let (X, d) be a Polish metric space and $p \in [0, \infty)$, then*

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

is a metric distance.

Definition 2.23 (Wasserstein distances and Wasserstein spaces). The metric W_p is called p -th Wasserstein distance, or Wasserstein distance of order p . The tuple $(\mathcal{P}_p(X), W_p)$ is called Wasserstein space and is denoted by the symbol $W_p(X)$.

Instead of $W_2(X)$ we will often just write $W(X)$.

Remark 2.24. The map $X \rightarrow \mathcal{P}(X)$, $x \mapsto \delta_x$ is an isometric embedding of the underlying Polish space X into the Wasserstein space on X , since $W_p(\delta_x, \delta_y) = d(x, y)$ for every $p \in [1, \infty)$.

Remark 2.25. We note that the p -th moment of μ is nothing but the p -th Wasserstein distance of μ to δ_{x_0} to the power of p : $\int_X d^p(x, x_0) d\mu(x) = W_p^p(\mu, \delta_{x_0})$.

Example 2.26. Let $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ and $\nu = \delta_y$ then

$$W_p^p(\mu, \nu) = \sum_{i=1}^n a_i d^p(x_i, y).$$

It is important to know that the Wasserstein distance W_p metrizes the weak convergence in $\mathcal{P}_p(X)$. This means that the weak convergence of $(\mu_k)_{k \in \mathbb{N}}$ to μ in $\mathcal{P}_p(X)$ is equivalent to $W_p(\mu_k, \mu) \rightarrow 0$. This is a useful property, but not unique to the Wasserstein distances. As a reminder, we give the definition of weak convergence.

Definition 2.27 (Weak convergence in $\mathcal{P}(X)$). A sequence $(\mu_k)_{k \in \mathbb{N}} \subset \mathcal{P}(X)$ is said to *converge weakly* to $\mu \in \mathcal{P}(X)$ if and only if $\int \varphi d\mu_k \rightarrow \int \varphi d\mu$ for any bounded continuous function φ on X . This is denoted by $\mu_k \rightharpoonup \mu$. A sequence $(\mu_k)_{k \in \mathbb{N}} \subset \mathcal{P}_p(X)$ is said to converge weakly to $\mu \in \mathcal{P}_p(X)$ if and only if for $x_0 \in X$ it is:

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

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

We further list some more important properties of $W_p(X)$.

Theorem 2.28 (Some properties of $W_p(X)$). 1) $W_p(X)$ is compact, in case X is. On the other hand one can show that whenever X is unbounded, then $W_p(X)$ cannot be locally compact.

2) W_p is continuous on $\mathcal{P}_p(X) \times \mathcal{P}_p(X)$. That means, if $\mu_k \rightharpoonup_p \mu$ and $\nu_k \rightharpoonup_p \nu$, then $W_p(\mu_k, \nu_k) \rightarrow W_p(\mu, \nu)$.

3) $W_p(X)$ is complete and separable.

We are going to use 2) to show Corollary 6.23 and 3) to show that Shape space is complete and separable as well (6.51).

2.3 Wasserstein geodesics and displacement interpolation

Let us now have a look at the geodesic structure of Wasserstein space.

Definition 2.29 (*Constant speed geodesic*). A curve $(\gamma_t)_{t \in [0,1]}$, $\gamma_0 \neq \gamma_1$, in a metric space (X, d) is called a *constant speed geodesic* or *metric geodesic* in case that

$$(2.3) \quad d(\gamma_t, \gamma_s) = |t - s|d(\gamma_0, \gamma_1) \quad \forall t, s \in [0, 1].$$

In Section 2.4 we will see that constant speed geodesics actually have a constant metric derivative.

We will often abbreviate curves $(\gamma_t)_{t \in [0,1]}$ by omitting the brackets and the interval of definition, i.e we will often just write γ_t instead.

Lemma 2.30. *A curve $(\gamma_t)_{t \in [0,1]}$, $\gamma_0 \neq \gamma_1$, is a geodesic if and only if*

$$d(\gamma_t, \gamma_s) \leq |t - s|d(\gamma_0, \gamma_1) \quad \forall t, s \in [0, 1].$$

Proof. Only one direction needs extra proof. Let $d(\gamma_t, \gamma_s) \leq |t - s|d(\gamma_0, \gamma_1) \forall t, s \in [0, 1]$. Suppose, there is a pair of times $t > s \in [0, 1]$ such that $d(\gamma_t, \gamma_s) < |t - s|d(\gamma_0, \gamma_1)$. Then

$$\begin{aligned} d(\gamma_0, \gamma_1) &\leq d(\gamma_0, \gamma_s) + d(\gamma_s, \gamma_t) + d(\gamma_t, \gamma_1) \\ &< sd(\gamma_0, \gamma_1) + (t - s)d(\gamma_0, \gamma_1) + (1 - t)d(\gamma_0, \gamma_1) \\ &= d(\gamma_0, \gamma_1), \end{aligned}$$

which is a contradiction. □

In particular, this implies that a given geodesic never crosses a point twice. We further have the following corollary.

Corollary 2.31. *If a curve $(\gamma_t)_{t \in [0,1]}$ is a constant speed geodesic, then for every $s < u < t$ it is*

$$(2.4) \quad d(\gamma_s, \gamma_t) = d(\gamma_s, \gamma_u) + d(\gamma_u, \gamma_t).$$

Thus, one can picture γ_t to be the shortest path between its endpoints. However, the converse implication is not true. We will actually find a counterexample in Section 4.2.

Proof. Let $s, t, u \in [0, 1]$, $s < u < t$. Then

$$\begin{aligned} d(\gamma_s, \gamma_t) &\leq d(\gamma_s, \gamma_u) + d(\gamma_u, \gamma_t) = (u - s)d(\gamma_0, \gamma_1) + (t - u)d(\gamma_0, \gamma_1) \\ &= (t - s)d(\gamma_0, \gamma_1) = d(\gamma_s, \gamma_t). \end{aligned}$$

□

Remark 2.32. Since $x \mapsto \delta_x$ is an isometry, for every constant speed geodesic $\gamma(t)$ in X , $\delta_{\gamma(t)}$ is a constant speed geodesic in $W_p(X)$.

Definition 2.33 (Geodesic space). A metric space (X, d) is called *geodesic* if for every $x, y \in X$ there exists a constant speed geodesic $(\gamma_t)_{t \in [0,1]}$ with $\gamma_0 = x$ and $\gamma_1 = y$.

Define for all $t \in [0, 1]$ the *evaluation maps* $e_t : \text{Geod}(X) \rightarrow X$, $\gamma \mapsto \gamma_t$. Here, $\text{Geod}(X)$ is the metric space of all constant speed geodesics on X , which is complete and separable, if endowed with the sup norm, as soon as X is complete and separable. Then we can formulate the following set of properties.

Theorem 2.34 (Wasserstein geodesics). • *Whenever (X, d) is geodesic, $W_2(X)$ is geodesic as well.*

- *Furthermore, a curve $(\mu_t)_{t \in [0,1]}$ is a constant speed geodesic in $W_2(X)$ if and only if there exists a measure $\mu \in \mathcal{P}_2(\text{Geod}(X))$ such that $(e_0, e_1)_{\#} \mu \in \text{Adm}(\mu_0, \mu_1)$ is an optimal plan and $\mu_t = (e_t)_{\#} \mu$.*
- *In particular, if X is a Hilbert space, μ_t is a constant speed geodesic in $W_2(X)$ if and only if there exists an optimal transport plan $\gamma \in \text{Adm}(\mu_0, \mu_1)$ such that*

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

- *If further γ is induced by a map T , this latter formula simplifies to*

$$\mu_t = ((1-t)\text{Id} + tT)_{\#} \mu_0$$

Remark 2.35. Curves of the form $\mu_t = ((1-t)\pi^1 + t\pi^2)_{\#} \gamma$ are called *displacement interpolation* from $\pi^1_{\#} \gamma$ to $\pi^2_{\#} \gamma$ through γ . The previous theorem thus tells that within the Wasserstein structure, it is more natural to interpolate measures on the level of measurable sets than on the level of measures itself: The curve $\mu_t = (1-t)\mu_0 + t\mu_1$, where interpolation is done by just shifting mass from one measure to the other, has infinite length in Wasserstein space and is, although being continuous, not absolutely continuous. In $((1-t)\pi^1 + t\pi^2)_{\#} \gamma$ the prescription is that a set A has the amount of mass that γ is assigning to the set of all those points (x, y) with interpolation $(1-t)x + ty$ being an element of A .

Remark 2.36. Displacement interpolation in the case $X = \mathbb{R}^n$, $c(x, y) = \|x - y\|^2$ and the optimal plan γ being induced by a map can also be obtained as the solutions of the *time-dependent minimization problem* as studied by Villani in [Vil03] and [Vil08]:

$$\inf_{F_t(x)} \left\{ \int_X C \left[(F_t(x))_{0 \leq t \leq 1} \right] d\mu(x) \mid T_0 = \text{Id}, T_1_{\#} \mu = \nu \right\},$$

where the infimum is taken over all functions $[0, 1] \rightarrow X$, $t \mapsto F_t(x)$ which are continuous and piecewise C^1 for $d\mu$ -almost all $x \in X$. $C[(z_t)]$ is thereby the cost for the replacement along the trajectory z_t , which is in our case $C[(z_t)] = \int_0^1 |\dot{z}_t|^2 dt$.

There is an even richer geodesic structure, in case X is a Riemannian manifold.

Theorem 2.37 (Wasserstein geodesics on manifolds). *Let M be a smooth Riemannian manifold*

- μ_t is a constant speed geodesic in $W_2(M)$ if and only if there exists a $\gamma \in \mathcal{P}_2(TM)$ such that $\int |v|^2 d\gamma(x, v) = W_2^2(\mu_0, \mu_1)$ and $(Exp(t))_\# \gamma = \mu_t$. Here, $Exp(t) : TM \rightarrow M$, $(x, v) \mapsto exp_x(tv)$.
- The joining constant speed geodesic of two measures in $W_2(M)$ is unique, provided one of the measures is absolutely continuous with respect to the volume measure.
- If μ_t is a constant speed geodesic in $W_2(M)$, then for every $t \in (0, 1)$ and $s \in [0, 1]$ there exists only one optimal transport plan from μ_t to μ_s and this plan is induced by a map which is locally Lipschitz.

Proposition 2.38. *If μ_t is a constant speed geodesic in $W_2(M)$, where M is a Riemannian manifold, such that μ_0 is absolutely continuous with respect to the volume measure, then for every $t \in (0, 1)$, μ_t is absolutely continuous with respect to the volume measure, too.*

We want to introduce one further notion that we are going to need later.

Definition 2.39 (Non branching space). A metric space (X, d) is called *non branching*, if the following map is injective for every $t \in (0, 1)$

$$\begin{aligned} Geod(X) &\longrightarrow X \times X \\ \gamma &\longmapsto (\gamma_0, \gamma_t). \end{aligned}$$

Proposition 2.40. *If (X, d) is a complete, separable, locally compact and non branching geodesic space, then also $W_p(X)$ is non branching for $p \in (0, \infty)$. The converse is also true: If $W_p(X)$ is non branching, then so is X .*

A proof of this can be found in [Vil08] (Corollary 7.32). According to [AG09], the local compactness condition on X is, however, not necessary.

2.4 Dynamical equations on $W_2(M)$

In the upcoming, we will only be concerned with $W_2(M)$, where (M, h) means a smooth, connected and complete Riemannian manifold with Riemannian metric tensor h . Furthermore, we equip the set of measurable sections in TM , which we will denote by $\Gamma(TM)$, with an L^2 -topology.¹ That means, for $v \in \Gamma(TM)$

¹A section $v : M \rightarrow TM$ is measurable whenever all components in every chart are measurable. Equivalently, if and only if v itself is measurable with respect to the Borel σ -algebras given on M and TM .

we define

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

and

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

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

The goal now is to identify a differentiable structure on $W_2(M)$. Unfortunately, there is no smooth structure in the traditional sense, e.g. in the sense of [KM97] where infinite dimensional manifolds are modeled on convenient vector spaces. So one has to try to find structures that resemble formally what one requires of a manifold structure. The differentiable structure on $W_2(M)$ that will be defined in the end thus consists of ad hoc definitions, accurately tailored to Optimal transport and the Wasserstein metric structure, which mimic conventional differentiable and Riemannian behavior. This could be seen as a misfeature of the theory, but this ad hoc calculus yields powerful tools to perform calculations and, which is particularly important for us, provides a very natural language to capture characteristics of physical theories.

The basic idea of a tangent vector at a given point is that it indicates the direction a (differentiable) curve will be going infinitesimally from that point. Then traditionally the set of all such vectors which can be found to be tangential to some curve at a given fixed point are collected in the tangent space at that point. In our situation, we could enlarge $\mathcal{P}_2(M)$ to the Banach space of signed measures where one could expect the tangent space at point μ to be filled with all the Radon measures with zero integral and nonnegative outside $\text{supp}(\mu)$. However, we would like to take a different approach here which is more adapted to the geometric structure given by the Wasserstein distance. Unfortunately, on $W_2(M)$ there is no notion of smooth curves - but there is a notion of metric geodesics. In case the transport plan for the optimal transport between two measures is induced by a map T , the interpolating geodesic on Hilbert spaces (see Theorem 2.34) can be written as $\mu_t = ((1-t)Id + tT)_\# \mu_0$, thus being of the form $\mu_t = F_{t\#} \mu_0$. Generally, on Riemannian manifolds optimal transport between μ_0 and μ_t can be achieved by $\mu_t = F_{t\#} \mu_0$, $F_t = \exp(t\nabla\varphi)$ (see e.g [Vil08], Chapter 12)². In these cases, F_t is injective and locally Lipschitz for $0 < t < 1$ ([Vil03], Subsubsection 5.4.1). It is known from the theory of characteristics for partial differential equations that curves of this kind solve the weak continuity equation, together with the vector field to which integral lines F_t corresponds.

²This formula in particular displays nicely that optimal transport happens along geodesics of the underlying metric space.

Definition 2.41 (Continuity equation). Given a family of vector fields $(v_t)_{t \in [0, T]}$, a curve $\mu_t : [0, T] \rightarrow W_2(M)$ is said to solve the *continuity equation*

$$(2.5) \quad \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$$

in the weak sense, in case

$$(2.6) \quad \int_0^T \int_M \left(\frac{\partial}{\partial t} \varphi(x, t) + h(\nabla \varphi(x, t), v_t(x)) \right) d\mu_t(x) dt = 0$$

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

Theorem 2.42. Let $(F_t)_{t \in [0, T]}$ be a family of maps on M such that $F_t : M \rightarrow M$ is a bijection for every $t \in [0, T]$, $F_0 = Id$ and both $(t, x) \mapsto F_t(x)$ and $(t, x) \mapsto F_t^{-1}(x)$ are locally Lipschitz on $[0, T] \times M$. Let further $v_t(x)$ be a family of velocity fields on M such that its integral lines correspond to the trajectories F_t , and μ be a probability measure. Then $\mu_t = F_{t\#}\mu$ is the unique weak solution in $\mathcal{C}([0, T], \mathcal{P}(M))$ of $\frac{d}{dt} \mu_t + \nabla \cdot (v_t \mu_t) = 0$ with initial condition $\mu_0 = \mu$. Here, $\mathcal{P}(M)$ is equipped with the weak topology.

Theorem 2.42 is taken from [Vil03] where it is labeled as Theorem 5.34.

Remark 2.43. In case $\mu_0 = \rho_0 d\lambda$, it is possible to prove that $\mu_t = F_{t\#}\mu_0$ is also absolutely continuous with respect to the Lebesgue measure. Let $v \in \mathcal{C}^1(\mathbb{R}^n)$, then

$$\rho_t(x) = \left(\frac{\rho_0}{\det(DF_t)} \circ F_t^{-1} \right)(x).$$

See [AC08] for a justification of this formula. In case $\nabla v = 0$, (that means, in case v is incompressible - compare Remark 2.49), this reduces to $\rho_t(F_t(x)) = \rho_0(x)$ (see [Vil03], Exercise 5.37), i.e. also the density is transported.

The question now is, whether one can characterize the class of curves on $W_2(M)$ that admit a velocity in the manner of 2.41. A satisfying answer is given by Theorem 2.47, taken from [AG09].

Definition 2.44 (Absolutely continuous curve). Let (E, d) be an arbitrary metric space and I an interval in \mathbb{R} . A function $\gamma : I \rightarrow E$ is called *absolutely continuous, a.c.*, if there exists a function $f \in L^1(I)$ such that

$$(2.7) \quad d(\gamma(t), \gamma(s)) \leq \int_t^s f(r) dr, \quad \forall s, t \in I, t \leq s.$$

Definition 2.45 (Metric derivative). The *metric derivative* $|\dot{\gamma}|(t)$ of a curve $\gamma : [0, 1] \rightarrow E$ at $t \in (0, 1)$ is given as the limit

$$(2.8) \quad |\dot{\gamma}|(t) = \lim_{h \rightarrow 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}.$$

It is known that for absolutely continuous curves γ , the metric derivative exists for a.e. t . It is an element of $L^1(0, 1)$ and, up to sets of zero Lebesgue-measure, the minimal function satisfying equation (2.7) for γ . In this sense absolutely continuous functions enable a generalization of the fundamental theorem of calculus to arbitrary metric spaces.

- Example 2.46.**
1. Every metric geodesic is absolutely continuous and $|\dot{\gamma}|(t) = d(\gamma(0), \gamma(1))$.
 2. Let $E = \mathbb{R}^n$ with the distance induced by the Euclidean norm $\|\cdot\|$, then $|\dot{\gamma}|(t) = \|\frac{d\gamma}{dt}(t)\|$ at any point $t \in (a, b)$ where γ is differentiable.
 3. In particular, every curve of Dirac measures $\delta_{\gamma(t)}$ in Wasserstein space is a.c. if and only if $\gamma(t)$ is a.c., in that case their respective metric derivatives coincide.

Theorem 2.47 (Differential characterization of a.c. curves). *Let $\mu_t : [0, 1] \rightarrow W_2(M)$ be an a.c. curve. Then there exists a Borel family of vector fields $(v_t)_{t \in [0, 1]}$ on M such that the continuity equation (2.6) holds and*

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

Conversely, if a curve $\mu_t : [0, 1] \rightarrow W(M)$ is such that there exists a Borel family of vector fields $(v_t)_{t \in [0, 1]}$ with $\|v_t\|_{L^2(\mu_t)} \in L^1(0, 1)$, together with which it satisfies (2.6), then there exists an a.c. curve $\tilde{\mu}_t$ being equal to μ_t for a.e. t and satisfying

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

□

It can be shown, see [Gig12], Theorem 1.31, that for metric geodesics, the velocity vector field is well defined for every $t \in (0, 1)$, not just for a.e. t .

In the following we call a pair (μ_t, v_t) which together solves the continuity equation an *a.c. couple*. We further want to call v_t an *accompanying vector field* for μ_t .

It is possible to formulate a similar theorem in terms of the more regular Lipschitz-continuous curves instead of a.c. curves, ([Vil08], Theorem 13.8). In this case v_t is an element of $L^\infty(dt, L^2(\mu_t))$ instead of $L^1(dt, L^2(\mu_t))$. But in respect of defining a tangent space on $W_2(M)$ it is more meaningful to work with absolutely continuous curves.

Starting from geodesics we have arrived at the larger class of absolutely continuous curves on $W_2(M)$ and we would like to think about the vector fields v_t as being tangential to μ_t . To relate this to the standard notion of tangent vectors, we give the following lemma.

Lemma 2.48. *Let μ_t be such that there is a smooth curve $\gamma : I \rightarrow \mathbb{R}^n$ and $\mu_t = \delta_{\gamma(t)}$. Then μ_t solves the continuity equation in the weak sense with v_t which is such that $v_t(x) = \dot{\gamma}(t)$ in case $x = \gamma(t)$ and $v_t(x) = 0$ otherwise.*

Proof.

$$\begin{aligned}
\int_0^1 \int_{\mathbb{R}^n} \left(\frac{\partial}{\partial t} \varphi \right) (x, t) d\delta_{\gamma(t)} dt &= \int_0^1 \left(\frac{\partial}{\partial t} \varphi \right) (\gamma(t), t) dt \\
&= \int_0^1 \frac{d}{dt} (\varphi(\gamma(t), t)) - \langle \nabla \varphi(\gamma(t), t), \dot{\gamma}(t) \rangle dt \\
&= - \int_0^1 \int_{\mathbb{R}^n} \langle \nabla \varphi, v_t \rangle d\delta_{\gamma(t)} dt.
\end{aligned}$$

□

The same proof can actually be given for any $v_t = V \in \Gamma^\infty(T\mathbb{R}^n)$ such that $V(\gamma(t)) = \dot{\gamma}$.

Remark 2.49. The continuity equation $\partial_t \rho + \nabla j = 0$ plays a prominent role in physics as it describes the dynamics of a conserved quantity ρ , for example charge, energy or mass. In our case, like it is in fluid dynamics, $j = \rho v$. Is D a neighborhood of some point $x \in M$, the divergence $\nabla j(x)$ of a vector field j at x can be interpreted as the infinitesimal rate of change of $\frac{vol(\Phi_t(D))}{vol(D)}$, Φ_t denoting the flow of j and the volume vol is taken with respect to the metric volume form. Accordingly, in case $\nabla j < 0$, the flow is contracting in the vicinity of x , creating higher densities and in case $\nabla j > 0$, Φ_t leads to expansion of the quantity and thus to lower densities. Gauss' theorem tells us that this is directly related to the volume which flows in or out from D across its surface. In case $\nabla j = 0$, the quantity under consideration is called *incompressible*. This happens whenever $\rho = const.$ or $v = curl B$. A flow is called *stationary* in case its vector field does not change with time, for example in case $j = curl B$.

One generator for conserved quantities and thus for quantities being subject to the continuity equation is Noether's theorem. It states that conserved quantities correspond to continuous symmetries in Lagrangian systems.

In this treatise, the physical context in which we will encounter the continuity equation again is quantum physics (see Section 4.1). The conserved quantity there being $|\psi|^2$, the squared norm of the wave function interpreted as probability density, and $j = |\psi|^2 \nabla S$, where $\psi = |\psi| e^{-iS}$. Whereas in quantum physics the continuity equation subjecting $|\psi|^2$ is treated within standard calculus, we are going to interpret it in the weak sense which enables us to use the Wasserstein geometry formalism to express quantum mechanical features.

After having found a dynamical equation governing displacement interpolation, one can also ask about the dynamics governed by the family of vector fields v_t which corresponds to the trajectories of transportation. The answer is found in the pressureless Euler equation $\frac{dv}{dt} + \nabla_v v = 0$, see [Vil03]. Physically, it describes the velocity of particles that travel along geodesics without mutual interaction. In the case which is of most interest to us, namely $v = \nabla S(x, t)$,

this reduces to a Hamilton-Jacobi equation

$$(2.9) \quad \frac{\partial S}{\partial t} + \frac{\|\nabla S\|^2}{2} = 0.$$

We will reencounter a very similar equation again in Chapter 4 when dealing with the Madelung equations. The flow map of a smooth solution to the system consisting of the continuity equation and equation (2.9) determines optimal transportation for the cost d^2 in case there exists a function $u(y)$ such that $-S(t = 0, x) = \inf_{y \in M} \frac{d^2(x, y)}{2} - u(y)$ (i.e. if $-S(t = 0, x)$ is $\frac{d^2}{2}$ -concave). Equation (2.9) can be solved using the Hopf-Lax formula

$$S(t, x) = \inf_{y \in M} \left(S(t = 0, y) + \frac{d^2(x, y)}{2t} \right).$$

A rigorous treatment of this aspect can be found in [Vil03], Subsection 5.4.9.

Before we move on to define a tangent space from what we have gathered so far, we provide a proposition about absolutely continuous curves which we will need later on.

Proposition 2.50. *Let μ_t^1 and μ_t^2 be a.c. curves. Then also the convex mixing $\mu_t := (1 - \lambda)\mu_t^1 + \lambda\mu_t^2$ with $0 \leq \lambda \leq 1$ is an a.c. curve.*

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

$$W(\mu_s^i, \mu_t^i) \leq \int_s^t g_i(\tau) d\tau.$$

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

$$\begin{aligned} \pi_{\#}^1((1 - \lambda)\gamma_1 + \lambda\gamma_2)(A) &= ((1 - \lambda)\gamma_1 + \lambda\gamma_2)((\pi^1)^{-1}(A)) \\ &= (1 - \lambda)\gamma_1((\pi^1)^{-1}(A)) + \lambda\gamma_2((\pi^1)^{-1}(A)) \\ &= ((1 - \lambda)\mu_s^1 + \lambda\mu_s^2)(A) = \mu_s(A). \end{aligned}$$

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

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

This means that

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

□

From the previous proof we can extract the following general corollary about p -th Wasserstein distances:

Corollary 2.51. *Let (X, d) be a metric space and $\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}$ four probability measures on X . Then,*

$$\begin{aligned}
W_p((1-\lambda)\mu_{11} + \lambda\mu_{12}, (1-\lambda)\mu_{21} + \lambda\mu_{22}) &\leq \sqrt[p]{(1-\lambda)} W_p(\mu_{11}, \mu_{21}) \\
&\quad + \sqrt[p]{\lambda} W_p(\mu_{12}, \mu_{22}).
\end{aligned}$$

2.5 Induced differentiable structure on $W_2(M)$

In this section we want to use what we have found out in Section 2.4 to define a notion for tangent spaces at elements of $W_2(M)$. To do so, it seems as if we would need to perform a paradigm shift. A transition from the Lagrangian point of view to the Eulerian point of view. From particles moving on trajectories to fields changing in time. And from classical calculus to weak calculus. Interestingly, this has similarities with the physical transition from classical mechanics to quantum physics. We will see further matching parallels later on.

As we have seen in Theorem 2.47, every absolutely continuous curve in $W_2(M)$ admits an $L^1(dt)$ family of $L^2(\mu_t)$ -vector fields together with which the continuity equation is satisfied. We want to think of this family of vector fields v_t as being tangential to the curve. However, v_t is not unique, there are many vector fields which allow for the same motion of the density: Adding another family w_t with the (t -independent) property $\nabla(w_t \mu_t) = 0$ to v_t does not alter the equation. Luckily, Theorem 2.47 provides a natural criterion to choose a unique element among the v_t 's. According to this theorem, there is at least one family v_t such that $|\dot{\mu}_t| = \|v_t\|_{L^2(\mu_t)}$ for almost all t , i.e. that is of minimal norm for almost all t . Linearity of (2.7) with respect to v_t and the strict convexity of the L^2 -norms ensure the uniqueness of this choice, up to sets of zero measure with respect to t . We want to call such a couple (μ_t, v_t) , where v_t is the unique minimal tangent family for an a.c. curve μ_t , a *tangent couple*.

It now seems reasonable to define the tangent space at point μ as the set of $v \in L^2(M, \mu)$ with $\|v\|_\mu \leq \|v + w\|_\mu$ for all $w \in L^2(M, \mu)$ such that

$\nabla(w\mu) = 0$. This condition for $v \in L^2(M, \mu)$, however, is equivalent to saying that $\int_M h(v, w) d\mu = 0$ for all $w \in L^2(M, \mu)$ with $\nabla(w\mu) = 0$. This in turn is equivalent to the following, which we will take as the definition of the tangent space.

Definition 2.52 (Tangent space $T_\mu W(M)$). The *tangent space* $T_\mu W(M)$ at point $\mu \in W(M)$ is defined as

$$(2.10) \quad T_\mu W(M) := \overline{\{\nabla\varphi \mid \varphi \in \mathcal{C}_c^\infty(M)\}}^{L^2(M, \mu)} \subset L^2(M, \mu).$$

We also give the definition of the normal space:

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

Similar results on the continuity equation and corresponding definitions of the tangent space are also possible for $p > 2$, in case X in $\mathcal{P}_p(X)$ is a Hilbert space. See [AGS08], Chapter 8.

Lemma 2.53. *Let (μ_t, v_t) be an a.c. couple such that $\|v_t\|_{L^2(\mu_t)} \in L^1(0, 1)$. Then (μ_t, v_t) is a tangent couple if and only if $v_t \in T_{\mu_t} W(M)$ for almost every $t \in (0, 1)$.*

Proof. Let $v_t \in T_{\mu_t} W(M)$ and let \tilde{v}_t be the tangent vector field along μ_t . Then $\nabla(v_t - \tilde{v}_t) = 0$. Since ∇ only acts on the spacial component, this equation can be considered to be true also only on M , for almost every $t \in (0, 1)$. This means that $v_t - \tilde{v}_t \in T_\mu^\perp W(M)$. Therefore, $v_t - \tilde{v}_t = 0$ and thus $v_t = \tilde{v}_t$ for a.e. t . \square

It can be shown (as stated in [Vil08], Chapter 13) that at least for measures μ which are absolutely continuous the vector space defined in equation (2.10) is isometric to the closed vector space generated by $d^2/2$ -convex functions ψ , equipped with the norm $\|\varphi\| := \|\nabla\varphi\|_{L^2(M, \mu)}$, which connects to Theorem 2.21.

In Section 2.4 we have seen that the velocity v_t field along a geodesic μ_t comes from time dependent optimal transport (see Proposition 1.32 in [Gig12] for more details). But we can see more generally that also tangent vector fields of a.c. curves can be obtained from optimal transport ([AG09], Proposition 2.32): Given a tangent couple (μ_t, v_t) in $W(\mathbb{R}^n)$ such that μ_t is regular for every t and let T_t^s be the optimal transport map from μ_s to μ_t . Then for a.e. $t \in [0, 1]$ it is, with respect to the limit in $L^2(\mu_t)$,

$$(2.11) \quad v_t = \lim_{s \rightarrow t} \frac{T_t^s - Id}{s - t}.$$

A similar statement is true also on manifolds and even in case there does not exist an optimal transport map between μ and ν (see [Gig12], Theorem 1.31).

It is not difficult to see that $\dim T_\delta W(M) = \dim M$, whereas in most of the cases $\dim T_\mu W(M) = \infty$.³ In general, it can be shown that as long as μ is concentrated on an at most countable set, it is $T_\mu W(M) = L^2(M, \mu)$ (see [Gig12], Remark 1.33). The following lemma displays that morally, the more points are contained in the support of the measure, the higher gets the dimension. On the other hand, every probability measure can be approximated by a sequence of measures with finite support (see [Vil08] Thm 6.18), so that in each neighborhood of every measure there is an element μ with $\dim T_\mu W(M) < \infty$.

Proposition 2.54. *Let $\mu_1, \mu_2 \in W(M)$ be such that $\text{supp}(\mu_1) \cap \text{supp}(\mu_2) = \emptyset$ and let $0 < \lambda < 1$. Then*

$$T_{(1-\lambda)\mu_1 + \lambda\mu_2} W(M) \cong T_{\mu_1} W(M) \oplus T_{\mu_2} W(M)$$

as Hilbert spaces with $\langle v_1 \oplus v_2, w_1 \oplus w_2 \rangle := \langle v_1, w_1 \rangle_{L^2(\mu_1)} + \langle v_2, w_2 \rangle_{L^2(\mu_2)}$.

For the proof of this we will use the following statement.

Lemma 2.55. *Let A, B be closed subsets of a metric space (X, d) such that $A \cap B = \emptyset$. Then there exists an open set $U \subset X$ such that $A \subset U$ and $B \cap U = \emptyset$.*

Proof of Lemma 2.55. The distance between a point $x \in X$ and a closed set $B \subset X$ is defined as $d(x, B) := \inf_{b \in B} d(x, b)$. It is $d(x, B) = 0$ if and only if $x \in B$. That means, for every $a \in \dot{A} := A \setminus \dot{A}$, where \dot{A} is the interior of A , $d_a := d(a, B) > 0$. Define U as

$$U := \dot{A} \cup \bigcup_{a \in \dot{A}} \mathcal{B}_{d_a/2}(a).$$

By $\mathcal{B}_{d_a/2}(a)$ we mean the open ball with radius $d_a/2$ centered around a . U is open as a union of open sets. Furthermore, $A \subset U$ and $U \cap B = \emptyset$. □

Proof of Proposition 2.54. First, let $v_1 \oplus v_2 \in T_{\mu_1} W(M) \oplus T_{\mu_2} W(M)$. Let \tilde{v}_i be any representative of $v_i \in T_{\mu_i} W(M) \subset L^2(\mu_i)$, $i = 1, 2$, and set $\hat{v}_i(x) := \tilde{v}_i(x)$ for all $x \in \text{supp}(\mu_i)$ and $\hat{v}_i(x) := 0$ otherwise. For $\mu := (1 - \lambda)\mu_1 + \lambda\mu_2$ it then holds, $\|\hat{v}_i\|_{L^2(\mu)} = \|\hat{v}_i\|_{L^2(\mu_i)} < \infty$ and $\hat{v}_1 + \hat{v}_2$ can be interpreted as an element of $L^2(\mu)$. This element is independent of the representative of v_i chosen in the beginning.

Let us now see that $\hat{v}_1 + \hat{v}_2$ is also an element of $T_\mu W(M)$. We know that for both $i = 1, 2$ there exists a sequence $(\varphi_n^i)_{n \in \mathbb{N}} \subset \mathcal{C}_c^\infty(M)$ such that $\|\hat{v}_i - \nabla \varphi_n^i\|_{L^2(\mu_i)} \rightarrow 0$. However, this is not necessarily true in the $L^2(\mu)$ -topology. For this, we need to find sequences $(\psi_n^i)_{n \in \mathbb{N}} \subset \mathcal{C}_c^\infty(M)$, such that $\|\hat{v}_i - \nabla \psi_n^i\|_{L^2(\mu_i)} \rightarrow 0$, but for which $\psi_n^i(x) = 0$ for each $x \in \text{supp}(\mu_j)$, $i \neq j$,

³There is an alternative definition of tangent space due to N. Gigli, which deals with this issue. In Appendix A.2.3 we briefly introduce this notion and explain why we prefer to work with the above stated definition.

and for each $n \in \mathbb{N}$. In that case, also $\|\hat{v}_i - \nabla\psi_n^i\|_{L^2(\mu)} \rightarrow 0$ would hold true, for both $i = 1, 2$, and with this, also $\|(\hat{v}_1 + \hat{v}_2) - (\nabla\psi_n^1 + \nabla\psi_n^2)\|_{L^2(\mu)} \rightarrow 0$. The latter is true because

$$\begin{aligned}
& \|(\hat{v}_1 + \hat{v}_2) - (\nabla\psi_n^1 + \nabla\psi_n^2)\|_{L^2(\mu)}^2 = \|(\hat{v}_1 - \nabla\psi_n^1) + (\hat{v}_2 - \nabla\psi_n^2)\|_{L^2(\mu)}^2 \\
&= \int_M h((\hat{v}_1 - \nabla\psi_n^1) + (\hat{v}_2 - \nabla\psi_n^2), (\hat{v}_1 - \nabla\psi_n^1) + (\hat{v}_2 - \nabla\psi_n^2)) d\mu \\
&= (1 - \lambda) \int_M h(\hat{v}_1 - \nabla\psi_n^1, \hat{v}_1 - \nabla\psi_n^1) d\mu_1 + \lambda \int_M h(\hat{v}_2 - \nabla\psi_n^2, \hat{v}_2 - \nabla\psi_n^2) d\mu_2 \\
&= (1 - \lambda) \|\hat{v}_1 - \nabla\psi_n^1\|_{L^2(\mu_1)}^2 + \lambda \|\hat{v}_2 - \nabla\psi_n^2\|_{L^2(\mu_2)}^2 \longrightarrow 0.
\end{aligned}$$

To construct such sequences $(\psi_n^i)_{n \in \mathbb{N}}$, we start with the sequences $(\varphi_n^i)_{n \in \mathbb{N}}$. The idea is to multiply each member of the sequence with a bump function such that it keeps its value on $\text{supp}(\mu_i)$ but falls off rapidly outside of this support such that it vanishes on the support of μ_j , $i \neq j$.

Let $n \in \mathbb{N}$. We know that $K_n^i := \text{supp}(\varphi_n^i) \cap \text{supp}(\mu_i)$ is compact. From Lemma 2.55 we know that there is an open set $U_n^i \subset X$ such that $K_n^i \subset U_n^i$, but $U_n^i \cap \text{supp}(\mu_j) = \emptyset$, $i \neq j$. Let $\chi_n^i \in C_c^\infty(M)$ such that $\chi_n^i(x) = 1$ for all $x \in K_n^i$ and $\chi_n^i(x) = 0$ for all $x \in M \setminus U_n^i$. (For the existence of such functions, see [Hör90].) Then $\chi_n^i \cdot \varphi_n^i \in C_c^\infty(M)$, $\text{supp}(\chi_n^i \cdot \varphi_n^i) \cap \text{supp}(\mu_j) = \emptyset$ for $i \neq j$ and $\chi_n^i \cdot \varphi_n^i(x) = \varphi_n^i(x)$ for all $x \in \text{supp}(\mu_i)$. Thus, $(\psi_n^i)_{n \in \mathbb{N}} := (\chi_n^i \cdot \varphi_n^i)_{n \in \mathbb{N}}$ is the sequence we were looking for and $\hat{v}_1 + \hat{v}_2 \in T_\mu W(M)$.

Let us now prove linearity of the map $v_1 \oplus v_2 \mapsto \hat{v}_1 + \hat{v}_2 \in L^2(\mu)$. For this let $(v_1 \oplus v_2) + (w_1 \oplus w_2) = (v_1 + w_1) \oplus (v_2 + w_2) \in T_{\mu_1} W(M) \oplus T_{\mu_2} W(M)$. We then have $\widehat{v_1 + w_1} + \widehat{v_2 + w_2} = \hat{v}_1 + \hat{w}_1 + \hat{v}_2 + \hat{w}_2 = (\hat{v}_1 + \hat{v}_2) + (\hat{w}_1 + \hat{w}_2)$, since the operation " + " is done pointwise.

Let now be $v \in T_\mu W(M) \subset L^2(\mu)$ and choose an arbitrary representative \hat{v} of v . Define \hat{v}_i to be $\hat{v}_i(x) := \hat{v}(x)$ for all $x \in \text{supp}(\mu_i)$ and $\hat{v}_i(x) = 0$ otherwise. Then $\|\hat{v}_i\|_{L^2(\mu_i)} = \|\hat{v}\|_{L^2(\mu)} < \infty$, $i = 1, 2$, and $\hat{v}_1 + \hat{v}_2$ is an element of the equivalence class v . Let $v_i = [\hat{v}_i] \in L^2(\mu_i)$, then $v_i \in T_{\mu_i} W(M)$ for similar reasons as above and v is the image of $v_1 \oplus v_2$ under the above constructed map. Injectivity is straightforward.

Lastly, we show the invariance of the scalar product.

$$\begin{aligned}
\langle v_1 \oplus v_2, w_1 \oplus w_2 \rangle &= \langle v_1, w_1 \rangle_{L^2(\mu_1)} + \langle v_2, w_2 \rangle_{L^2(\mu_2)} \\
&= \langle \hat{v}_1, \hat{w}_1 \rangle_{L^2(\mu_1)} + \langle \hat{v}_2, \hat{w}_2 \rangle_{L^2(\mu_2)} \\
&= \langle \hat{v}_1, \hat{w}_1 \rangle_{L^2(\mu)} + \langle \hat{v}_2, \hat{w}_2 \rangle_{L^2(\mu)} \\
&= \langle \hat{v}_1 + \hat{v}_2, \hat{w}_1 + \hat{w}_2 \rangle_{L^2(\mu)}
\end{aligned}$$

□

Corollary 2.56. *Let $(\mu_t^1, v_t^1), (\mu_t^2, v_t^2)$ be tangent couples such that for all $t \in [0, 1]$, $\text{supp}(\mu_t^1) \cap \text{supp}(\mu_t^2) = \emptyset$. Then the tangent vector field along $\mu_t := (1 - \lambda) \mu_t^1 + \lambda \mu_t^2$, $0 < \lambda < 1$, is $v_t^1 \oplus v_t^2 \in T_{\mu_t^1} W(M) \oplus T_{\mu_t^2} W(M) \cong T_{\mu_t} W(M)$.*

Proof. We already know from Proposition 2.50 that μ_t is an a.c. curve. Given the assumptions on the supports of μ_t^1 and μ_t^2 , we now want to explicitly calculate the tangent vector field along μ_t . Using the same notation as in Proposition 2.54, we have

$$\begin{aligned}
& \frac{d}{dt}\mu_t + \nabla((v_t^1 \oplus v_t^2)\mu_t) \\
&= \int \int \left(\frac{\partial}{\partial t}\varphi(x, t) + h(\nabla\varphi(x, t), (\hat{v}_t^1 + \hat{v}_t^2)) \right) d\mu_t dt \\
&= (1 - \lambda) \int \int \left(\frac{\partial}{\partial t}\varphi(x, t) + h(\nabla\varphi(x, t), \hat{v}_t^1) \right) d\mu_t^1 dt \\
&\quad + \lambda \int \int \left(\frac{\partial}{\partial t}\varphi(x, t) + h(\nabla\varphi(x, t), \hat{v}_t^2) \right) d\mu_t^2 dt \\
&= (1 - \lambda) \left(\frac{d}{dt}\mu_t^1 + \nabla(\hat{v}_t^1\mu_t^1) \right) + \lambda \left(\frac{d}{dt}\mu_t^2 + \nabla(\hat{v}_t^2\mu_t^2) \right) = 0.
\end{aligned}$$

Now, since $(\mu_t, v_t^1 \oplus v_t^2)$ is an a.c. couple and $v_t^1 \oplus v_t^2 \in T_{\mu_t}W(M)$ for a.e. t , according to Lemma 2.53 it only remains to see that $\|v_t^1 \oplus v_t^2\|_{L^2(\mu_t)} \in L^1(0, 1)$.

This is true, since by definition, $\|v_t^1 \oplus v_t^2\|_{L^2(\mu_t)} = \sqrt{\|v_t^1\|_{L^2(\mu_t^1)} + \|v_t^2\|_{L^2(\mu_t^2)}} \leq \|v_t^1\|_{L^2(\mu_t^1)} + \|v_t^2\|_{L^2(\mu_t^2)} \in L^1(0, 1)$. \square

We call the disjoint union of all tangent spaces,

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

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

$$\begin{array}{ccc}
TW(M) & \xrightarrow{B} & TW(N) \\
\pi^M \downarrow & & \downarrow \pi^N \\
W(M) & \xrightarrow{F} & W(N)
\end{array}$$

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

Now that we have found a meaningful definition for tangent spaces, the next step is to try to define a Riemannian structure on $W_2(M)$. Intuition comes from the following formula which is due to J.-D. Benamou and Y. Brenier ([BB99]). It shows that the Wasserstein distance W_2 , having been defined through the, static, optimal transport problem, can be recovered by a dynamic formula, reminiscing the length functional on Riemannian manifolds, defining the Riemannian metric distance.

Theorem 2.57 (Benamou-Brenier formula). *Let $\mu, \nu \in \mathcal{P}_2(M)$, then it is*

$$(2.12) \quad W(\mu, \nu) = \inf_{(\mu_t, v_t)} \int_0^1 \|v_t\|_{L^2(\mu_t)} dt,$$

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

Remark 2.58. Instead of equation (2.12) one can similarly show that $W(\mu, \nu)^2 = \inf_{(\mu_t, v_t)} \int_0^1 \|v_t\|_{L^2(\mu_t)}^2 dt$. So W^2 can also be seen, up to a factor 1/2, as the minimal kinetic energy required to get from μ to ν .

Because of formula (2.12), we want to interpret the expression $\int_0^1 \|v_t\|_{L^2(\mu_t)} dt$ as the length of a curve μ_t . It can be shown (as stated in [Vil08], Chapter 13) that the infimum is achieved if and only if μ and ν allow for an optimal plan. The minimizing curve will then be a tangent couple where μ_t is a geodesic.

Proof. We basically restate the proof given in [AG09]. Using the second part of Theorem 2.47 we first have

$$W(\mu, \nu) \leq \int_0^1 |\dot{\mu}_t| dt \leq \int_0^1 \|v_t\|_{L^2(\mu_t)} dt.$$

For the reverse inequality it is sufficient to consider a constant speed geodesic μ_t connecting μ and ν . According to the first part of Theorem 2.47 there is a family of vector fields v_t such that (μ_t, v_t) is an a.c. couple and $\|v_t\|_{L^2(\mu_t)} \leq |\dot{\mu}_t| = W(\mu, \nu)$ for a.e. $t \in [0, 1]$. Because of this we thus have

$$W(\mu, \nu) \geq \int_0^1 \|v_t\|_{L^2(\mu_t)} dt.$$

□

Being inspired by this resemblance of formulas, we define the following.

Definition 2.59 (*Formal Riemannian tensor on $W_2(M)$*). The *formal Riemannian metric tensor* H_μ on $W(M)$ at point $\mu \in W(M)$ is defined as

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

Indeed, since $\|v_t\|_{L^2(\mu_t)} = \sqrt{\int_M h(v, v) \, d\mu} = \sqrt{H_\mu(v, v)}$, we now have

$$(2.13) \quad W(\mu, \nu) = \inf_{(\mu_t, \nu_t)} \int_0^1 \sqrt{H_\mu(v_t, v_t)} \, dt.$$

The tuple $(T_\mu W(M), H_\mu)$ constitutes a Hilbert Space.

Gigli [Gig08], however, emphasizes that Definition 2.52 does not allow for a traditional Riemannian structure on $W_2(M)$ since the natural exponential map $v \mapsto \exp_\mu(v) := (Id + v)_\# \mu$ has injectivity radius 0 for every μ .

Chapter 3

Differentiable maps between Wasserstein spaces

In this chapter, we want to propose a notion of differentiability for maps between Wasserstein spaces $W(M)$ and $W(N)$, where M and N are Riemannian manifolds.

3.1 Absolutely continuous maps

Since $W(M)$ and $W(N)$ are no "real" manifolds, we cannot compose a map between Wasserstein spaces with charts, to be able to apply Euclidean calculus. Instead, let us recall the fact that a map $f : M \rightarrow N$ is differentiable iff it maps differentiable curves to differentiable curves. Having only a notion of absolutely continuous curves, which are metrically differentiable almost everywhere and which lie at the fundament of the construction of tangent spaces, we start with formulating the following definition.

Definition 3.1 (*Absolutely continuous map*). A map $F : W(M) \rightarrow W(N)$ is called *absolutely continuous*, or, *a.c.*, if the curve $F(\mu_t) \subset W(N)$ is absolutely continuous up to redefining $t \mapsto \mu_t$ on a zero set, whenever $\mu_t \subset W(M)$ is absolutely continuous.

We will build our notion of differentiable maps between Wasserstein spaces on this idea of absolutely continuous maps. Before we continue to do so, let us first find some conditions with which maps are absolutely continuous. For this, we want to recall the notion of *proper* maps. Later, in Chapter 6, we will, for similar reasons as we do here, also deal with proper actions and proper spaces.

Definition 3.2 (*Proper map*). A continuous map $f : X \rightarrow Y$ between a Hausdorff space X and a locally compact Hausdorff space Y is called *proper*, if for compact subsets $K \subset Y$, the preimage $f^{-1}(K) \subset X$ is compact in X .

For example, every map $f : X \rightarrow Y$ that is injective, continuous and open is proper.¹

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

Theorem 3.3. *Let $F : W(M) \rightarrow W(N)$ be given as $F(\mu) = f_{\#}\mu$, $f : M \rightarrow N$ being smooth and proper and such that $\sup_{x \in M} \|df_x\| < \infty$. Then F is absolutely continuous and for every tangent couple (μ_t, v_t) , the tuple $(F(\mu_t), dF_{\mu_t}(v_t))$ is an a.c. couple, where*

$$(3.1) \quad dF_{\mu_t}(v_t)_y := \int_{f^{-1}(y)} df_x(v_{tx}) d\mu_t^y(x)$$

for almost every t and for $y \in f(M)$. Here, $df_x : T_x M \rightarrow T_{f(x)} N$ denotes the tangent map of f at point x , v_{tx} means the vector field v_t at point $x \in M$ and $\mu_t^y(x)$ is defined through the disintegration theorem. The latter means $d\mu_t(x) = d\mu_t^y(x) df_{\#}\mu_t(y)$ (see Appendix A.1.3).² For all $y \notin f(M)$, we set $dF_{\mu_t}(v_t)_y = 0$ (or any other value).

Since f is not injective in general, $df(v)_y$ is generally multi valued³. So we take the mean value over all the vectors $df_x(v_{tx})$ as the image vector $dF_{\mu_t}(v_t)_y$ of the vector field v_t at point y , where x stands for the elements of the fiber $f^{-1}(y)$. In case f is injective, $dF_{\mu}(v)$ reduces to $df(v)$ for every μ , which then can be regarded as full-fledged vector field.

Our naming of the vector field along $F(\mu_t)$, $dF_{\mu_t}(v_t)$ is, of course, very suggestive. Indeed, since the map $(v, \mu) \mapsto dF_{\mu}(v)$ is linear in v , Theorem 3.3 supports a natural definition for a notion of differentiability for absolutely continuous maps F . However, before we give such a definition, we need to make some further preparatory observations.

Let us first continue with proving Theorem 3.3.

Proof. Let μ_t be an a.c. curve. Using Theorem 2.47, we want to prove that there exists a family of vector fields $(\tilde{v}_t)_{t \in [0,1]}$ with $\int_0^1 \|\tilde{v}_t\|_{L^2(F(\mu_t))} dt < \infty$, such that $(F(\mu_t), \tilde{v}_t)$ is an a.c. couple.

Let $(v_t)_{t \in [0,1]}$ be the tangent vector field of μ_t . For each t for which $v_t \in T_{\mu_t} W(M)$ (i.e. almost everywhere) and for $y \in f(M)$ we define,

¹Definitions of proper maps may vary a bit from reference to reference. In Appendix A.1.2 we give a more general definition for proper maps between arbitrary topological spaces, for the interested reader. The definitions of locally compact and Hausdorff spaces can be looked up in A.1.1.

²Note that what in A.1.3 appears as lower index y , now appears as upper index y since here were additionally dealing with the t -dependence of μ_t .

³Although $df_x : T_x M \rightarrow T_{f(x)} M$ is well defined for every x as a mapping between tangent spaces, it is not well defined as a mapping between vector fields as long as f is not injective.

$$dF_{\mu_t}(v_t)_y := \int_{f^{-1}(y)} df(v_{tx}) d\mu_t^y(x).$$

For all $y \notin f(M)$, we set $dF_{\mu_t}(v_t)_y = 0$. We will prove that $dF_{\mu_t}(v_t)$ is an example of such vector fields \tilde{v}_t we are looking for.

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

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

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

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

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

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

For simplicity, the variable t has been omitted in the last calculation.

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

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

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

Remark 3.4. The vector field $dF_\mu(v)$, and thus the differentiability of $F = f_{\#}$ at point $\mu \in W(M)$ as we will define it later, depends on the differentiability of f at every point $x \in M$. It thus does not seem so easy, to find a more general version of Theorem 3.3 in terms of conditions on f that do not guarantee smoothness and compactness of the support of $\varphi \circ f$. However, while it would probably not be so fruitful to experiment with milder conditions on the boundedness of df ,

one could, nevertheless, speculate on weakening the other conditions. But it is inevitable that $\varphi \circ f$ has to allow for at least an approximation of test-functions. There are several statements about density of test-functions but which don't completely fit to our situation. Demanding for example $f \in \mathcal{C}^1(M, N)$, yields $\varphi \circ f \in L^1(\mu)$ for every $\mu \in W(M)$. There is a theorem, [Aub82] §5, saying that on a smooth Riemannian manifold M , $\overline{\mathcal{C}_c^\infty(M)} = L^p(\mu)$, $1 \leq p < \infty$, where $\mu = dh$ is the Riemannian measure. However, as soon as μ is an arbitrary positive Radon measure, only $\overline{\mathcal{C}_c(M)} = L^p(\mu)$ has been established in that reference. There are similar, more general theorems on (bounded or open) subsets of \mathbb{R}^n and for densities which obey certain regularity conditions ([Kuf85],[Kil94]), but it seems, there is no statement that is completely covering our case.

3.2 About the image of dF_μ

For Theorem 3.3 we did not need to test whether $dF_\mu(v) \in T_{F(\mu)}W(M)$ for all $v \in T_\mu W(N)$, since we only needed $(F(\mu_t), dF_{\mu_t}(v_t))$ to be an a.c. couple. But is it still true, in case (μ_t, v_t) is a tangent couple?

Let us first properly assert that $dF_\mu(v) \in L^2(\mu)$. In the rest of this section let $F : W(M) \rightarrow W(N)$ be as in Theorem 3.3 and $dF_\mu(v)$ as in formula (3.1), unless stated otherwise.

Lemma 3.5. *For every $\mu \in W(M)$ and $v \in T_\mu W(M)$, $dF_\mu(v) \in L^2(F(\mu))$.*

Proof. As in the proof of Theorem 3.3, $\|dF_\mu(v)\|_{L^2(F(\mu))} \leq C\|v\|_{L^2(\mu)}$, which is finite. \square

Now that we know that $dF_\mu(v)$ is always an element of $L^2(F(\mu))$, we can consider formula (3.1) as the prescription for a map between $T_\mu W(M)$ and $L^2(F(\mu))$. It is useful to know that this map is always bounded.

Theorem 3.6 (Boundedness of dF). *For each $\mu \in W(M)$, $dF_\mu : T_\mu W(M) \rightarrow L^2(F(\mu))$ is bounded with*

$$(3.2) \quad \|dF_\mu\| \leq \operatorname{ess\,sup}_{x \in M}^\mu \|df_x\|.$$

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

The right-hand side of equation (3.2) is finite since we demanded $\sup_{x \in M} \|df_x\|$ to be finite.

Proof. Taking similar steps as in the proof of Theorem 3.3, we have:

$$\begin{aligned}
\|dF_\mu\|^2 &= \left(\sup_{\|v\|_{T_\mu W(M)}=1} \|dF_\mu(v)\|_{L^2(F(\mu))} \right)^2 \\
&= \sup_{\|v\|_{T_\mu W(M)}=1} \|dF_\mu(v)\|_{L^2(F(\mu))}^2 \\
&= \sup_{\|v\|_{T_\mu W(M)}=1} \int_N \left\| \int_{f^{-1}(y)} df_x(v_x) d\mu^y(x) \right\|_{T_y N}^2 df_\# \mu(y) \\
&\leq \sup_{\|v\|_{T_\mu W(M)}=1} \left(\int_M \|v_x\|_{T_x M}^2 d\mu \cdot \operatorname{ess\,sup}_{x \in M}^\mu \|df_x\|^2 \right) \\
&= \left(\sup_{\|v\|_{T_\mu W(M)}=1} \int_M \|v_x\|_{T_x M}^2 d\mu \right) \cdot \operatorname{ess\,sup}_{x \in M}^\mu \|df_x\|^2 \\
&= \operatorname{ess\,sup}_{x \in M}^\mu \|df_x\|^2.
\end{aligned}$$

□

We have proven an inequality in Theorem 3.6. However, there are indeed functions F such that equality holds for every μ .

Example 3.7. Let $g : M \rightarrow M$ be a Riemannian isometry, i.e. $g^*h = h$, where h is the Riemannian metric tensor on M . Then, for $F = g_\#$ and for all $\mu \in W(M)$, $\|dF_\mu\| = \operatorname{ess\,sup}_{x \in M}^\mu \|df_x\| = 1$.

This is, because on the one hand, for all $x \in M$, $\|dg_x\| = \sup_{\|v\|=1} \|dg_x(v)\| = 1$, since dg is an isometry between the tangent spaces $T_x M$ and $T_{g(x)} M$ and the norm is taken with respect to the Riemannian tensor. On the other hand,

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

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

Lemma 3.8. *Let $\mu = \delta_x$, for $x \in M$. Then $dF_\mu(v) \in T_{F(\mu)} W(N)$ for all $v \in T_\mu W(M)$.*

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

Lemma 3.9. *Let $g : M \rightarrow M$ be a Riemannian isometry, i.e. $g^*h = h$, and $v = \nabla\varphi \in T_\mu W(M)$, $\varphi \in \mathcal{C}_c^\infty(M)$. Then for every $\mu \in W(M)$, $dF_\mu(v) = dg(v) = \nabla(\varphi \circ g^{-1}) \in T_{F(\mu)} W(M)$.*

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

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

□

Since we know from Theorem 3.6 that $dg_{\#\mu}$ is bounded and therefore continuous for every $\mu \in W(M)$, we can infer the following more general statement.

Corollary 3.10. *Let $g : M \rightarrow M$ be a Riemannian isometry and $T_\mu W(M) \ni v = \lim_{n \rightarrow \infty} \nabla\varphi_n$. Then $dg(v) = \lim_{n \rightarrow \infty} \nabla(\varphi_n \circ g^{-1}) \in T_{F(\mu)}W(M)$.*

Example 3.11. A simple example for an isometry on \mathbb{R}^3 is $f(x_1, x_2, x_3) = (x_1, x_3, x_2)$. For a vector field $v(x) = (v^1(x), v^2(x), v^3(x))$ it is then, for every $\mu \in W(\mathbb{R}^3)$, $dF_\mu(v) = dg(v) = (v^1(g(x)), v^3(g(x)), v^2(g(x)))$. Furthermore, we can compute directly that $\nabla(\varphi \circ g^{-1}) = (\partial_{x_1}\varphi, \partial_{x_3}\varphi, \partial_{x_2}\varphi) = dg(\nabla\varphi)$.

On \mathbb{R}^3 , a vector field v is conservative, i.e. the gradient of a scalar function, if and only if its curl $\nabla \times v$ is zero. Recall that this can be reformulated in terms of 1-forms: On a general Riemannian manifold (M, h) , there is a duality between vector fields v and 1-forms v^\flat by the formula $v^\flat(\cdot) := h(v, \cdot)$. The curl of a vector field and the exterior differential d^1 of a 1-form on a three-dimensional manifold are related by the formula $\text{curl} = \sharp \star d^1 \flat$, where \sharp is the inverse procedure to \flat and \star the Hodge star operator. In particular, $\text{curl} v = 0$ if and only if $d^1 v^\flat = 0$. For the Euclidean space \mathbb{R}^3 this implies $\nabla \times v = 0$ if and only if $d^1 v^\flat = 0$.

However, in our situation we are more generally interested in determining when v is an L^2 -limit of conservative vector fields. So instead of looking at the de Rham complex, we want to make use of the so called L^2 -de Rham complex on \mathbb{R}^3 (see for example [Lüc02] for details on the following).

To be able to do this, let $\Omega_c^p(\mathbb{R}^3)$ be the set of smooth p -forms on \mathbb{R}^3 with compact support. On this set we consider the L^2 -inner product

$$\langle \omega, \xi \rangle_{L^2(p, h)} := \int_{\mathbb{R}^3} \omega \wedge \star \xi.$$

This inner product depends on the choice of the Riemannian tensor h on \mathbb{R}^3 . Let $L^2\Omega^p(\mathbb{R}^3, h)$ denote the Hilbert space completion of $\Omega_c^p(\mathbb{R}^3)$.

Now let us look at the following linear operators from the L^2 -de Rham complex on \mathbb{R}^3 :

$$\begin{aligned} d^0 : \Omega_c^0(\mathbb{R}^3) &\longrightarrow L^2\Omega^1(\mathbb{R}^3, h) \\ d^1 : \Omega_c^1(\mathbb{R}^3) &\longrightarrow L^2\Omega^2(\mathbb{R}^3, h) \end{aligned}$$

and their respective minimal closures \bar{d}^0 and \bar{d}^1 . It is $\text{im}(\bar{d}^0) \subset \text{ker}(\bar{d}^1)$. And since $\text{ker}(\bar{d}^1)$ is closed (being the kernel of a closed operator), also $\overline{\text{im}(\bar{d}^0)}^{L^2(1, h)} \subset \text{ker}(\bar{d}^1)$.

The principal idea is then that finding an element ω in the domain of \bar{d}^1 (an element which is smooth, L^2 and $\bar{d}^1(\omega)$ is again L^2) which is not in its kernel, should correspond to finding an element which is not in some tangent space $T_{F(\mu)}W(\mathbb{R}^3)$. However, the L^2 -structures in $T_{F(\mu)}W(\mathbb{R}^3) = \overline{\{\nabla\varphi\}}^{L^2((\mathbb{R}^3, h), F(\mu))}$ and in $L^2\Omega^p(\mathbb{R}^3, h)$ (and thus, in particular in $\overline{im(d^0)}^{L^2(1, h)}$), or rather in its dual with respect to h , are a priori not related. In both cases, for the scalar product two vector fields are first inserted into the Riemannian tensor h , but whereas in the second case, this is integrated with respect to the metric volume form, in the first case, this is integrated with respect to a probability measure. To adapt to this, we formulate the following lemma.

Lemma 3.12. *Let η be the Euclidean metric tensor on \mathbb{R}^3 , $\nu : \mathbb{R}^3 \rightarrow \mathbb{R}_{>0}$ a positive function on \mathbb{R}^3 , $h := \nu(x)^2\eta$ a new Riemannian tensor and $d\nu = \nu(x)^2dvol_\eta$, where $dvol_\eta$ denotes the measure induced by the metric volume form vol_η . Then*

$$\overline{\{\nabla^\eta\varphi \mid \varphi \in C_c^\infty\}}^{L^2((\mathbb{R}^3, \eta), \nu)} \cong \overline{\{\nabla^h\varphi \mid \varphi \in C_c^\infty\}}^{L^2((\mathbb{R}^3, h), vol_h)}$$

as Hilbert spaces. Here, ∇^η and ∇^h denote the gradients that are taken with respect to the metrics η and h , respectively. In particular,

$$w = \lim_{n \rightarrow \infty} \nabla^\eta \varphi_n \text{ in } L^2((\mathbb{R}^3, \eta), \nu) \Leftrightarrow \frac{1}{\nu^2} w = \lim_{n \rightarrow \infty} \nabla^h \varphi_n \text{ in } L^2((\mathbb{R}^3, h), dvol_h).$$

Proof. Let us first point out that $\nabla^\eta\varphi = \nu^2\nabla^h\varphi$ for all differentiable functions φ , so that the isomorphism as vector spaces is naturally given by $\nabla^\eta\varphi \mapsto \frac{1}{\nu^2}\nabla^\eta\varphi$. It remains to check whether the scalar product is preserved. For this we note that $dvol_h = \nu^3dvol_\eta$ (see for example [Bes87] for validation). Then

$$\begin{aligned} \int_{\mathbb{R}^3} h(\nabla^h\varphi, \nabla^h\psi) dvol_h &= \int_{\mathbb{R}^3} \nu^2 h(\nabla^h\varphi, \nabla^h\psi) \nu dvol_\eta \\ &= \int_{\mathbb{R}^3} \nu^4 \eta(\nabla^h\varphi, \nabla^h\psi) d\nu = \int_{\mathbb{R}^3} \eta(\nabla^\eta\varphi, \nabla^\eta\psi) d\nu. \end{aligned}$$

Now, this isometry can continuously be extended to the closure of $\{\nabla^\eta\varphi\}$. \square

So, finding a vector field which is not a member of $\overline{\{\nabla^h\varphi\}}^{L^2((\mathbb{R}^3, h), vol_h)}$, corresponds to finding a vector field which is not a member of $\overline{\{\nabla^\eta\varphi\}}^{L^2((\mathbb{R}^3, \eta), \nu)}$. Based on this, to find elements that are not L^2 -limits of gradients, we look at the L^2 -de Rham complex on a Riemannian manifold of the form $(\mathbb{R}^3, h = \nu^2\eta)$.

Lemma 3.13.

$$w = \lim_{n \rightarrow \infty} \nabla^h \varphi_n \text{ in } L^2((\mathbb{R}^3, h), dvol_h) \iff w^\flat = \lim_{n \rightarrow \infty} d\varphi_n \text{ in } L^2\Omega^1(\mathbb{R}^3, h),$$

where $\varphi_n \in C_c^\infty(\mathbb{R}^3)$ and the duality between w and w^\flat is with respect to $h = \nu^2\eta$.

Proof. This follows from

$$\int_{\mathbb{R}^3} h(v, w) \, d\text{vol}_h = \int_{\mathbb{R}^3} w^\flat \wedge \star v^\flat.$$

(Consult for example [Fec06] for calculations on this.) \square

We conclude the following: If $\nabla \times w \neq 0$ holds true for $w := df(v)$, $v = \nabla\varphi$, $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^3)$ and f a smooth map, we know that also $dw^\flat \neq 0$, where w^\flat is the dual 1-form of w with respect to the Euclidean metric. But we need to know whether the dual 1-form of $\frac{1}{\nu^2}w$ with respect to h is in the kernel of d or not. Luckily, both forms coincide, i.e. $h(\frac{1}{\nu^2}w, \cdot) = \frac{\nu^2}{\nu^2}\eta(w, \cdot) = w^\flat(\cdot)$. So if w^\flat is smooth, L^2 and $d^{-1}(\omega)$ is again L^2 , w^\flat is in the domain of d^{-1} , but not in its kernel, which means that it cannot be an element of $\overline{\text{im}(d^0)}^{L^2(1, h)}$. With Lemma 3.13, then, $\frac{1}{\nu^2}w$ is not a member of $\overline{\{\nabla^h\varphi\}}^{L^2((\mathbb{R}^3, h), \text{vol}_h)}$ and Lemma 3.12 tells us that w is not part of $T_\nu W(\mathbb{R}^3)$, in case ν is a probability measure.

We want to apply this consideration in the following corollary.

Corollary 3.14. *Consider the map $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, $(x, y, z) \mapsto (x, y, x)$. Then there is a function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^3)$ such that $df(\nabla\varphi)$ is not an $L^2(\nu)$ -limit of gradients, where ν is of the form $d\nu = \nu^2 d\lambda$ and $\nu : \mathbb{R}^3 \rightarrow \mathbb{R}_{>0}$ such that $\int_{\mathbb{R}^3} 1 d\nu < \infty$.*

Proof. Let us choose $\varphi(x) := \exp\left(\frac{1}{\|x\|^2 - 1}\right) \in \mathcal{C}_c^\infty(\mathbb{R}^3)$ for all $\|x\| < 1$, $x = (x, y, z)$, and $\varphi(x) := 0$ otherwise. Then $\nabla \times df(\nabla\varphi) \neq 0$. For $v = \nabla\varphi$ it is

$$\begin{aligned} df(v)(f(x)) &= (v^1(f(x)), v^2(f(x)), v^3(f(x))) \\ &= (\varphi_{,1}(x, y, x), \varphi_{,2}(x, y, x), \varphi_{,1}(x, y, x)). \end{aligned}$$

Here, $\varphi_{,i}$ denotes the partial derivative of φ with respect to the i -th component. For validation of our statement, we only need to compute the third component of the curl of $df(v)$:

$$\begin{aligned} (\nabla \times df(v))^3(f(x)) &= df(v)_{,1}^2 - df(v)_{,2}^1 = v^2(f(x))_{,1} - v^1(f(x))_{,2} \\ &= (\varphi_{,2} \circ f)_{,1} - (\varphi_{,1} \circ f)_{,2} \\ &= 4xy \frac{4x^2 + 2y^2 - 1}{(2x^2 + y^2 - 1)^4} \exp \frac{1}{2x^2 + y^2 - 1}. \end{aligned}$$

This last expression is unequal to zero for example for $x = y = z = 1/2$. The final conclusion follows from the remarks above. \square

The counterexample we have found in the proof of Corollary 3.14 is not a counterexample for $\mu = \delta$, as we have seen in Lemma 3.8, since all vector fields are regarded as elements of an L^2 -space. However, as soon as, for example, $\text{supp}(\varphi) \subset \text{supp}(\mu)$, $dF_\mu(\nabla\varphi) \neq \nabla\tilde{\varphi}$.

3.3 Differentiable maps between Wasserstein spaces

Since the map in Corollary 3.14 is smooth, $\sup_{x \in M} \|df_x\| < \infty$, but, however, not proper, it is unclear whether in the situation of Theorem 3.3 $dF_\mu(v) \in T_{F(\mu)}W(N)$ is always achieved. But for a differential of F we need this always to be true. Luckily, $L^2(\nu) = T_\nu W(N) \oplus T_\nu^\perp W(N)$ for every $\nu \in W(N)$ and to fix our issue, we simply compose dF with a projection onto $T_{F(\mu)}W(N)$.

Definition 3.15. We call P^μ the orthogonal linear projection

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

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

Proposition 3.16. *For every a.c. couple (μ_t, v_t) , $(\mu_t, P^{\mu_t}(v_t))$ is a tangent couple.*

Proof. Let (μ_t, v_t) be an a.c. couple, then, for $v_t = v_t^\top + v_t^\perp$ whenever possible, it is

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

Thus, (μ_t, v_t^\top) is an a.c. couple. With Lemma 2.53 it is a tangent couple. \square

This means that, even if dF_μ does not hit $T_{F(\mu)}W(M)$, at least $P^{F(\mu)} \circ dF_\mu : T_\mu W(M) \rightarrow T_{F(\mu)}W(M)$ is a linear and bounded map between $T_\mu W(M)$ and $T_{F(\mu)}W(M)$ and for every tangent couple (μ_t, v_t) , $(F(\mu_t), (P^{F(\mu_t)} \circ dF_{\mu_t})(v_t))$ is a tangent couple, too.

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

Definition 3.17 (Differentiable map between Wasserstein spaces). An absolutely continuous map $F : W(M) \rightarrow W(N)$ is called *differentiable* in case for every $\mu \in W(M)$ there exists a bounded linear map $dF_\mu : T_\mu W(M) \rightarrow T_{F(\mu)}W(N)$ such that for every tangent couple (μ_t, v_t) the image curve $dF_{\mu_t}(v_t)$, where well-defined, is a tangent vector field of $F(\mu_t)$. In this way a bundle map⁴ $dF : TW(M) \rightarrow TW(N)$ is defined which we want to call the *differential* of F .

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

Remark 3.18. The reader might be surprised that we only give a global definition of differentiability, without having started with a pointwise definition. The latter is difficult, if at all possible, since the tangent vector fields v_t are only defined for a.e. $t \in [0, 1]$, so a pointwise evaluation of these is not well-defined. The situation would change if one would be able to speak about *continuous* curves of tangent vector fields, but it doesn't seem to be so easy to make this notion

⁴In our sense of the word "bundle".

precise. For differing t, t' the vector fields v_t and $v_{t'}$ are elements of different tangent spaces which is why the usual notion of continuity cannot be trivially applied.

Remark 3.19. For the sake of Definition 3.17 it would be enough to just demand “enough” dF_μ such that the given condition is fulfilled. However, later in Section 7.1, we need a map dF_μ at every point $\mu \in W(M)$, to ensure the meaningfulness of the definition of the pullback of the formal Riemannian metric tensor.

Note again that $dF_{\mu_t}(v_t)$ is only well-defined almost everywhere, since v_t is. But this is not harmful to our definition since in particular also the tangent vectors of $F(\mu_t)$ are only well-defined almost everywhere. But in this same manner, Definition 3.17 does not guarantee uniqueness of dF in a strict sense. (Here we mean that $dF = \widetilde{dF}$ whenever $dF_\mu(v) = \widetilde{dF}_\mu(v)$ for all $(\mu, v) \in TW(M)$.) But, after all, one can say that dF is unique up to a “negligible” set.

Definition 3.20 (Negligible set). A subset $Z \subset TW(M)$ is called *negligible* whenever for every tangent couple (μ_t, v_t) the set $\{t \in (0, 1) \mid (\mu_t, v_t) \in Z\}$ is of Lebesgue measure zero.

Remark 3.21. This definition respects the $L^1(dt)$ -nature of the v_t 's in the sense that changing any v_t on a set of measure zero does not change the measure of the set $\{t \in (0, 1) \mid (\mu_t, v_t) \in Z\}$.

Proposition 3.22 (Uniqueness of the differential). *The differential dF of a differentiable map $F : W(M) \rightarrow W(N)$ is unique up to a redefinition on a negligible set $Z \subset TW(M)$.*

Proof. Let dF and \widetilde{dF} be two pointwise linear bundle maps, dF being the differential of an a.c. map F . It is to show that dF and \widetilde{dF} are both a differential of F if and only if $\{(\mu, v) \in TW(M) \mid dF_\mu(v) \neq \widetilde{dF}_\mu(v)\}$ is negligible.

Let dF and \widetilde{dF} be different only on a negligible set. In this case, for each tangent couple (μ_t, v_t) the image velocities $dF_{\mu_t}(v_t)$ are different from the ones of $\widetilde{dF}_{\mu_t}(v_t)$ only on a null set and thus still equal the tangent vector fields along $F(\mu_t)$ almost everywhere. Let on the other hand dF and \widetilde{dF} both fulfill the conditions of Definition 3.17. By definition, for each tangent couple (μ_t, v_t) both $dF_{\mu_t}(v_t)$ and $\widetilde{dF}_{\mu_t}(v_t)$ are equal almost everywhere to the tangent vectors along $F(\mu_t)$. Thus, for every tangent couple (μ_t, v_t) , $\{t \in (0, 1) \mid dF_{\mu_t}(v_t) \neq \widetilde{dF}_{\mu_t}(v_t)\}$ has Lebesgue measure zero. \square

Proposition 3.23. 1.) $T_\mu(W(M)) \setminus \{0\}$ is negligible. But $T_\mu(W(M))$ isn't.

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

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

4.) The following is an equivalence relation on the set of mappings between tangent bundles on Wasserstein spaces:

$$F \sim G := \{(\mu, v) \in TW(M) \mid F(\mu, v) \neq G(\mu, v)\} \text{ is negligible.}$$

Remark 3.24. Let dF be a differential of a map $F : W(M) \rightarrow W(N)$. Then there are members of its equivalence class $[dF]$ which are not a differential of F since since not every member has to be linear. Restricting, however, the equivalence relation onto the subset of pointwise linear maps between tangent bundles of Wasserstein spaces solves this issue. In this case $[dF]$ contains precisely all the possible differentials of F . Whenever we refer to a representative of dF , we mean an element of the latter equivalence class. Later, in Chapter 7, we will mostly fix one distinguished representative to be able to perform well-defined operations.

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

$T_\mu(W(M))$ cannot be negligible since $\mu_t = \mu$ is absolutely continuous with metric derivative 0.

- 2.) This follows from the fact that any countable union of sets of measure zero again is of measure zero.
- 3.) Let N be a subset of a negligible set and (μ_t, v_t) an a.c. curve with a fixed representative v_t . The amount of times where $(\mu_t, v_t) \in N$ can only be a subset of a set of zero measure. Since the Lebesgue measure is a complete measure this subset itself is measurable and in particular of measure zero.
- 4.) This follows from 1.) and 2.)

□

Corollary 3.25. 1.) In case $F = f_\#$ and f is as in Theorem 3.3, F is differentiable with $dF_\mu = P^{F(\mu)} \circ \hat{dF}_\mu$, where $P^{F(\mu)}$ is the orthogonal projection onto $T_{F(\mu)}N$ from Proposition 3.16 and

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

as in formula (3.1). In case f is a Riemannian isometry, the additional projection P is not necessary, as we have seen in Corollary 3.10. Then, $dF_\mu = df$ for all $\mu \in W(M)$.

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

- 3.) Let $F : W(M) \rightarrow W(N)$ and $G : W(N) \rightarrow W(O)$ be two differentiable maps. Then also $G \circ F : W(M) \rightarrow W(O)$ is differentiable with $d(G \circ F)_\mu(v) = (dG_{F(\mu)} \circ dF_\mu)(v)$ up to a negligible set.
- 4.) Whenever F is differentiable, bijective with differentiable inverse F^{-1} , then dF is also invertible with inverse $d(F^{-1})$, up to a negligible set.

Proof. 1.) This follows from Theorem 3.3 and Proposition 3.16.

2.) This is immediate.

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

4.) According to above it is up to a negligible set,

$$(d(F^{-1})_{F(\mu)} \circ dF_\mu)(v) = d(F^{-1} \circ F)_\mu(v) = d(id)_\mu(v) = v$$

and the other way round. □

Remark 3.26. Let us emphasize that this type of differentiability is highly tailored to the structure given by Optimal transport. It knowingly does not fit into the framework of [KM97]. Nevertheless, let us mention that also in this reference, the notion of differentiable maps between infinite dimensional manifolds is established via the property that differentiable curves should be mapped onto differentiable curves.

In the examples, we so far have only been concerned with maps $F : W(M) \rightarrow W(N)$ which are induced by maps $f : M \rightarrow N$. Now one could wonder how a map F which is not of this type could look like and what its differentiability properties are. The following lemma gives a simple criterion to decide when a map F is not a pushforward of measures.

Lemma 3.27. *Whenever there is an $f : M \rightarrow N$ such that $F = f_\#$, for every $x \in M$ it is $F(\delta_x) = \delta_{f(x)}$.*

Proof. Let $x \in M$ and $A \subset N$. Then $f_\#(\delta_x)(A) = \delta_x(f^{-1}(A)) = \delta_{f(x)}(A)$. The last equality sign is true since whenever $x \in f^{-1}(A)$ it is $f(x) \in A$ and other way round. □

Example 3.28. • If $F(\mu) = \mu_0$ is a constant map such that $\mu_0 \neq \delta_{y_0}$, $y_0 \in N$, then there exists no map $f : M \rightarrow N$ such that $F = f_{\#}$. In case $F(\mu) = \delta_{y_0}$, it is $F = f_{\#}$ with $f(x) = y_0 \forall x \in M$.

- Let $F_i : W(M) \rightarrow W(N)$, $i = 1, 2$, such that they do not coincide on $\{\delta_x \mid x \in M\}$. The mixing of measures $F := (1 - \lambda)F_1 + \lambda F_2$ for $0 < \lambda < 1$, then, cannot be a pushforward of measures.

Remark 3.29. Another way to think about this issue is the following: Every map $F : W(M) \rightarrow W(N)$ has a decomposition into a map $\tilde{F} : W(M) \rightarrow \mathcal{P}(M \times N)$ with $\pi_{\#}^1 \tilde{F}(\mu) = \mu$ and the map $\pi_{\#}^2 : \mathcal{P}(M \times N) \rightarrow W(N)$, i.e. $F = \pi_{\#}^2 \circ \tilde{F}$. Certainly, \tilde{F} is not unique, but one can always choose $\tilde{F}(\mu) = \mu \otimes F(\mu)$. (Other way round, of course, $r_{\#}^2 \circ \tilde{F}$ is always a map from $W(M)$ to $W(N)$.) Thus, F is a pushforward with respect to a map f if and only if there exists a map \tilde{F} in such a way that $\tilde{F}(\mu) = (Id, f)_{\#} \mu$. According to [AG09], Lemma 1.20 this is equivalent to saying that for every $\tilde{F}(\mu)$ there exists a $\tilde{F}(\mu)$ -measurable set $\Gamma \subset M \times N$ on which $\tilde{F}(\mu)$ is concentrated such that for μ -a.e. x there exists only one $y = f(x) \in M$ with $(x, y) \in \Gamma$. And in this case, $\tilde{F}(\mu) = (Id, f)_{\#} \mu$.

Let us now try to find out whether the maps given in Example 3.28 are differentiable.

Lemma 3.30. *Any constant map $F : W(M) \rightarrow W(N)$, $\mu \mapsto \mu_0$ is differentiable with $dF = 0$ up to a negligible set.*

Proof. Let (μ_t, v_t) be any tangent couple on M , then $(F(\mu_t), dF_{\mu_t}(v_t)) = (\mu_0, 0)$ fulfills the continuity equation:

$$\begin{aligned} & \int_0^1 \int_N \frac{\partial}{\partial t} \varphi(x, t) + h(\nabla \varphi(x, t), 0) d\mu_0 dt = \int_0^1 \int_N \frac{\partial}{\partial t} \varphi(x, t) d\mu_0 dt \\ & = \int_0^1 \frac{\partial}{\partial t} \int_N \varphi(x, t) d\mu_0 dt = \int_N \varphi(x, 1) d\mu_0 - \int_N \varphi(x, 0) d\mu_0 = 0. \end{aligned}$$

□

Lemma 3.31. *Let $F_i : W(M) \rightarrow W(N)$, $i = 1, 2$, be arbitrary a.c. maps. Then, for $0 \leq \lambda \leq 1$, also $F := (1 - \lambda) F_1 + \lambda F_2$ is a.c.*

Proof. Let μ_t be an a.c. curve. Then by definition $F_i(\mu_t)$, $i = 1, 2$, are a.c. curves. From Proposition 2.50 we now know that also $F(\mu_t)$ is an a.c. curve. □

The question now is, of course, whether there are conditions under which $F = (1 - \lambda) F_1 + \lambda F_2$ is also differentiable. Since F is a.c., given both F_i are a.c., we know that F maps a.c. curves to a.c. curves. We know that along each of these a.c. image curves there have to be tangent vector fields. But we cannot know, a priori, whether there exists a linear bundle map, mapping curves of tangent vector fields to corresponding curves of tangent vector fields. Finding an expression for a general image tangent vector field would be helpful

for finding such a bundle map. But this seems to be possible, at least without greater effort, only in very special cases.

Corollary 3.32. *Let F_1, F_2 be a.c. and $F := (1 - \lambda) F_1 + \lambda F_2$, $0 < \lambda < 1$. If an a.c. curve μ_t is such that $\text{supp}(F_1(\mu_t)) \cap \text{supp}(F_2(\mu_t)) = \emptyset$ for almost every t , the tangent vector along $F(\mu_t)$ is $v_t^1 \oplus v_t^2$, where v_t^i is the tangent vector field along $F_i(\mu_t)$.*

Proof. This follows from Lemma 3.31 and Corollary 2.56. \square

Corollary 3.33. *In case the F_i are differentiable and F is such that for each a.c. curve μ_t , $\text{supp}(F_1(\mu_t)) \cap \text{supp}(F_2(\mu_t)) = \emptyset$ for almost all t , then F is differentiable with $dF_\mu(v) = (dF_1)_\mu(v) \oplus (dF_2)_\mu(v)$.* \square

A simple example for this kind of map would be a map with $F_i(\mu) = \mu_i = \text{const.}$, where $\text{supp}(\mu_1) \cap \text{supp}(\mu_2) = \emptyset$. But since this is a constant map, we already know from Lemma 3.30 that F is differentiable with $dF_\mu(v) = 0$. A less trivial example is the following.

Example 3.34. Let $f_i : M \rightarrow M_i$, $i = 1, 2$, be as in Theorem 3.3. Both maps induce a pushforward of measures from M to $M_1 \dot{\cup} M_2$: $F_i := (f_i)_\# : W(M) \rightarrow W(M_1 \dot{\cup} M_2)$ where $\text{supp}(F_i(\mu)) \subset M_i$ for all $\mu \in W(M)$. Then, according to Corollary 3.33, $F = (1 - \lambda) F_1 + \lambda F_2$ is differentiable with $dF_\mu(v) = (dF_1)_\mu(v) \oplus (dF_2)_\mu(v)$. Here, the $(dF_i)_\mu(v)$ are as in formula (3.1). \square

In all other cases, things get complicated. A direct computation shows the following Lemma.

Lemma 3.35. *Let μ_t be an a.c. curve which is the convex combination of two other a.c. curves: $\mu_t = (1 - \lambda) \mu_t^1 + \lambda \mu_t^2$, $0 < \lambda < 1$. Let v_t^1 and v_t^2 be the corresponding tangent vector fields. Then the velocity vector fields of μ_t , v_t , are the solutions of the following differential equation:*

$$(3.0) \quad -\nabla \left((1 - \lambda) \mu_t^1 v_t^1 + \lambda \mu_t^2 v_t^2 \right) = \nabla (\mu_t v_t).$$

Corollary 3.36. *In case $M = \mathbb{R}$ and $\mu_t^i = \rho^i(x, t) d\lambda$, define the vector field v by*

$$v = -\frac{1}{(1 - \lambda)\rho^1 + \lambda\rho^2} \left((1 - \lambda) \rho^1 v_t^1 + \lambda \rho^2 v_t^2 \right)$$

at points x where $\rho^i(x) \neq 0$ for both $i = 1, 2$ and

$$v^+ = v_t^1 \oplus v_t^2$$

otherwise. Then v is together with $\mu_t = (1 - \lambda) \mu_t^1 + \lambda \mu_t^2$ a solution of equation (3.0).

Proof. In one dimension, equation (3.0) reduces to

$$-\partial_x \left((1 - \lambda) v^1 \rho^1 + \lambda v^2 \rho^2 \right) = \partial_x \left(((1 - \lambda) \rho^1 + \lambda \rho^2) v \right).$$

This means we have

$$-((1 - \lambda) v^1 \rho^1 + \lambda v^2 \rho^2 + C) = ((1 - \lambda) \rho^1 + \lambda \rho^2) v,$$

which is equivalent to

$$v = -\frac{1}{(1 - \lambda)\rho^1 + \lambda\rho^2} ((1 - \lambda) \rho^1 v_t^1 + \lambda \rho^2 v_t^2 + C).$$

□

Remark 3.37. What we can see from the preceding examination is that the notion of differentiable maps that we've built on top of the formal differentiable structure on $W(M)$ is not compatible with the convex structure of $W(M)$ other than in special cases ($F = (1 - \lambda) F_1 + \lambda F_2$ still being absolutely continuous, which is not clear for maps F with the property $F(\lambda\mu + (1 - \lambda)\nu) = \lambda F(\mu) + (1 - \lambda)F(\nu)$, though). This is different from considering $\mathcal{P}(M)$ as a convex subspace of the vector space of signed measures, where one can apply convenient calculus [KM97]. So pushing the formal calculus further in the manner that we have, one can see how it deviates from what we assume calculus has to provide, even though it matches some other formulas pretty well.

Chapter 4

Quantum dynamics on $W_2(M)$

This chapter aims at investigating a special solution of the free Schrödinger equation with the tools of optimal transport. To set the physical background, we begin with a briefing on quantum physics, on the aspects that interest us most.

4.1 Elements of quantum dynamics

For our account on quantum physics, we extensively rely on [\[BBCK92\]](#) and [\[Sta01\]](#). Of course, there is a lot more to be said about the history of quantum mechanics, its non-classical properties, its interpretation and philosophy and also about its mathematical formulations. But naturally, this is beyond our scope and already treated extensively in many textbooks.

On quantum physics

Quantum physics has started to be developed in the beginning of the 20th century to explain processes in the atomic realm which could not be explained by classical theories. Nowadays, regarding the range of explicable experiments and the precision of predictions, quantum physics is considered to be one of the best physical theories.

The upshot of the early developments of quantum physics was that both light and matter can have particle- as well as wave-like properties. The quantitative correspondence between these both aspects of light has been formulated by Albert Einstein in 1905, the one of matter has been proposed by Louis De Broglie in his doctoral thesis in 1924. But also other quantities of the atomic world have been found to be quantized, such as the spatial orientation of angular momentum, as demonstrated by Otto Stern and Walther Gerlach in 1921, and the energy electrons can have and emit in an atom, as explained by Niels Bohr

in 1913. There is a constant of nature characterizing all quantized systems and this is Planck’s constant h , discovered by Max Planck in 1900.

As a consequence of the wave-particle duality of matter, Werner Heisenberg concluded that it is principally impossible to simultaneously perform precise measurements on position and momentum of a particle, which is expressed in his famous uncertainty relation $\Delta x \Delta p \geq \hbar/2$, where $\hbar = h/2\pi$. This also explains why Newton’s law is not applicable in the quantum realm: the exact knowledge of position and momentum at an initial time is mandatory there. But more generally, the comprehension grew that the extent to which an observation affects a phenomenon cannot be minimized arbitrarily by rearranging the experimental setup, so that the role of the observer and of measurement is distinguished in quantum physics.

Related to this is a property of quantum theory that makes it possible for us to apply the theory of optimal transport: predictions regarding outcomes of measurements can only be made in a probabilistic way. In particular, and most important for us, the predicted location of a particle is given by a Lebesgue-probability density on \mathbb{R}^3 .

Wave mechanics

The basic equation of quantum physics is the *Schrödinger equation*, which was introduced by Erwin Schrödinger in 1926 in a series of four papers, all entitled with “Quantisierung als Eigenwertproblem” (engl. “Quantization as an eigenvalue problem”) ([Sch26a],[Sch26b],[Sch26c],[Sch26d]). And although the mathematical apparatus describing quantum features has developed enormously since, we will concern ourselves only with *wave mechanics*, i.e. “that portion of quantum theory that is based on the Schrödinger wave equation” ([BBCK92]).

The Schrödinger equation for a particle with mass m in an external potential $V(x, t)$ is given by

$$(4.1) \quad i\hbar \frac{\partial}{\partial t} \psi(x, t) = \left(-\frac{\hbar^2}{2m} \Delta + V(x, t) \right) \psi(x, t),$$

where Δ denotes the Laplace operator with respect to the x -variable. It is a partial differential equation for functions ψ on $\mathbb{R}^3 \times \mathbb{R}$ and its solutions are called *wave functions* and considered to represent the state of the particle.

Let us for now assume that the potential V does not depend on t and that the wave function has the form $\psi(x, t) = \alpha(t)\psi(x)$. From equation (4.1) we can then extract separate equations for both $\alpha(t)$ and $\psi(x)$:

$$(4.2) \quad i\hbar \frac{\partial}{\partial t} \alpha(t) = E\alpha(t)$$

$$(4.3) \quad \left(-\frac{\hbar^2}{2m} \Delta + V(x) \right) \psi(x) = E\psi(x),$$

where E is a constant.

Both, (4.2) and (4.3), are equations for eigenfunctions of the differential operators $i\hbar \frac{\partial}{\partial t}$ and $-\frac{\hbar^2}{2m}\Delta + V(x)$, respectively. Whereas the general solution of (4.2) is quick to state, $\alpha(t) = Ae^{-iEt/\hbar}$, $A \in \mathbb{R}$, it is more involved to solve equation (4.3), the so called *time-independent Schrödinger equation*. We will, however, later be only interested in the case $V = 0$, i.e. in the case of free particles. Due to the linearity of (4.1), its general solutions are superpositions of solutions of the form $\psi(x, t) = \alpha(t)\psi(x)$.

One can find that $-\frac{\hbar^2}{2m}\Delta$ plays the quantum role of the kinetic energy $\mathbf{p}^2/2m$, so the operator $H := -\frac{\hbar^2}{2m}\Delta + V(x)$ is in its structure analogous to the classical Hamilton function and is therefore called (*quantum*) *Hamiltonian*. Its respective eigenvalues E are thereby interpreted as the energy of the corresponding particle and application of H is thus seen as a measurement of the energy of the system whose state is represented by ψ . More generally, it was discovered that it makes sense to relate the application of (self-adjoint) operators on ψ and measurements on the system described by ψ with each other, where the corresponding eigenvalues of the operator are the possible outcomes of the measurement. After the performance of such a measurement, the wave function is regarded to necessarily *collapse* into a state which is an eigenfunction of the corresponding operator.

Probabilistic interpretation of the wave function

In 1926 Max Born proposed to interpret $\psi(x, t)$ in a probabilistic way ([Bor26a]): the squared modulus $\rho(x, t) := |\psi(x, t)|^2$ of the wave function should describe the probability density of finding the particle at point x at time t . The condition $\int \rho(x, t) d\lambda(x) = 1$ should thereby ensure that the probability of finding the particle somewhere in space is one. Born further argued (in [Bor26b]) that if ψ is written as a linear combination of orthonormal functions ψ_n , i.e. $\psi = \sum_n c_n \psi_n$, the squared modulus of the coefficients $|c_n|^2$ can be interpreted as the probabilities for finding the system in the state ψ_n . In particular, if ψ_n is an eigenfunction of a self-adjoint operator A , $|c_n|^2$ is the probability of measuring the eigenvalue corresponding to ψ_n . Along these lines, $\int \bar{\psi} A \psi d\lambda$ can be interpreted as the expectation value of the outcome of a measurement of A , where $\bar{\psi}$ denotes the complex conjugation of ψ .

The probability density $\rho(x, t)$ satisfies the continuity equation, i.e. a local law of conservation of probability, together with the *current density of probability* $j(x, t) := \frac{\hbar}{2mi} (\bar{\psi}(x, t)\nabla\psi(x, t) - \nabla\bar{\psi}(x, t)\psi(x, t))$:

$$(4.4) \quad \frac{d}{dt}\rho(x, t) + \nabla \cdot j(x, t) = 0.$$

(Compare Section 2.4, in particular Remark 2.49.)

Writing the wave function in polar form, i.e. $\psi(x, t) = R(x, t)e^{iS(x, t)/\hbar}$, where $R(x, t) = |\psi(x, t)| = \sqrt{\rho(x, t)}$ and $S(x, t)$ a real function (the *phase* of

the wave function), yields

$$j(x, t) = \rho(x, t) \frac{1}{m} \nabla S(x, t).$$

Substituting the polar form in (4.1), one obtains, after some calculus, two equations: The continuity equation for ρ and $j = \rho \frac{1}{m} \nabla S$ as in (4.4) and a modified Hamilton-Jacobi equation with generating function S ,

$$(4.5) \quad \frac{\partial}{\partial t} S(x, t) + \frac{1}{2m} \|\nabla S(x, t)\|^2 + V(x) = \frac{\hbar^2}{2m} \frac{\Delta \sqrt{\rho(x, t)}}{\sqrt{\rho(x, t)}}.$$

Equation (4.5) differs from a Hamilton-Jacobi equation only in the term

$$Q(x, t) := -\frac{\hbar^2}{2m} \frac{\Delta \sqrt{\rho(x, t)}}{\sqrt{\rho(x, t)}},$$

which is called the *quantum potential* and which vanishes in the classical limit $\hbar \rightarrow 0$.

The continuity equation (4.4) together with the quantum Hamilton-Jacobi equation (4.5) were first derived from the Schrödinger equation by Erwin Madelung in 1926 ([Mad27]) and are therefore called *Madelung equations*. It is important to note that, however, not every solution of the Madelung equations yields a solution of the Schrödinger equation. To read more about this, we refer, for example, to [Wal94].

If we compare equation (4.5) for $V = 0$ with equation (2.9), we see that the Madelung equations (4.4) and (4.5) can be considered as the quantum version of the equations for optimal transport (we will mention this analogy again Section 4.3).

In Section 2.4 we have seen that the continuity equation characterizes absolutely continuous curves in $W_2(M)$ (Theorem 2.47). Thus, if it would hold for $d\mu_t := \rho(x, t)d\lambda(x)$ that $\int \|x\|^2 d\mu_t < \infty$, i.e. if $\mu_t \in W(\mathbb{R}^3)$ for all t , and $\|\nabla S(x, t)\|_{L^2(\mu_t)} \in L^1(0, \infty)$, then we would know that the curve of probability measures defined by the Schrödinger equation is absolutely continuous in $W(\mathbb{R}^3)$. We will elaborate on that below.

Motion in expectation and the spreading of wave packets

The expected motion of quantum particles behaves in a surprisingly classical manner. Let $\langle x(t) \rangle$ be the expected position and $\langle v(t) \rangle$ the expected velocity of the quantum particle at time t :

$$\begin{aligned} \langle x(t) \rangle &= \int_{\mathbb{R}^3} x \, d\mu_t, \\ \langle v(t) \rangle &= \frac{1}{m} \int_{\mathbb{R}^3} \nabla S(x, t) \, d\mu_t. \end{aligned}$$

Then, for a free particle it is $\frac{d}{dt}\langle v(t) \rangle = 0$ (so $\langle v(t) \rangle \equiv \langle v \rangle$ from now on) and

$$\langle x(t) \rangle = \langle v \rangle t + \langle x(t=0) \rangle.$$

Thus, interestingly, the expected velocity $\langle v(t) \rangle$ does not depend on time and in particular it holds $\frac{d}{dt}\langle x(t) \rangle = \langle v \rangle$.

More generally, for a particle under the influence of the force $F(x, t) = -\nabla V(x, t)$ it holds

$$(4.6) \quad \frac{d}{dt}\langle x(t) \rangle = \frac{1}{m}\langle p(t) \rangle \quad \text{and} \quad \frac{d}{dt}\langle p(t) \rangle = \langle F(x, t) \rangle,$$

where

$$\langle p(t) \rangle = \frac{\hbar}{i} \int_{\mathbb{R}^3} \bar{\psi}(x, t) \nabla \psi(x, t) d\lambda(x)$$

and

$$\langle F(x, t) \rangle = - \int \bar{\psi}(x, t) (\nabla V(x, t)) \psi(x, t) d\lambda(x).$$

Note that $\frac{d}{dt}\langle p(t) \rangle = F(\langle x(t) \rangle, t)$ does not hold true, unless the force is linearly dependent on position, as in the harmonic oscillator or a constant force.

Equations (4.6) are called Ehrenfest equations and can be generalized to arbitrary self-adjoint operators.

But not only the time-development of the expectation value of the position of a free particle is interesting, also its variance $\sigma_x^2(t)$ is, where

$$\sigma_x^2(t) := \int_{\mathbb{R}^3} (x - \langle x(t) \rangle)^2 d\mu_t.$$

Defining

$$\sigma_v^2 := \int_{\mathbb{R}^3} \left\| \left(\frac{\hbar}{im} \nabla - \langle v \rangle \right) \psi(x, t) \right\|^2 d\lambda(x),$$

which does not depend on time, the following formula holds true:

$$(4.7) \quad \sigma_x^2(t) = \sigma_v^2 \cdot (t - t_1)^2 + \sigma_x^2(t_1).$$

Hereby is t_1 the instant in time at which $\sigma_x^2(t)$ attains its minimum value. Equation (4.7) describes the spreading of the wave function of a free particle: starting at t_1 the variance $\sigma_x^2(t)$ grows monotonously by the term $\sigma_v^2 \cdot (t - t_1)^2$. A consequence of the spreading of the wave packet is a growing decrease of information about the location of the particle.

The Gaussian wave packet

It is possible to give a general solution to the initial value problem of the free Schrödinger equation. Is $\psi(x', 0)$ the wave function at initial time $t = 0$, then

$$(4.8) \quad \psi(x, t) = \int_{\mathbb{R}^3} K(x - x', t) \psi(x', 0) d\lambda(x'),$$

where the integration kernel K is given by

$$K(x - x', t) = \left(\frac{m}{i\hbar t}\right)^{3/2} \exp\left(i\frac{m}{2\hbar t}(x - x')^2\right).$$

K obeys the initial condition $K(x - x', t = 0) = \delta(x - x')$ and is a regular solution of the free Schrödinger equation for $t \neq 0$. It can be interpreted as an elementary wave emitted at $t = 0$ from the point x' . The mathematically simplest solution of the Schrödinger equation is a monochromatic plane wave $e^{\frac{i}{\hbar}(px - p^2 t/2m)}$, which is however not a physical solution because it is not normalizable.

The simplest initial condition $\psi(x, 0)$ for the calculation of the integral in equation (4.8) is a *Gaussian function*. The most general form of this function is

$$(4.9) \quad \psi(x, 0) = N \exp\left(-\frac{1}{2}(x - C)M(x - C)\right),$$

where $M \in \mathbb{C}^{3 \times 3}$ is a complex 3×3 matrix, $C \in \mathbb{C}^3$ and N a normalization constant.

Substituting (4.9) into (4.8) and performing an integration yields

$$\psi(x, t) = N \left(\det\left(\mathbf{1} + \frac{i\hbar t}{m}M\right)\right)^{-1/2} \exp\left(-\frac{1}{2}(x - C)\left(M^{-1} + \frac{i\hbar t}{m}\mathbf{1}\right)^{-1}(x - C)\right),$$

where $\mathbf{1}$ is the identity element in $\mathbb{C}^{3 \times 3}$. The simplest case is the spherical case, where $C = 0$ and $M_{ij} = \delta_{ij}l^{-2}$. Then,

$$(4.10) \quad \psi(x, t) = N \left(1 + \frac{i\hbar t}{ml^2}\right)^{-3/2} \exp\left(-\frac{1}{2}\left(l^2 + \frac{i\hbar t}{m}\right)^{-1} \|x\|^2\right).$$

The probability density in that case is

$$(4.11) \quad \rho(x, t) = |N|^2 \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2} \exp\left(-\left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1} \|x\|^2\right).$$

4.2 Optimal transport for a solution of the free Schrödinger equation

Let us now investigate solutions of the free Schrödinger equation of the form (4.10) with the tools of Optimal transport. Therefore, let $\rho(x, t)$ be as in (4.11) and μ_t be defined by $d\mu_t = \rho(x, t)d\lambda(x)$. We will see that with the coefficient $N^2 := (l^2\pi)^{-3/2}$, μ_t is a probability measure. The following questions then arise naturally:

- 1) Is μ_t an absolutely continuous curve?

- 2) Is the gradient of the phase of $\psi(x, t)$ an element of $T_{\mu_t} W(\mathbb{R}^3)$?
- 3) How does the optimal transport map from μ_s to μ_t look like?
- 4) What is the cost for transporting optimally from μ_s to μ_t ? I.e. what is the Wasserstein distance between μ_s and μ_t ?

In this section we will tackle these questions one after the other in the Wasserstein space $W(\mathbb{R}^3) = (\mathcal{P}_2(\mathbb{R}^3), W(\cdot, \cdot))$, where

$$W(\mu, \nu) = \inf_{\pi \in \text{Adm}(\mu, \nu)} \int_{\mathbb{R}^3 \times \mathbb{R}^3} \|x - y\|^2 d\pi(x, y),$$

i.e. the cost function is $d^2(x, y) = \|x - y\|^2$.

Lemma 4.1. *If $N^2 = (l^2\pi)^{-3/2}$, μ_t is a probability measure for every $t \in \mathbb{R}_{>0}$. I.e.*

$$\mu_t = (l^2\pi)^{-3/2} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2} \exp\left(-\left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1} \|x\|^2\right) d\lambda(x) \in \mathcal{P}(\mathbb{R}^3).$$

Proof. Define $Q := \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2}$ and $A := 2\left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1}$. Then, according to [Zha],

$$\begin{aligned} & Q \int_{\mathbb{R}^3} \exp\left(-\frac{1}{2} A \|x\|^2\right) d\lambda(x) = Q \sqrt{\frac{(2\pi)^3}{A^3}} \\ &= \pi^{3/2} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2} \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{3/2} \\ &= \pi^{3/2} \left(\frac{1}{l^2} \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)\right)^{-3/2} \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{3/2} \\ &= (l^2 \pi)^{3/2}. \end{aligned}$$

□

Lemma 4.2. $\mu_t \in W(\mathbb{R}^3)$ for every $t \in \mathbb{R}_{\geq 0}$.

Proof. Integrals of the form $\int_{\mathbb{R}^3} \|x\|^2 \exp(-a\|x\|^2) d\lambda(x)$ are finite, so the integral $\int_{\mathbb{R}^3} \|x\|^2 d\mu_t$ is finite, too. □

Let us continue with the question whether μ_t is an absolutely continuous curve in $W(\mathbb{R}^3)$. In Section 4.1, we have already seen that the continuity equation is satisfied by $(\mu_t, \nabla S(x, t))$, in the strong sense (equation (4.4)), where S is the phase of the wave function, which means that it also holds in the weak sense. According to Theorem 2.47, we have proven absolute continuity of μ_t , as soon as we have shown $\|\nabla S(x, t)\|_{L^2(\mu_t)} \in L^1(0, \infty)$.

Remark 4.3. The condition in Theorem 2.47 is $\|v_t\|_{L^2(\mu_t)} \in L^1(0, 1)$, since μ_t was defined for $t \in [0, 1]$. However, here we have to adapt to a curve on $[0, \infty)$.

Lemma 4.4. *The gradient of the phase of the wave function (4.10) is given by*

$$\nabla S(x, t) = \frac{\hbar^2 t}{m} \left(l^4 + \frac{\hbar^2 t^2}{m^2} \right)^{-1} x.$$

Instead of calculating ∇S directly, after extracting S from ψ , we want to use the formula $j(x, t) = \rho(x, t) \frac{1}{m} \nabla S(x, t)$ to find an expression for ∇S .

Proof. Let us calculate $j(x, t) = \frac{\hbar}{2mi} (\bar{\psi} \nabla \psi - \nabla \bar{\psi} \psi)$. We use the following abbreviations: $A := \left(1 + \frac{i\hbar t}{m l^2}\right)^{-3/2}$ and $B := \left(l^2 + \frac{i\hbar t}{m}\right)^{-1}$, so that $\psi(x, t) = NA \exp\left(-\frac{1}{2}B\|x\|^2\right)$. Then

$$\begin{aligned} \nabla \psi(x, t) &= -BAN \exp\left(-\frac{1}{2}B\|x\|^2\right) x, \\ \nabla \bar{\psi}(x, t) &= -\bar{B}\bar{A}N \exp\left(-\frac{1}{2}B\|x\|^2\right) x, \\ \bar{\psi} \nabla \psi &= -B|A|^2 N^2 \exp\left(-\frac{1}{2}(B + \bar{B})\|x\|^2\right) x \\ \nabla \bar{\psi} \psi &= -\bar{B}|A|^2 N^2 \exp\left(-\frac{1}{2}(B + \bar{B})\|x\|^2\right) x. \end{aligned}$$

Therefore,

$$\frac{\hbar}{2mi} (\bar{\psi} \nabla \psi - \nabla \bar{\psi} \psi) = -\frac{\hbar}{2mi} N^2 |A|^2 (B - \bar{B}) \exp\left(-\frac{1}{2}(B + \bar{B})\|x\|^2\right) x.$$

Furthermore,

$$\begin{aligned} |A|^2 &= \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2} \\ B + \bar{B} &= 2l^2 \left(l^4 + \frac{\hbar^2 t^2}{m^2}\right)^{-1} = 2 \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1} \\ B - \bar{B} &= -2i \frac{\hbar t}{m} \left(l^4 + \frac{\hbar^2 t^2}{m^2}\right)^{-1}. \end{aligned}$$

So that, because $\rho(x, t) = N^2 |A|^2 \exp\left(-\frac{1}{2}(B + \bar{B})\|x\|^2\right)$,

$$j(x, t) = -\frac{\hbar}{2mi} \rho(x, t) (B - \bar{B}) x.$$

It follows that

$$\begin{aligned} \nabla S(x, t) &= -\frac{\hbar}{2i} (B - \bar{B}) x \\ &= \frac{\hbar^2 t}{m} \left(l^4 + \frac{\hbar^2 t^2}{m^2}\right)^{-1} x. \end{aligned}$$

□

Proposition 4.5.

$$\|\nabla S(x, t)\|_{L^2(\mu_t)} = \sqrt{\frac{3}{2}} \frac{\hbar^2 t}{m l^6} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-2}.$$

Proof. Define $C := \left(\frac{\hbar^2 t}{m} \left(l^4 + \frac{\hbar^2 t^2}{m^2}\right)^{-1}\right)$, $B := (l^2 \pi)^{-3/2} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2}$ and $A := 2 \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1}$. Then, again with [Zha],

$$\begin{aligned} \|\nabla S(x, t)\|_{L^2(\mu_t)}^2 &= \int_{\mathbb{R}^3} \|\nabla S(x, t)\|_{\mathbb{R}^3}^2 d\mu_t(x) = C^2 \int_{\mathbb{R}^3} \|x\|^2 d\mu_t(x) \\ &= C^2 \int_{\mathbb{R}^3} \|x\|^2 B \exp\left(-\frac{1}{2} A \|x\|^2\right) d\lambda(x) \\ &= C^2 B \sqrt{\frac{(2\pi)^3}{2^3 \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-3}}} \frac{3}{2} \frac{2}{2} \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right) \\ &= \frac{3}{2} \left(\frac{\hbar^2 t^2}{m l^4}\right)^2 \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-7/2} l^{-3} \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1/2} \\ &= \frac{3}{2 l^4} \left(\frac{\hbar^2 t^2}{m l^4}\right)^2 \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-4}. \end{aligned}$$

□

Corollary 4.6.

$$\int_0^\infty \|\nabla S(x, t)\|_{L^2(\mu_t)} dt = \sqrt{\frac{3}{2}} \frac{m}{2 l^2} < \infty.$$

This expression is in particular independent of \hbar .

Proof. Define $a := \sqrt{\frac{3}{2}} \frac{\hbar^2}{m l^6}$ and $b := \frac{\hbar^2}{m^2 l^4}$. Then

$$\begin{aligned} \int_0^\infty \|\nabla S(x, t)\|_{L^2(\mu_t)} dt &= \int_0^\infty \frac{at}{(1 + bt^2)^2} dt \\ &= \frac{a}{2b}. \end{aligned}$$

□

Corollary 4.7. *The curve μ_t is absolutely continuous up to a redefinition on a null set in time.* □

Let us now approach the question of how the probability distributions of the location of the particle at several times can be transported optimally onto each other. A natural transport map seems already to be given, namely the (time-dependent) flow of $\frac{1}{m} \nabla S$. But is it also optimal? In any case, it is worth

finding out, how much it costs when the probabilities are transported with this flow and whether there is a map that transports cheaper than the map that is naturally given by the Schrödinger equation.

We first recall the definition of the flow of time-dependent vector fields.

Definition 4.8. Let $V(x, t)$ be a time-dependent vector field on \mathbb{R}^3 , i.e. $V : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$. An *integral curve* of V is given by a function $\varphi(x_0, t_0, \cdot) : \mathbb{R} \rightarrow \mathbb{R}^3$, $t \mapsto \varphi(x_0, t_0, t)$ which satisfies

$$\dot{\varphi}(x_0, t_0, t) = V(\varphi(x_0, t_0, t), t) \quad \text{and} \quad \varphi(x_0, t_0, t_0) = x_0,$$

where the dot denotes the differentiation with respect to t . A *flow map* of V is then given by

$$F_s : \mathbb{R}^3 \times \mathbb{R} \longrightarrow \mathbb{R}^3, \quad (x, t) \mapsto \varphi(x, s, t).$$

The map $F : (\mathbb{R}^3 \times \mathbb{R}) \times \mathbb{R} \rightarrow (\mathbb{R}^3 \times \mathbb{R})$, $(x, s, t) \mapsto (\varphi(x, s, t), s + t)$ is a flow map in the time-independent sense, i.e. it satisfies the group law in the last variable.

Lemma 4.9. *The (time-dependent) flow of the vector field $\frac{1}{m}\nabla S(x, t)$ is given by*

$$F_s(x, t) = \sqrt{\frac{1 + \frac{\hbar^2}{m^2 l^4} t^2}{1 + \frac{\hbar^2}{m^2 l^4} s^2}} x.$$

Proof. Define $V(x, t) := \frac{1}{m}\nabla S(x, t)$. Functions of the form

$$\varphi(x_0, t_0, t) = \exp\left(\frac{1}{2} \ln\left(l^4 + \frac{\hbar^2 t^2}{m^2}\right) + C\right) x$$

are solutions of $\dot{\varphi}(x_0, t_0, t) = V(\varphi(x_0, t_0, t), t)$. Additionally, the initial condition $\varphi(x_0, t_0, t_0) = x_0$ holds, if the constant C is chosen to be $C = -\frac{1}{2} \ln\left(l^4 + \frac{\hbar^2 t_0^2}{m^2}\right)$. Application of calculation rules for the logarithm yields formula (4.9). \square

Corollary 4.10. *For every pair of times s and t , $\mu_t = F_s(\cdot, t)_\# \mu_s$.*

We could infer this statement from Theorem 2.42. However, for the sake of directness, we will perform the explicit calculation. For this, we will need the following lemma which we give without proof.

Lemma 4.11. *Let the measure μ be defined by $d\mu = \rho(x)d\lambda(x)$ and let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a bijective measurable map. Then, for a measurable set A ,*

$$T_\# \mu(A) = \int_A (\rho \circ T^{-1})(x) dT_\# \lambda(x).$$

In particular, in case $T \in ISO(\mathbb{R}^3)$, the density of $T_\# \mu$ with respect to the Lebesgue measure is given by $(\rho \circ T^{-1})(x)$.

Proof of Corollary 4.10. First we see that

$$F_s(\cdot, t) \# \lambda = \left(\frac{1 + \frac{\hbar^2}{m^2 l^4} t^2}{1 + \frac{\hbar^2}{m^2 l^4} s^2} \right)^{-3/2} \lambda.$$

Furthermore,

$$\begin{aligned} & \rho(F_s^{-1}(x, t), s) \\ &= N^2 \left(1 + \frac{\hbar^2}{m^2 l^4} s^2 \right)^{-3/2} \exp \left(- \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2} \right)^{-1} \frac{1 + \frac{\hbar^2 s^2}{m^2 l^4}}{1 + \frac{\hbar^2 t^2}{m^2 l^4}} \|x\|^2 \right) \\ &= N^2 \left(1 + \frac{\hbar^2}{m^2 l^4} s^2 \right)^{-3/2} \exp \left(-l^{-2} \left(1 + \frac{\hbar^2 s^2}{m^2 l^4} \right)^{-1} \frac{1 + \frac{\hbar^2 s^2}{m^2 l^4}}{1 + \frac{\hbar^2 t^2}{m^2 l^4}} \|x\|^2 \right) \\ &= N^2 \left(1 + \frac{\hbar^2}{m^2 l^4} s^2 \right)^{-3/2} \exp \left(- \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2} \right)^{-1} \|x\|^2 \right) \\ &= \rho(x, t). \end{aligned}$$

So in total, using Lemma 4.11, $\mu_t = F_s(\cdot, t) \# \mu_s$. \square

Theorem 4.12. *The map $F_s(\cdot, t)$ gives the optimal transport from μ_s to μ_t for each pair $t, s \in [0, \infty)$.*

Remark 4.13. In the proof of Theorem 4.12 we are going to use the cost function $d^2/2$, as it allows us to use Proposition 2.9, which is very convenient. However, T is an optimal map with respect to the cost function $d^2/2$ if and only if it is an optimal map with respect to d^2 : Let $W^{d^2/2}(\cdot, \cdot)$ be the Wasserstein distance with respect to $d^2/2$ and $W^{d^2}(\cdot, \cdot)$ the one with respect to d^2 , then if we assume T to be optimal with respect to $d^2/2$,

$$\begin{aligned} W^{d^2/2}(\mu, \nu)^2 &= \int_X \frac{d^2(x, T(x))}{2} d\mu(x) = \inf_{\pi \in \text{Adm}(\mu, \nu)} \int_{X \times X} \frac{d^2(x, y)}{2} d\pi(x, y) \\ &= \frac{1}{2} \inf_{\pi \in \text{Adm}(\mu, \nu)} \int_{X \times X} d^2(x, y) d\pi(x, y) \\ &= \frac{1}{2} W^{d^2}(\mu, \nu)^2. \end{aligned}$$

So $\int_X d^2(x, T(x)) d\mu(x) = W^{d^2}(\mu, \nu)^2$ which means that T is optimal with respect to d^2 , too. Similarly the other way round. In particular, μ_t is a geodesic with respect to $W^{d^2/2}$ if and only if it is a geodesic with respect to W^{d^2} . Whenever we write $W(\cdot, \cdot)$, i.e. whenever we omit the superscript, we refer to the cost function d^2 .

Proof of Theorem 4.12. Define $C := \sqrt{\frac{1 + \frac{\hbar^2}{m^2 l^4} t^2}{1 + \frac{\hbar^2}{m^2 l^4} s^2}}$ and $\bar{\varphi}(x) := \frac{1}{2} C \|x\|^2$. (In the notation of C and φ we neglect the fact that they depend on the parameters s

and t .) Then $F_s(x, t) = Cx$ and

$$F_s(x, t) = \nabla \bar{\varphi}(x).$$

This, in particular, means that $F_s(x, t) \in \partial^- \bar{\varphi}(x)$, i.e. $F_s(x, t)$ is an element of the subdifferential of $\bar{\varphi}$ at point x . If we further define $\varphi(x) := \frac{1-C}{2} \|x\|^2$, we see that $\bar{\varphi}(x) = \frac{\|x\|^2}{2} - \varphi(x)$. From Proposition 2.9 we can conclude that φ is c -concave and that $F_s(x, t) \in \partial^{c+} \varphi(x)$. So, if it were that condition (2.2) holds for μ_s and $\mu_t = F_s(\cdot, t) \# \mu_s$, we could infer from Remark 2.11 that $F_s(\cdot, t)$ is optimal from μ_s to μ_t . Regarding condition (2.2), we first see that $\frac{\|x-y\|^2}{2} \leq \frac{\|x\|^2}{2} + \frac{\|y\|^2}{2}$. Since integrals of the type $\int_{\mathbb{R}^3} \|x\|^2 \exp(-B\|x\|^2) d\lambda(x)$, $B \in \mathbb{R}_{>0}$, are finite, we know that $\frac{\|x\|^2}{2} \in L^1(\mu_{s,t})$. \square

The finding of the optimal transport maps enables us to determine whether $\frac{1}{m} \nabla S(x, t)$ is a family of tangent vector fields along μ_t .

Corollary 4.14. *It is $\frac{1}{m} \nabla S(x, t) \in T_{\mu_t} W(\mathbb{R}^3)$ for all $t \in [0, \infty)$.*

Proof. We are going to use formula (2.11). Since from $\|v_n - v\|_{\mathbb{R}^3}^2 \rightarrow 0$ it follows $\|v_n - v\|_{L^2(\mu_t)} \rightarrow 0$, we will only determine the limit in \mathbb{R}^3 . Let $c := \frac{\hbar^2}{m^2 l^4}$, then

$$\lim_{s \rightarrow t} \frac{F_t(x, s) - x}{s - t} = \lim_{s \rightarrow t} \frac{\sqrt{\frac{1+cs^2}{1+ct^2}} x - x}{t - s} = \lim_{s \rightarrow t} \frac{\sqrt{\frac{1+cs^2}{1+ct^2}} - 1}{t - s} x.$$

Calculating the following limit in \mathbb{R} yields:

$$\lim_{s \rightarrow t} \frac{\sqrt{\frac{1+cs^2}{1+ct^2}} - 1}{t - s} = \frac{cs}{cs^2 + 1} = \frac{\hbar^2 t}{m} \left(l^4 + \frac{\hbar^2 t^2}{m^2} \right)^{-1},$$

which is precisely the prefactor of $\frac{1}{m} \nabla S(x, t)$. From this we can infer the claimed statement. \square

From Theorem 2.47 we can infer the following corollary.

Corollary 4.15. *The metric derivative $|\dot{\mu}_t|$ of μ_t equals $\|\nabla S(x, t)\|_{L^2(\mu_t)}$, i.e.*

$$|\dot{\mu}_t| = \sqrt{\frac{3}{2}} \frac{\hbar^2 t}{m l^6} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4} \right)^{-2}.$$

\square

In particular, μ_t cannot be a constant speed geodesic, as its metric derivative is not constant. However we will see that it is still a geodesic in the sense of formula (2.4), i.e. in the sense of shortest paths.

We close this section with finally calculating the cost for the optimal transportation from μ_s to μ_t .

Theorem 4.16. For each $s, t \in [0, \infty)$,

$$W(\mu_s, \mu_t) = \sqrt{\frac{3}{2}} \left| \sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 t^2}{m^2 l^4}} \right|.$$

Proof. Let us again first use the cost function $d^2/2$ and let again $C := \sqrt{\frac{1 + \frac{\hbar^2}{m^2 l^4} t^2}{1 + \frac{\hbar^2}{m^2 l^4} s^2}}$ (dependent on the parameter s and t). Then

$$\begin{aligned} W^{d^2/2}(\mu_s, \mu_t)^2 &= \frac{1}{2} \int_{\mathbb{R}^3} \|x - F_s(x, t)\|^2 d\mu_s(x) \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \|x - Cx\|^2 d\mu_s(x) = \frac{1}{2} \int_{\mathbb{R}^3} (1 - C)^2 \|x\|^2 d\mu_s(x) \\ &= \frac{1}{2} \int_{\mathbb{R}^3} (1 - C)^2 |N|^2 \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right)^{-3/2} \exp\left(-\left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right)^{-1} \|x\|^2\right) d\lambda(x). \end{aligned}$$

Define now $Q := (1 - C)^2 |N|^2 \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right)^{-3/2}$ and $A := 2 \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right)^{-1}$. Then, with [Zha],

$$W^{d^2/2}(\mu_s, \mu_t)^2 = \sqrt{\frac{(2\pi)^3}{A^3}} \frac{3}{2} \frac{Q}{A} = \frac{3\sqrt{(2\pi)^3}}{2} \frac{Q}{A^{5/2}}.$$

Let us first determine the factor $\frac{Q}{A^{5/2}}$ separately.

$$\begin{aligned} \frac{Q}{A^{5/2}} &= \frac{(1 - C)^2 N^2 \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right)^{-3/2}}{2^{5/2} \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right)^{-5/2}} = \frac{(1 - C)^2 N^2}{2^{5/2}} \frac{\left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right)^{5/2}}{\left(\frac{1}{l^2} \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right)\right)^{3/2}} \\ &= \frac{(1 - C)^2 N^2}{2^{5/2}} l^3 \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right). \end{aligned}$$

Substituting C and N yields

$$\begin{aligned} \frac{Q}{A^{5/2}} &= \frac{l^3 (l^2 \pi)^{-3/2}}{2^{5/2}} \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right) \left(1 - \sqrt{\frac{1 + \frac{\hbar^2 t^2}{m^2 l^4}}{1 + \frac{\hbar^2 s^2}{m^2 l^4}}}\right)^2 \\ &= \frac{l^5}{(l^2 \pi)^{3/2} 2^{5/2}} \left(\sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}}\right)^2 \left(1 - \sqrt{\frac{1 + \frac{\hbar^2 t^2}{m^2 l^4}}{1 + \frac{\hbar^2 s^2}{m^2 l^4}}}\right)^2 \\ &= \frac{l^5}{l^3 (\pi)^{3/2} 2^{5/2}} \left(\sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 t^2}{m^2 l^4}}\right)^2. \end{aligned}$$

So in total,

$$\begin{aligned} \frac{3\sqrt{(2\pi)^3}}{2} \frac{Q}{A^{5/2}} &= \frac{3\sqrt{(2\pi)^3}}{2} \frac{l^2}{(\pi)^{3/2} 2^{5/2}} \left(\sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 t^2}{m^2 l^4}} \right)^2 \\ &= \frac{3}{4} l^2 \left(\sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 t^2}{m^2 l^4}} \right)^2. \end{aligned}$$

□

Corollary 4.17. For μ_t as before and $s, t, u \in \mathbb{R}_{\geq 0}$, $s < u < t$, we have

$$W(\mu_s, \mu_t) = W(\mu_s, \mu_u) + W(\mu_u, \mu_t),$$

for all $t \in \mathbb{R}_{\geq 0}$.

Proof.

$$\begin{aligned} W(\mu_s, \mu_u) + W(\mu_u, \mu_t) &= \sqrt{\frac{3}{2}} \left(\sqrt{1 + \frac{\hbar^2 u^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}} \right) + \\ &\quad \sqrt{\frac{3}{2}} \left(\sqrt{1 + \frac{\hbar^2 t^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 u^2}{m^2 l^4}} \right) = W(\mu_s, \mu_t). \end{aligned}$$

□

4.3 Madelung equations in terms of Wasserstein geometry

For completeness, we want to use this section to briefly mention the article [vR12] by M.-K. von Renesse. Following up a work by Lott [Lot07], Renesse formally used notions for higher order calculus on $W(\mathbb{R}^3)$ to be able to express the Madelung equations (4.4) and (4.5) in the following unified way¹:

$$(4.12) \quad \nabla_{v_t}^W v_t = -\nabla^{L^2(\mu_t)} \left(V(\mu_t) + \frac{\hbar^2}{8} I(\mu_t) \right),$$

where

$$V(\mu) := \int_{\mathbb{R}^3} V(x) d\mu,$$

with $V(x)$ the classical potential from the Schrödinger equation (4.1) and

$$I(\mu) := \int_{\mathbb{R}^3} \|\nabla \ln(\rho(x))\|^2 d\mu,$$

¹A thorough mathematical treatment of second order analysis on Wasserstein space has been carried out by Gigli in [Gig12].

the so called *Fisher information* for $d\mu = \rho(x)d\lambda$. Furthermore, the symbol $\nabla_{v_t}^W v_t$ means the formal covariant derivative in $W(\mathbb{R}^3)$ of v_t along v_t , which we are not going to explain in more detail. Finally, $\nabla^{L^2(\mu)}$ denotes the gradient with respect to the canonical $L^2(\mu)$ scalar product.

The most stunning aspect of this formula is obviously the classical coat of this equation: (4.12) has the form of Newton's second law of motion, where the potential of the force is given as the sum of the mean value of the classical potential $V(x)$ and the Fisher Information $I(\mu)$. It is also interesting to note that in particular, the Fisher information precisely corresponds to the quantum potential in formula (4.5) and therefore is the part in the equation that makes the dynamics "quantum". According to [vR12], taking the limit $\hbar \rightarrow 0$, solutions of the classical Newtonian equation $\gamma(t)$ solve (4.12) after applying the canonical isometric embedding from \mathbb{R}^3 to $W(\mathbb{R}^3)$, $\gamma_t \mapsto \delta_{\gamma_t}$.

To close this chapter, we calculate the value of the Fisher information along the curve μ_t studied in Section 4.2.

Corollary 4.18. *Let μ_t be as in (4.11) with $N^2 = (l^2\pi)^{-3/2}$. Then for every $t \in \mathbb{R}_{\geq 0}$,*

$$I(\mu_t) = \frac{6}{l^4} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-4} = 4 \left(\frac{ml^2}{\hbar^2 t}\right)^2 \|\nabla S(x, t)\|^2.$$

For $t = 0$ it is then, $I(\mu_0) = \frac{6}{l^4}$. In particular, $I(\mu_t)$ is monotonously decreasing.

That $I(\mu_t)$ is monotonously decreasing makes sense since μ_t is the squared norm of a wave packet that is spreading, which means that the information about the location of the particle is decreasing.

Let us also recall that $\|\nabla S(x, t)\|_{L^2(\mu_t)} = |\dot{\mu}_t|$, so that also $I(\mu_t) = 4 \frac{l^2 m}{\hbar^2 t} |\dot{\mu}_t|$.

Proof. Define $A := \left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1}$, $B := (l^2\pi)^{-3/2} \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2}$ and $C := \left(\frac{\hbar^2 t}{m} \left(l^4 + \frac{\hbar^2 t^2}{m^2}\right)^{-1}\right)$. Then $C = \frac{\hbar^2 t}{ml^2} A$ and

$$\begin{aligned} I(\mu_t) &= \int_{\mathbb{R}^3} \|\nabla \ln(\rho(x, t))\|^2 d\mu(x) \\ &= \int_{\mathbb{R}^3} \|\nabla (\ln B + \ln(\exp(-A\|x\|^2)))\|^2 d\mu(x) \\ &= \int_{\mathbb{R}^3} A^2 \|\nabla \|x\|^2\|^2 d\mu(x) = \int_{\mathbb{R}^3} A^2 \|2x\|^2 = 4A^2 \int_{\mathbb{R}^3} \|x\|^2 d\mu(x) \\ &= 4 \left(\frac{ml^2}{\hbar^2 t}\right)^2 C^2 \int_{\mathbb{R}^3} \|x\|^2 d\mu(x). \end{aligned}$$

Comparing the last expression with the calculation of the proof of Proposition 4.5, we get the claimed formula. \square

Chapter 5

Philosophical interlude

So far, we have introduced the mathematical and physical background for the inquiries in the two forthcoming chapters. However, the motivation for these investigations comes from a set of philosophical ideas on which we want to elaborate a bit in this interlude.

It is commonly believed that space and matter are concepts that correspond to things in the real world. However, in different theories of nature, on different levels of description, the perception to what precisely these concepts correspond and of their mutual dependence can differ significantly. But it seems that in a given theoretical framework the dependency of these concepts can only be in the following three ways:

- 1- Space and matter are distinct physical entities without mutual influence of any kind.
- 2- Space and matter are distinct physical entities that affect each other the one or the other way.
- 3- Space and matter are such that one of the two notions can completely be derived from the other.

Whereas possibilities one and two are already well-established in physical theories, we here want to follow up the third possible way. The mathematical inquiry of the forthcoming chapters is in particular regarded as a first attempt to investigate such an idea.

The first form of dependency is assumed in Newtonian mechanics as well as in Quantum mechanics, where (Euclidean) space serves as an absolute background structure for the motion of matter. Since space and matter do not influence each other in these theories, the physical entity that the concept of space refers to is principally inaccessible by experiment. In particular it is impossible to determine absolute positions. In [\[Mac83\]](#), Ernst Mach argued that notions like this should not be regarded as elements of reality and therefore should not be used in theories of nature:

”Über den absoluten Raum und die absolute Bewegung kann niemand etwas aussagen, sie sind bloße Gedankendinge, die in der Erfahrung nicht aufgezeigt werden können. Alle unsere Grundsätze der Mechanik sind, wie ausführlich gezeigt worden ist, Erfahrungen über relative Lagen und Bewegungen der Körper. Sie konnten und durften auf den Gebieten, auf welchen man sie heute als gültig betrachtet, nicht ohne Prüfung angenommen werden. Niemand ist berechtigt, diese Grundsätze über die Grenzen der Erfahrung hinaus auszudehnen. Ja diese Ausdehnung ist sogar sinnlos, da sie niemand anzuwenden wüßte.”

The second form of dependency is implemented in General relativity, where the distribution of matter determines the geometry of space-time and vice versa.¹

As mentioned above, in this treatise we want to go in the direction of the third form of dependency and assume that the concept of matter is more fundamental than the concept of space. Space, then, is regarded as an emergent element of reality. Or, paraphrasing this with the words of Julian Barbour [Bar01], ”The world does not contain things, it is things.” One way to realize such an assumption quantitatively is within a relational theory where space is what is spanned by all the particles.² A relational theory, in general, does not treat absolute notions, such as single or multiple-particle trajectories in absolute space, but it subjects properties of objects in relation to other objects, such as relative distances.³ This point of view has in particular been promoted by Ernst Mach, as noted above. However, other than performing calculations within the framework of classical mechanics and concluding that it is mathematically very convenient to start with absolute space and to have absolute quantities, Mach did not give any formal prescription of how to quantify his principle of relative dynamics. A more formal implementation of his ideas was presented by Julian Barbour and Bruno Bertotti in [BB82]. Their idea was that an n -particle configuration forms a *shape* which constitutes space. The evolution of these shapes is determined by a principle they call *best matching* - shapes that resemble one another most closely are *matched* together which generates a notion of time evolution measuring the change of the shapes. In the next chapters we are also going to talk about shapes and about a related idea to best matching which we will call *shape distance*. While being similar in the name and in spirit - indeed Barbour's Shape dynamics has sparked the idea for our notion of Shape space -

¹A further possible interpretation of General relativity could be that this theory does not speak about space-time and distribution of matter at the same time, but rather that space-time and matter are two possible levels of description which can be converted into each other with the help of the the Einstein-equations. This view then would correspond to the third mode of dependency.

²It is commonly said that Albert Einstein initially tried to make General relativity a relational theory, which he gave up eventually.

³Dropping the ontological requirement for space, another possible reading of Newtonian and Quantum mechanics is, to view the mathematical structure that served as space, i.e. \mathbb{R}^3 , merely as a coordinate system with respect to which the motion is described. In this sense, these theories, for one-particle, can be seen as effective relational theories.

these both approaches differ significantly, mathematically as well as physically. Whereas Shape dynamics in [BB82] is dealing with multi-particle classical dynamics, we are here concerned with one-particle quantum dynamics, since we additionally want to assume that matter fundamentally has to be described by quantum theory. There is ongoing research about quantizing Julian Barbour's Shape dynamics for classical systems which is, however, still different from our approach.⁴

The notion of shape that we are going to consider in the upcoming chapters is that of the shape of the wave function of a (non-relativistic) quantum mechanical one-particle system. We, a priori, explicitly leave open the question regarding a realism of the wave function or its probability distribution. Is the wave function denoted by $\psi(x, t)$, then $|\psi(x, t)|^2$ is the associated probability distribution and $\mu := |\psi(x, t)|^2 d\lambda(x)$ is a probability measure on \mathbb{R}^3 . Rotation and translation of μ affects its orientation and position on \mathbb{R}^3 , but it has no influence on the relative distribution of density. Accordingly, all those measures that can be attained from μ by some series of rotations and translations can be considered to be the same, in this sense. A shape will thus be any collection of measures that is constituted such that each two members can be transformed into each other by a series of rotations and translations and, conversely, any series of rotations and translations performed on a member of the collection yields another member.⁵ This concept of shape can straightforwardly be generalized to probability measures in Wasserstein space $W_p(X)$, where X is an arbitrary metric space, substituting \mathbb{R}^3 . Instead of rotations and translations one applies the action of the isometry group G of X . We will see that in terms of push-forwards of measures, G acts isometrically on $W_p(X)$, too. In abstract terms, a shape, then, is an element of the space $W_p(X)/G$, i.e. it is an equivalence class of measures where the equivalence relation is taken with respect to the action of G on $W_p(X)$. The Shape distance between two shapes will then be the smallest Wasserstein distance between members of each equivalence class that is possible to find, similarly to best matching. We will call $W_p(X)/G$ together with the Shape distance the Shape space.

The idea behind this definition of shape is that space as the absolute background structure, is, so to say, modded out. What remains of it mathematically, solely serves as an anchor for ψ and as an auxiliary means for the manifestation of shape, i.e. of the relative configuration. In particular, motion does not take place on (absolute) space anymore, motion is expressed in terms of change of the shape and shape is all there is.

Let us point out that in this treatise we are not aiming at working out how

⁴Shape dynamics has developed to be a gravitational theory, describing the evolution of the conformal geometry of space, being dynamically equivalent to the ADM formalism of General relativity. Here we are referring to Shape dynamics for classical systems as described in [BB82].

⁵This is equivalent to saying that a shape is a collection of measures that contains one measure μ such any transformation of μ in terms of rotations and translations is also an element of the collection. This definition, however, creates the illusion of the existence of one special element which is not the case. Indeed, in case there exists one such measure, every other measure in this collection has the same property.

any traditional notion of space can be deduced from our definition of shapes. Rather, as a first step, we want to show that it is possible to describe quantum motion without resorting to an absolute background structure.

One might wonder, why we want to regard equivalence classes of $\mu := |\psi|^2$ as shapes, rather than equivalence classes of ψ itself. It is true that the isometry group G of \mathbb{R}^3 also acts isometrically on L^2 by composition ($(g, \psi) \mapsto \psi \circ g^{-1}$). However, as we have seen in Chapter 4, the dynamics of μ is explicitly, and conveniently, driven by the Wasserstein distance (on the one hand, plus the Fisher information on the other). There is no such connection between the dynamics of ψ and the L^2 -metric.

Our notion of shape is meaningful only due to the extendedness of the wave function, or, as a consequence, due to the extendedness of its probability density. On the contrary, a classical particle at point x can be regarded as the Dirac measure δ_x at that point. Its shape consists of all the Dirac measures on \mathbb{R}^3 , which is why in particular, there cannot be a non-trivial dynamics for the particle on Shape space. Also, going back to the notion of shapes as relative positions of particles, it is a priori not clear what relative positions of quantum particles could be, due to their theoretical representation by functions, different from just points in space. Additionally it is questionable whether a meaningful definition could be used for a relational *dynamics*. Also, to derive from this a notion of physical space would not be possible in the manner that was used in the classical case.

Although our definition for shape is a priori only for one particle, mathematically it is straightforward to generalize this to the many-particle case. This is because other than in the classical scenario, where multiple particles each are tracked by their own trajectory, many-particles in Quantum dynamics are described by associating just one wave function to the whole collection of particles. However, while classical many-particles are collectively treated on the same (position) space, multiple quantum particles each roam alone in their own, configuration, space. So, generalizing our approach to many particles creates a different philosophical situation to start with, since position space and configuration space do not fall together anymore. (See for example [NA13] to read more about philosophical issues with the usage of $3N$ -dimensional configuration space to describe many-particle Quantum mechanics.) But there are also physical issues, raised by the application of, for example, the full isometry group of the $3N$ -dimensional product space, since this would lead to far off from the common picture of multi-particles. As a resolution, it might be appropriate to restrict to the action of a suitable subgroup of the full isometry group. For example the one consisting only of products $g_1 \times \dots \times g_N$ where each g_i is an isometry of \mathbb{R}^N , or subgroups that preserve other additional, physical, structures.

Since our notion of shape is not one that is based on relative positions of multiple particles, our considerations might not be relational at a first glance. But still, our shape comprises its relative distribution of density, which completely amounts for its motion as it is inscribed in Wasserstein geometry and due to the workings of the Fisher information. In this sense and due to the absence of absolute motion, our reasoning could indeed be regarded as relational. This,

however, should not be confused with the already established notion *Relational quantum mechanics*. The latter is an interpretation of Quantum mechanics, having been put forward in particular by Carlo Rovelli in [Rov96]. It proposes that the state of a system is not an absolute quantity, but instead observer-dependent. Regarding each measurement as a quantum interaction, the state that is measured depends on the relation of the system to the other system that is performing the measurement (i.e. the observer). Thus, measurements taken by different observers are assumed to lead to different outcomes.

Chapter 6

Shape space

In this whole chapter let (X, d) be a Polish metric space, G a subgroup of its isometry group $ISO(X)$ and $p \in \mathbb{N}_{\geq 1}$.

6.1 Shape distances $\mathcal{S}_p(X)$

Definition 6.1 (Isometry of metric spaces). An *isometry* between two metric spaces (X_1, d_1) and (X_2, d_2) is a surjective map $g : X_1 \rightarrow X_2$ such that $d_2(g(x), g(y)) = d_1(x, y) \forall x, y \in X_1$.

Whereas injectivity of an isometry follows from preservation of the distance, surjectivity has to be added by hand: Consider for example in the sequence space l^2 the map $l^2(\mathbb{R}) \rightarrow l^2(\mathbb{R}), (x_n)_n \mapsto (y_n)_n := (x_{n-1})_n$ with $y_1 = 0$ on the complete metric space $l^2(\mathbb{R})$. This map is clearly distance-preserving, but it is not surjective.

Definition 6.2 (Isometry group). The set of isometries on X forms a group with respect to composition of maps, the so called *isometry group* $ISO(X)$.

The group G acts in a natural way on X :

$$\begin{aligned} G \times X &\rightarrow X \\ (g, x) &\mapsto g(x). \end{aligned}$$

But it also acts naturally on $\mathcal{P}_p(X)$ by the pushforward of measures.

Definition 6.3. We define the action of G on $\mathcal{P}_p(X)$ in the following way:

$$\begin{aligned} G \times \mathcal{P}_p(X) &\rightarrow \mathcal{P}_p(X) \\ (g, \mu) &\mapsto g\mu := g\#\mu. \end{aligned}$$

It is quick to see that this definition indeed fulfills the right conditions to be an action. The following Lemma guarantees that the action is well-defined.

Lemma 6.4. Let $\mu \in \mathcal{P}_p(X)$ and $g \in G$, then $g\#\mu \in \mathcal{P}_p(X)$, too.

Proof. For $\mu \in \mathcal{P}_p(X)$ and $x'_0 \in X$ with $x_0 = gx'_0$ it is

$$\begin{aligned} \int_X d(x_0, x)^p d(g\mu)(x) &= \int_X d(x_0, gx)^p d\mu(x) = \int_X d(gx'_0, gx)^p d\mu(x) \\ &= \int_X d(x'_0, x)^p d\mu(x) < \infty. \end{aligned}$$

□

Definition 6.5 (Shape distance). The p -th Shape distance $D_p(\mu, \nu)$ of $\mu, \nu \in \mathcal{P}_p(X)$ is defined as

$$D_p(\mu, \nu) := \inf_{g \in G} W_p(g\mu, \nu).$$

In the following we will show that this *Shape distance* is a pseudometric on $\mathcal{P}_p(X)$ modulo G . If X is such that the closed balls are compact and the action of G on X is proper then $D_p(\cdot, \cdot)$ is indeed a metric, as we will also see below. The more general case where G is the invariance group of an arbitrary cost function, not necessarily a metric, would not lead us to a (pseudo) metric in general.

We recall the definition of a pseudometric.

Definition 6.6 (Pseudometric). A function $d : X \times X \rightarrow \mathbb{R}_{\geq 0}$ between a set X is called a *pseudometric* in case for all $x, y, z \in X$, $d(x, x) = 0$, $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$. Other than for a metric, it is not demanded that $d(x, y) = 0$ if and only if $x = y$.

Definition 6.7. Two elements in $\mathcal{P}_p(X)$ are defined to be *equivalent* if there exists an element in G which maps them into each other:

$$\mu \sim \nu \Leftrightarrow \exists g \in G : \mu = g\nu.$$

The quotient space with respect to this equivalence relation is denoted by $\mathcal{S}_p(X) := \mathcal{P}_p(X) / \sim$.

Definition 6.8 (Shape space). The space $\mathcal{S}_p(X)$, together with $D_p(\cdot, \cdot)$ is called *Shape space*. If $p = 2$, we will often just write $\mathcal{S}(X)$ instead of $\mathcal{S}_2(X)$

Remark 6.9. It is evident that Definition 6.7 indeed yields an equivalence relation on $\mathcal{P}_p(X)$. Also, we observe that the following map is well defined, where we denote the equivalence class of μ by $[\mu]$:

$$\begin{aligned} D_p(\cdot, \cdot) : \mathcal{S}_p(X) \times \mathcal{S}_p(X) &\longrightarrow \mathbb{R} \\ ([\mu], [\nu]) &\longmapsto D_p(\mu, \nu). \end{aligned}$$

For this reason, we always mean a map on $\mathcal{S}_p(X) \times \mathcal{S}_p(X)$ when we write expressions such as $D_p(\mu, \nu)$. We state it explicitly when we mean a map on $\mathcal{P}_p(X)$.

Theorem 6.10. *The p -th Shape distance $D_p(\cdot, \cdot)$ is a pseudometric on $\mathcal{S}_p(X)$.*

To prove this theorem we formulate the following two lemma:

Lemma 6.11. *For topological spaces X, Y let $f, g : X \rightarrow Y$ be injective measurable maps such that $f(X)$ and $g(X)$ are closed. Also, let $\widetilde{Adm}(f_{\#}\mu, g_{\#}\nu) := \{\Pi \in \mathcal{P}(f(X) \times g(X)) \mid \pi_{\#}^1 \Pi = f_{\#}\mu, \pi_{\#}^2 \Pi = g_{\#}\nu\}$. Then $(f \times g)_{\#}(\cdot)$ is a bijective map from $Adm(\mu, \nu)$ to $\widetilde{Adm}(f_{\#}\mu, g_{\#}\nu)$ with inverse map $(f^{-1} \times g^{-1})_{\#}(\cdot)$.*

Proof. First we need to see that the maps are well-defined. For this let $\Pi \in Adm(\mu, \nu)$. It is $\text{supp}((f \times g)_{\#}\Pi) \subset f(X) \times g(X)$, since the images of X under g and f are closed, which means that we can naturally identify $(f \times g)_{\#}\Pi$ as an element of $\mathcal{P}(f(X) \times g(X))$. Furthermore it is

$$(\pi_{\#}^1(f \times g)_{\#}\Pi)(A) = \Pi(f^{-1}(A) \times X) = \mu(f^{-1}(A)) = f_{\#}\mu(A),$$

for $A \in \mathcal{B}(f(X))$. A Similar calculation holds true for the projection onto the second component and for $(f_1^{-1} \times f_2^{-1})_{\#}(\cdot)$. The latter map is well-defined since every map is surjective on its image. For being inverse we observe that for $A, B \in \mathcal{B}(X)$

$$((f_1^{-1} \times f_2^{-1})_{\#}(f_1 \times f_2)_{\#}\Pi)(A, B) = ((f_1 \times f_2)_{\#}\Pi)(f_1(A) \times f_2(B)) = \Pi(A, B)$$

and the other way round. \square

Remark 6.12. For every $\Pi \in Adm(f_{\#}\mu, g_{\#}\nu)$ it is $\text{supp} \Pi \subset f(X) \times g(X)$ and in particular

$$\begin{aligned} W_p(f_{\#}\mu, g_{\#}\nu)^p &= \inf_{\Pi \in Adm(f_{\#}\mu, g_{\#}\nu)} \int_{Y \times Y} d^p(x, y) d\Pi(x, y) \\ &= \inf_{\Pi \in \widetilde{Adm}(f_{\#}\mu, g_{\#}\nu)} \int_{f(X) \times g(X)} d^p(x, y) d\Pi(x, y). \end{aligned}$$

Of course, $\widetilde{Adm}(f_{\#}\mu, g_{\#}\nu) = Adm(f_{\#}\mu, g_{\#}\nu)$ in case f and g are also surjective.

Lemma 6.13. *For all $g \in G$ it is*

$$W_p(g\mu, \nu) = W_p(\mu, g^{-1}\nu).$$

Proof. According to Lemma 6.11 the map $(g \times g)_{\#}(\cdot)$ gives a bijective correspondence of the sets $Adm(g\mu, \nu)$ and $Adm(\mu, g^{-1}\nu)$. We can use this to write

$$\begin{aligned} W_p(g\mu, \nu)^p &= \inf_{\Pi \in Adm(g\mu, \nu)} \int_X d(x, y)^p d\Pi(x, y) \\ &= \inf_{\bar{\Pi} \in Adm(\mu, g^{-1}\nu)} \int_X d(x, y)^p d[(g \times g)_{\#}\bar{\Pi}](x, y) \\ &= \inf_{\bar{\Pi} \in Adm(\mu, g^{-1}\nu)} \int_X d(x, y)^p d\bar{\Pi}(x, y) = W_p^p(\mu, g^{-1}\nu). \end{aligned}$$

\square

Corollary 6.14. From Lemma 6.13 it follows that G acts isometric also on the p -th Wassersteinspace: $W_p(g\mu, g\nu) = W_p(\mu, g^{-1}g\nu) = W_p(\mu, \nu)$. \square

Corollary 6.15. Let $G_{\#} := \{g_{\#} : W(M) \rightarrow W(M) \mid g \in G\}$. It is a subgroup of $ISO(W(M))$ and group-isomorphic to G .

Proof. $G_{\#} \subset ISO(W(M))$ follows from 6.14. Subgroup properties are immediate. The group homomorphism between G and $G_{\#}$ is given by $\varphi : G \rightarrow G_{\#}; g \mapsto g_{\#}$. It is surjective by definition and injective due to the following reasoning: Assume that $g_1, g_2 \in G$ are such that $g_{1\#} = g_{2\#}$. Then for every $x \in M$ we have

$$\delta_{g_1(x)} = g_{1\#}\delta_x = g_{2\#}\delta_x = \delta_{g_1(x)},$$

which means that $g_1(x) = g_2(x)$. \square

Proof of Theorem 6.10. Let $[\mu], [\nu]$ and $[\sigma]$ be elements of $\mathcal{S}_p(X)$. Of course $D_p(\mu, \nu) = 0$ if $[\mu] = [\nu]$ with optimal $g = 1_G$ and $D_p(\mu, \nu) \geq 0$ in any other case, by definition. Then we show that D is symmetric: Since W_p is symmetric and from Lemma 6.13 we know that

$$D_p(\mu, \nu) = \inf_{g \in G} W_p(g\mu, \nu) = \inf_{g \in G} W_p(\mu, g\nu) = \inf_{g \in G} W_p(g\nu, \mu) = D_p(\nu, \mu).$$

For the validity of the triangle inequality we see that for all $\sigma \in W_p(M)$ and all $g_1, g_2 \in G$,

$$\begin{aligned} D_p(\mu, \nu) &= \inf_{g \in G} W_p(g\mu, \nu) \leq W_p(g_2g_1\mu, \nu) = W_p(g_1\mu, g_2^{-1}\nu) \\ &\leq W_p(g_1\mu, \sigma) + W_p(\sigma, g_2^{-1}\nu) = W_p(g_1\mu, \sigma) + W_p(g_2\sigma, \nu). \end{aligned}$$

Therefore,

$$\begin{aligned} D_p(\mu, \nu) = \inf_{g \in G} W_p(g\mu, \nu) &\leq \inf_{g_1, g_2 \in G} (W_p(g_1\mu, \sigma) + W_p(g_2\sigma, \nu)) \\ &= \inf_{g_1 \in G} W_p(g_1\mu, \sigma) + \inf_{g_2 \in G} W_p(g_2\sigma, \nu) \\ &= D_p(\mu, \sigma) + D_p(\sigma, \nu). \end{aligned}$$

\square

Actually, we have proven something far more general. The only properties we have used for Theorem 6.10 is that the group action of g^{-1} behaves as a sort of adjoint action to the action of g with respect to the distance on the original metric space and that the formula for the prospective distance on the quotient space is given as an infimum over all possible distances elements of two orbits can have. We will formulate this finding in the following corollary.

Corollary 6.16. Let the metric space (X, d) and G be such that $d(gx, y) = d(x, g^{-1}y)$ for every $g \in G$ and $D(x, y) = \inf_{g \in G} d(gx, y)$. Then D is a pseudo-metric on X/G .

Remark 6.17. The condition $d(gx, y) = d(x, g^{-1}y)$ is necessary for D to be well-defined on X/G and lets us freely put the g at one of the two arguments.

Remark 6.18. Note that it is not clear under which conditions $D(\cdot, \cdot)$ would be symmetric if G would not be a group.

Remark 6.19. $D(\cdot, \cdot)$ is not a metric in this generality, because a minimizer for $\inf_{g \in G} d(\cdot, \cdot)$ need not exist. But as we will see, it is possible to show this minimizing property for proper actions.

Remark 6.20. One might wonder if one can bring $W_p(\mu, g\nu)$ and $W_p(\mu, \nu)$ in any algebraic relationship to each other, for every $\mu, \nu \in \mathcal{P}(X)$ and fixed g . However, a simple homogeneous relationship can not be established, as one can see at the example X being the real numbers, $p = 1$ and $gx := x + 2$. If $\mu = \delta_x$ and $\nu = \delta_y$, then $W(\mu, \nu) = d(x, y)$. Choose here for example $x = 3$ and $y = 0$, then $3 \cdot W(\mu, g\nu) = W(\mu, \nu)$. If, on the other hand, $x = 2$ then $0 \cdot W(\mu, \nu) = W(\mu, g\nu)$.

6.2 Existence of a minimizer for the Shape distance

The goal of this section is to prove the existence of a minimizer for the Shape distance-problem. The basic idea is that the problem should be reduced to having a continuous map on a compact subset of the isometry group and the minimizer of this map should be exactly the minimizer of the shape distance-problem. This is why we start with the following lemma and its important corollary.

Lemma 6.21. *Let G be a topological group with continuous group action on X . Then the group action of G on $W_p(X)$ is continuous as well.*

Proof. Let $((g_i), (\mu_i))_{i \in I}$ be a net in $G \times W_p(X)$ with $(g_i, \mu_i) \rightarrow (g, \mu)$ (see Appendix A.1.1). As we know from Section 2.1, the Wasserstein distance metrizes weak convergence, i.e. we have to show that $W_p(g_i\mu_i, g\mu) \rightarrow 0$ as $(g_i, \mu_i) \rightarrow (g, \mu)$. For this we see that

$$W_p(g_i\mu_i, g\mu) = W_p(\mu_i, g_i^{-1}g\mu) \leq W_p(\mu_i, \mu) + W_p(\mu, g_i^{-1}g\mu) \quad \forall i \in I.$$

We know already that $\mu_i \rightarrow \mu$. But it is also true that $g_i^{-1}g\mu \rightarrow \mu$: Since $g_i \rightarrow g$ in G and G acts continuously on X , the functions $g_i^{-1}g(x)$ converge pointwise against $Id_G(x)$. Using dominated convergence theorem yields $g_i^{-1}g\mu \rightarrow \mu$ and thus the continuity of the Wasserstein distance on $W_p(X)$. \square

Remark 6.22. In case $G \subset ISO(X)$, there is always at least one topology on G such that the action on X is continuous, namely the compact-open topology (see A.1.1 for a definition). Generally, the following holds: Is X a locally compact and Y any topological space and is further $H \subset C(X, Y)$, where $C(X, Y)$ is the set of continuous functions $f : X \rightarrow Y$. Then one can show ([Sch69], p.74) that the compact-open topology is the coarsest topology on H such that the map $H \times X \rightarrow Y$, $(\Phi, x) \mapsto \Phi(x)$ is continuous.

Corollary 6.23. *Let G and (X, d) be as in Lemma 6.21. Then the map $G \rightarrow \mathbb{R}; g \mapsto W_p(g\mu, \nu)$ is continuous for every choice of $\mu, \nu \in W_p(X)$.*

Proof. We know from Theorem 2.28 that $W_p(\cdot, \cdot)$ is continuous on $W_p(X) \times W_p(X)$. With this, $W_p(\cdot, \nu)$ is continuous on $W_p(X)$ for every choice of ν . Since we know from Lemma 6.21 that the group action of G on $W_p(X)$ is also continuous, we have continuity as a composition of continuous functions. \square

The map considered in Corollary 6.23 is of course the map for which we want to find the existence of a minimizer. The complement of the compact set on which this map is supposed to be considered for this should consist of all those elements in G which transports the two given measures “too far away”, so that the distance between them is increasing instead of decreasing. For this, it is useful to consider spaces on which most of the mass of probability measures is concentrated on compact balls, to have better control over their Wasserstein distances. To synthesize this, the action should “take care” of compact sets appropriately. The property we will use for this, similarly to Chapter 3, is *proper*.

Definition 6.24 (Proper action). Let X be a metric space. A continuous action $G \times X \rightarrow X$ of a group G on X is called *proper*, if the map $G \times X \rightarrow X \times X, (g, x) \mapsto (gx, x)$ is proper.

For a definition of proper maps, see Definition 3.2.

Remark 6.25. The definition for an action to be proper is equivalent to saying that for every $K_1, K_2 \subset X$ compact the set $\{g \in G \mid gK_1 \cap K_2 \neq \emptyset\}$ is compact in G .

One can show that if G acts proper on X then X/G is a Hausdorff space.

Definition 6.26 (Proper metric space). A metric space is called *proper* if every bounded closed set is compact.

Remark 6.27. • A metric space (X, d) being proper is equivalent to $x \mapsto d(x_0, x)$ being a proper map for any $x_0 \in X$.

- A proper metric space is automatically locally compact (which is why $W_p(X)$ cannot be proper in general). This fact enables us to use, for our situation suitable, Definition 6.24 of proper actions on locally compact spaces.
- Any proper space is complete.
- One can view “proper” as a finiteness condition, recalling that finite dimensional Hilbert spaces are always proper, but in infinite dimensions one can show that the closed unit ball is not sequentially compact.

Now we have all the ingredients we need to formulate the theorem.

Theorem 6.28 (Existence of a minimizer). *Let the Polish metric space (X, d) be proper. Additionally let the topology and the action of G be such that it acts properly on X . In this case a minimizer for the problem $\inf_{g \in G} W_p(g\mu, \nu)$ exists for every choice of $\mu, \nu \in \mathcal{P}_p(X)$.*

Corollary 6.29. *In the setting of Theorem 6.28 the Shape distance $D_p(\cdot, \cdot)$ is a metric.*

Proof of Corollary 6.29. Since we already know that $D_p(\cdot, \cdot)$ is a pseudometric, we only have to make sure that if $D_p(\mu, \nu) = 0$ then automatically $[\mu] = [\nu]$. From Theorem 6.28 we know that there exists an element $g_{min} \in G$ such that $0 = D_p(\mu, \nu) = W_p(g_{min}\mu, \nu)$. Since $W_p(\cdot, \cdot)$ is a metric we can conclude that $g_{min}\mu = \nu$ and thus $[\mu] = [\nu]$. \square

Remark 6.30. If there exists no minimizer of $\inf_{g \in G} W_p(g\mu, \nu)$ in general, one can always convert the pseudometric $D_p(\cdot, \cdot)$ into a metric via the so called metric identification. For this procedure one identifies points which have zero pseudodistance with each other.

We will use the following definition in the proof for the existence of a minimizer.

Definition 6.31. Let X be a topological space. A sequence of subsets $(K_n)_{n \in \mathbb{N}} \subset X$ is called *exhaustion of X* if $\bigcup_{n \in \mathbb{N}} K_n = X$ and $K_n \subset K_{n+1}$.

Proof of Theorem 6.28. If X is bounded, then it is automatically compact and then also G is compact. So from now on, let X be unbounded.

According to the prerequisites, for every $R \in \mathbb{R}$, $x \in X$: $\overline{B_R(x)}$ is compact and we have an exhaustion of X by balls centered around an arbitrary element $x \in X$. Now let $\mu, \nu \in \mathcal{P}_p(X)$ and $x \in X$. Using the Monotone Convergence Theorem, for every $\epsilon > 0$ there exists an element $R \in \mathbb{R}$ such that $\mu(B_R(x)) > 1 - \epsilon$ and $\nu(B_R(x)) > 1 - \epsilon$. Then let $C := W_p(\mu, \nu)^p$ and choose $g \in G$ such that $\overline{gB_{R'}(x)} \cap \overline{B_{R'}(x)} = \emptyset$, where $R' = kR$, $k > \frac{1}{2R} \sqrt[p]{\frac{C}{1-2\epsilon}} + 1$. To fulfill this, we have to choose $\epsilon < 1/2$.

If no such g exists, we know automatically that G is compact and the proof ends here. Otherwise, for $B' := B_{R'}(x)$, it is

$$\begin{aligned} & W_p(g\mu, \nu)^p \\ &= \inf_{\Pi \in \text{Adm}(g\mu, \nu)} \left(\int_{gB \times B} d^p(x, y) d\Pi(x, y) + \int_{X \times X \setminus gB \times B} d^p(x, y) d\Pi(x, y) \right) \\ &\geq (2(R' - R))^p \inf_{\Pi \in \text{Adm}(g\mu, \nu)} \Pi(gB \times B) \geq (2(k-1)R)^p (1-2\epsilon) \\ &> C \end{aligned}$$

This means that the minimizer has to be found in the subset $\{g \in G \mid \overline{gB_{R'}(x)} \cap \overline{B_{R'}(x)} \neq \emptyset\} \neq \emptyset$. Since all closed balls of (X, d) are compact and we

assumed the group action to be proper, this subset is compact, so the existence of a minimizer is guaranteed. \square

Remark 6.32. One might have the idea that, instead of demanding that there exists an exhaustion of X by compact balls, it is enough to say that there exists a compact exhaustion, i.e. to say that X is σ -compact. But in this case, it is not clear how to control the distances of elements of the set $gK_{n'}$ to elements of the set $K_{n'}$, where K_n are the elements of the exhaustion.

To confirm intuition we additionally give the following proposition, which is implicitly used in the preceding proof:

Proposition 6.33. *Let $\mu \in \mathcal{P}_p(X)$ and $g \in G$. Then $\text{supp}(g\mu) = g \text{supp}(\mu)$. In particular, two measures cannot be equivalent to each other if each support is not an isometric image of the respective other support.*

Proof. Let $g \in G$, $\mu \in \mathcal{P}_p(X)$. To establish this proposition, we need the following two identities:

- 1.) $\{A \in \mathcal{B}(X) \mid \mu(g^{-1}(A^c)) = 0\} = \{gA \in \mathcal{B}(X) \mid \mu(A^c) = 0\}$.
- 2.) Let I be an index family, we then have $\bigcap_{i \in I} gA_i = g \bigcap_{i \in I} A_i$ for $A_i \subset X$.

Thereby is $gA := \{ga \mid a \in A\}$, for $A \subset X$. Proof of the first identity: Let $A \in \mathcal{B}(X)$, then $A = gg^{-1}A$, since g is bijective. Then we have $\mu((g^{-1}A)^c) = \mu(g^{-1}A^c) = 0$. For the other way round, let $gA \in \mathcal{B}(X)$, then $\mu(g^{-1}(gA)^c) = \mu(A^c) = 0$. We will use the second identity without further justification. Now we can see that

$$\begin{aligned} \text{supp } g\mu &= \overline{\bigcap_{g\mu(A^c)=0} A} = \overline{\bigcap_{\mu(g^{-1}A^c)=0} A} = \overline{\bigcap_{\mu(A^c)=0} gA} \\ &= g \overline{\bigcap_{\mu(A^c)=0} A} = g \overline{\bigcap_{\mu(A^c)=0} A} = g \text{supp } \mu \end{aligned}$$

\square

6.3 The Shape space on Riemannian manifolds

For applications in physical scenarios, it is important to know whether Theorem 6.28 is applicable to the Euclidean space, or more generally to Riemannian manifolds.

In the realm of Riemannian geometry, one a priori distinguishes between two notions of isometry. Since every connected Riemannian manifold (M, h) is a metric space, being equipped with the Riemannian distance, there is on the one hand the notion of isometry in the sense defined in the beginning of this chapter. On the other hand, a diffeomorphism $f : (M_1, h_1) \rightarrow (M_2, h_2)$ is

called an isometry, in case it satisfies $h_1 = f^*h_2$, where $*$ denotes the pullback of tensors. Fortunately, every such diffeomorphism in this latter sense is also an isometry in the metric sense (whenever $(M_1, h_1) = (M_2, h_2)$). The theorem by Myers and Steenrod guarantees that the other way round is true as well [MS39].

We further note that Riemannian manifolds are separable and locally compact metric spaces by definition and the metric topology coincides with the manifold topology. However, in general Riemannian manifolds are not complete, i.e. Polish, but for our purposes it is of course necessary to require this from now on.

There are two further things to check. First, whether every complete Riemannian manifold is proper and secondly, whether the action of the isometry group is proper. Regarding the first issue we can make use of the following Theorem by Hopf and Rinow:

Theorem 6.34 (Hopf-Rinow theorem, [RH64]). *Let (M, h) be a Riemannian manifold. Then the following is equivalent:*

- i) M is a complete metric space.
- ii) Every closed bounded set is compact.

Now we have to have a look at the isometry group of (M, h) and its action on M .

Theorem 6.35 ([DvW28]). *Let (M, h) be a connected Riemannian manifold, then its isometry group G , equipped with the compact-open topology, is a locally compact topological group with continuous group action on M . In case M is compact, so is G .*

A proof of this can also be found in [SK63]. We again refer to A.1.1 for a definition of the compact-open topology.

We now state the next significant theorem.

Theorem 6.36. *Let (M, h) be a connected Riemannian manifold and let its isometry group G be equipped with the compact-open topology. Then the action of G on M is proper.*

A proof of this can be found in [Sch08], p. 17, Theorem 2.22.

Corollary 6.37. *In case the Polish space X is a complete connected Riemannian manifold and its isometry group is equipped with the compact-open topology, a minimizer for the Shape distance-problem exists and $D_p(\cdot, \cdot)$ is a metric distance on $\mathcal{S}_p(X)$.*

Example 6.38. The Euclidean space is of course a complete and connected Riemannian manifold.

Remark 6.39. • According to [Sch08], in case M is the union of more than one connected components, the action of the isometry group need not to be proper in general.

- The isometry group of a connected Riemannian manifold M is a closed subset of the group of diffeomorphisms of M with respect to the compact-open topology (see [SK63] p. 48, or [Sch08]).

Remark 6.40. Infinite dimensional (separable) Banach spaces don't serve as an example for our situation since closed unit balls are not compact in general. In this sense, one can regard the condition of the compactness of closed balls in Theorem 6.28 as a sort of finiteness condition (compare also Remark 6.27).

6.4 Fisher information on $\mathcal{S}_p(X)$

In Section 4.3 we have seen that the Fisher information is an important ingredient to describe Quantum dynamics in terms of Wasserstein geometry. It is thus interesting to know, whether it is well defined on $\mathcal{S}_p(X)$.

Proposition 6.41. *The Fisher information $I(\mu) = \int_M \|\nabla \ln(\mu)\|^2 d\mu$ on a Riemannian manifold M is invariant under the action of G on $W_p(X)$.*

Proof of Proposition 6.41. Since the Fisher Information is defined on the space of measures which are absolutely continuous with respect to the volume measure vol we have $d\mu = \mu d\text{vol}$. According to Lemma 4.11, $g\mu$ is again absolutely continuous with respect to vol , with $d(g\mu) = (\mu \circ g^{-1})d\text{vol}$. For $g \in G$ we then have

$$I(g\mu) = \int_M \|\nabla \ln(\mu \circ g^{-1})\|^2 d(g\mu) = \int_M \|\nabla \ln(\mu \circ g^{-1} \circ g)\|^2 d\mu = \int_M \|\nabla \ln(\mu)\|^2 d\mu.$$

□

Remark 6.42. Proposition 6.41 shows that the Fisher information is well-defined on $\mathcal{S}_p(X)$. A possible interpretation of this might be that the quantity the Fisher Information 'measures' is an internal property of the 'system' μ .

Remark 6.43. Let $V : X \rightarrow \mathbb{R}$ be a μ -integrable function then $V(\mu)$ is not G -invariant in general since $V(x)$ is not: $V(g\mu) = \int_X V(x) d(g\mu) = \int_X (V \circ g)(y) d(\mu)$.

6.5 Some properties of Shape space

Theorem 6.44. *Let the metric space (X, d) and the subgroup of the isometry group G be such that $d(gx, y) = d(x, g^{-1}y)$ for every $g \in G$. Let furthermore $D(x, y) := \inf_{g \in G} d(gx, y)$ be a metric on X/G in such a way that the infimum is always realized by an element $g \in G$. Then for $\pi : X \rightarrow X/G, x \mapsto [x]$ it is equivalent:*

- 1.) A is open in $(X/G, D)$
- 2.) $\pi^{-1}(A)$ is open in (X, d) .

Remark 6.45. Compare the conditions in this theorem (and the following) to the ones in 6.16 and in 6.48.

Corollary 6.46. *The topology on the Shape space coming from the Shape distance is the final topology with respect to the projection map π (i.e. the quotient topology). In particular, π is continuous.* \square

Proof of Theorem 6.44. Since in metric spaces the open r -Balls are a basis of the topology, we only have to prove 6.44 for $A = B_r^{X/G}([x]) := \{[y] \in X/G \mid D([x], [y]) < r\}$. Then we have for $B_r^X(x) := \{y \in X \mid d(x, y) < r\}$,

$$\begin{aligned} \pi^{-1}\left(B_r^{X/G}([x])\right) &= \{y \in X \mid [y] \in B_r^{X/G}([x])\} \\ &= \{y \in X \mid \exists g \in G : d(gx, y) < r\} \\ &= \bigcup_{g \in G} \{y \in X \mid d(gx, y) < r\} \\ &= \bigcup_{g \in G} B_r^X(gx). \end{aligned}$$

For the other direction let $A \subset X/G$ such that $\pi^{-1}(A) = \bigcup_{i \in I} B_{r_i}^X(x_i)$. Then

$$A = \pi\left(\bigcup_{i \in I} B_{r_i}^X(x_i)\right) = \bigcup_{i \in I} \pi(B_{r_i}^X(x_i)).$$

So the question reduces to whether $\pi(B_r^X(x)) \in \mathcal{O}_{X/G}$ for an arbitrary $x \in X$ and $r \in \mathbb{R}$:

$$\begin{aligned} \pi(B_r^X(x)) &= \pi(\{y \in X \mid d(x, y) < r\}) \\ &= \{\pi(y) \in X/G \mid d(x, y) < r\} \\ &= \{[y] \in X/G \mid \exists \tilde{y} \in [y] : d(x, \tilde{y}) < r\} \\ &= \{[y] \in X/G \mid D([x], [y]) < r\} \\ &= B_r^{X/G}(x) \end{aligned}$$

\square

Theorem 6.47. *Let the situation be as in Theorem 6.44. Then $(X/G, D)$ is complete if X is complete.*

Proof. Let $([x_n])_{n \in \mathbb{N}} \subset X/G$ be a Cauchy sequence. The idea of the proof is to define a Cauchy sequence in X such that its induced sequence in the space of equivalence classes is a subsequence of the original Cauchy sequence $([x_n])_{n \in \mathbb{N}}$. This is of use, since a Cauchy sequence converges if it has a converging subsequence.

Now, for $\epsilon_0 = (1/2)^0$ there is an $N_0 \in \mathbb{N}$ such that for every $m, n > N_0$, $D([x_n], [x_m]) < \epsilon_0$. Define y_0 to be an arbitrary element of $[x_{m_0}]$, for an $m_0 \geq N_0$. Further, for $\epsilon_1 = (1/2)^1$ there is an $N_1 \in \mathbb{N}$ such that for every $m, n > N_1 \geq$

m_0 , $D([x_n], [x_m]) < \epsilon_1$. Define y_1 to be an element of $[x_{m_1}]$, for an $m_1 \geq N_1$ such that $d(y_0, y_1) = D([y_0], [y_1])$, which is smaller 1 in particular, according to the choice of y_0 and y_1 . Successively, the sequence $(y_n) \subset X$ is chosen according to this scheme. In particular, for every $m \in \mathbb{N}$ it is $d(y_0, y_m) < \sum_{i=0}^{\infty} \epsilon_i < \sum_{i=0}^{\infty} (1/2)^i < \infty$. Since $\epsilon_i > 0 \forall i$ and $\lim_{n \rightarrow \infty} \epsilon_i = 0$, we have $\lim_{n \rightarrow \infty} \sum_{i=n}^{\infty} \epsilon_i = 0$ monotoneously and therefore for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for every $m, n > N$, $d(y_n, y_m) < \epsilon$. Since X is complete, there is an $y \in X$ such that $\lim_{n \rightarrow \infty} y_n = y$. Finally, $\lim_{n \rightarrow \infty} D([y_n], [y]) \leq \lim_{n \rightarrow \infty} d(y_n, y) = 0$. \square

Remark 6.48. As we have already pointed out in Remark 4.11, the condition $d(gx, y) = d(x, g^{-1}y)$ is necessary for D to be well-defined on X/G and lets us freely put the g in one of the two arguments. And the existence of a minimum for $D(x, y) = \inf_{g \in G} d(gx, y)$ is needed in the proof to ensure the existence of the chosen elements y_i , $i > 0$.

For the next, resulting, corollary we will additionally use that the p -th Wasserstein space $W_p(X)$ is complete, whenever (X, d) is (which we assume always to be the case).

Corollary 6.49. $(\mathcal{S}_p(X), D_p)$ is complete. \square

Proposition 6.50. Let the situation be like above. Then, if (X, d) is separable then $(X/G, D)$ is separable as well.

Proof. Let $A \subset X$ be countable and dense. Then $[A] := \{[a] \in X/G \mid a \in A\}$ is also countable. Regarding denseness, let $M \in X/G$. Then for every $x \in M$ there exists a sequence $(x_n)_{n \in \mathbb{N}} \subset A$ such that $d(x_n, x) \rightarrow 0$. But then also $D([x_n], [x]) \rightarrow 0$. \square

Corollary 6.51. $\mathcal{S}_p(X)$ is a Polish space. \square

In the following, by $(X, d_X) \cong (Y, d_Y)$ we mean that the metric spaces (X, d_X) and (Y, d_Y) are isometric, i.e. there is a bijective isometry between these spaces. Also, we denote the isometry group of (X, d_X) by G^X and the Shape Distance on $W_p(X)/G^X$ by $D_p^{G^X}$. G^Y and $D_p^{G^Y}$ analogously.

Lemma 6.52. Let (X, d_X) and (Y, d_Y) be Polish spaces. If $I : (X, d_X) \hookrightarrow (Y, d_Y)$ is a distance-preserving map, i.e. $d_Y(I(x), I(y)) = d_X(x, y)$ for all $x, y \in X$, then $I_{\#} : W_p(X) \hookrightarrow W_p(Y)$ is a distance-preserving map as well.

Proof. We are going to use 6.11, with that same notation. For this, it is neces-

sary that X is complete such that $I(X)$ is closed in Y

$$\begin{aligned}
W_p^Y(I_{\#}\mu, I_{\#}\nu)^p &= \inf_{\pi \in \text{Adm}(I_{\#}\mu, I_{\#}\nu)} \int_{Y \times Y} d_Y^p(x, y) d\pi(x, y) \\
&= \inf_{\pi \in \widetilde{\text{Adm}}(I_{\#}\mu, I_{\#}\nu)} \int_{I(X) \times I(X)} d_Y^p(x, y) d\pi(x, y) \\
&= \inf_{\pi \in \text{Adm}(\mu, \nu)} \int_{I(X) \times I(X)} d_Y^p(x, y) dI_{\#}\pi(x, y) \\
&= \inf_{\pi \in \text{Adm}(\mu, \nu)} \int_{X \times X} d_X^p(x, y) d\pi(x, y) \\
&= W_p^X(\mu, \nu).
\end{aligned}$$

□

Corollary 6.53. *From $(X, d_X) \cong (Y, d_Y)$ follows $W_p(X) \cong W_p(Y)$.*

Proof. Every distance preserving map I is injective, surjectivity of $I_{\#}$ follows from I being bijective: Let $\mu \in \mathcal{P}(Y)$, then $I_{\#}^{-1}\mu$ is mapped to μ by $I_{\#}$. □

Proposition 6.54. *If $I : (X, d_X) \hookrightarrow (Y, d_Y)$ is a distance-preserving map such that $I(X)$ is closed, then $[I_{\#}] : (W_p(X)/G^X, D_p^{G^X}) \hookrightarrow (W_p(Y)/G^Y, D_p^{G^Y})$, $[\mu] \mapsto [I_{\#}\mu]$ is a distance-preserving map as well.*

Proof. We start by showing that $[I_{\#}]$ is well-defined. For this we first note that the map $G^X \rightarrow G^{I(X)}$, $g \mapsto I \circ g \circ I^{-1}$ is a well-defined bijective map. $G^{I(X)}$ is meant to be the full isometry group of the metric space $I(X)$, which can be identified with a subgroup of G^Y by the natural group homomorphism mapping $g \in G^{I(X)}$ to the isometry on Y which maps y to y for $y \in Y \setminus I(X)$ and which has the same values as g otherwise. Let $\nu \in [\mu]$. That means there is an element $g_X \in G^X$ such that $\nu = g_X\mu$. Let $A \subset Y$ be measurable. Then for $g_{I(X)} \in G^{I(X)}$ being such that $g_X^{-1}I^{-1} = I^{-1}g_{I(X)}$ it is

$$I_{\#}\nu(A) = I_{\#}g_X\mu(A) = \mu(g_X^{-1}I^{-1}(A)) = \mu(I^{-1}g_{I(X)}(A)) = g_{I(X)}^{-1}I_{\#}\mu(A) \in [I_{\#}\mu].$$

By using Lemma 6.52 we now show that $[I_{\#}]$ indeed gives a distance-preserving map between the Shape spaces: Let $\mu \neq \nu$, then

$$\begin{aligned}
D_p^Y([I_{\#}\mu], [I_{\#}\nu]) &= \inf_{g_Y \in G^Y} d^Y(g_Y I_{\#}\mu, I_{\#}\nu) = \inf_{g_Y \in G^{I(X)}} d^Y(g_Y I_{\#}\mu, I_{\#}\nu) \\
&= \inf_{g_X \in G^X} d^Y(I_{\#}g_X\mu, I_{\#}\nu) = \inf_{g_X \in G^X} d^X(g_X\mu, \nu) \\
&= D^X([\mu], [\nu]).
\end{aligned}$$

□

Corollary 6.55. $(X, d_X) \cong (Y, d_Y) \Rightarrow (W_p(X)/G^X, D_p^X) \cong (W_p(Y)/G^Y, D_p^Y)$. *That means, in the given situation the Shape spaces for the full isometry groups are isometric.*

Proof. Since every isometry is injective, it only remains to show that $[I_{\#}]$ is surjective. This follows directly from the surjectivity of the map $I_{\#}$: Let $[\eta] \in W_p(X)/G^Y$. Then there is an element $\mu \in W_p(X)$ such that $\eta = I_{\#}\mu$. That means, $[\eta] = [I_{\#}\mu] = [I_{\#}]([\mu])$. \square

Remark 6.56. • Proposition 6.54 still holds true in case either G^Y is substituted by a subgroup which itself contains $G^{I(X)}$ as a subgroup. Or both, G^X and G^Y , are substituted by subgroups such that $g \mapsto I \circ g \circ I^{-1}$ is a bijection between these subgroups. For Corollary 6.55 the situation can naturally only be changed to the latter scenario.

- These series of results mean that Shape spaces only depend on the metric structure of the underlying space X and the chosen subgroup of the isometry group.

6.6 Geodesic structure of Shape space

As we have seen in Section 2.1, Wasserstein spaces have an interesting geodesic structure. Naturally, the question arises in which respects the geodesic structure of the Shape spaces is related.

In this section, let (X, d) , $G \subset ISO(X)$ be non-trivial and the action of G on X be such that Theorem 6.28 holds, i.e. such that D_p is a metric distance. When speaking of geodesics, we always mean constant speed geodesics (as in Definition 2.29).

Proposition 6.57. *Let $(\mu_t)_{t \in [0,1]}$ be a geodesic in $W_p(X)$ s.th. $D_p([\mu_0], [\mu_1]) = W_p(\mu_0, \mu_1) \neq 0$. Then $([\mu_t])_{t \in [0,1]}$ is a geodesic as well in $\mathcal{S}_p(X)$.*

Proof. For every $t, s \in [0, 1]$ we have

$$D_p([\mu_t], [\mu_s]) \leq W_p(\mu_t, \mu_s) = |t - s|D_p([\mu_0], [\mu_1])$$

which is equivalent to the definition of constant speed geodesics, according to Lemma 2.30. \square

To be able to formulate a more general statement, we give the following lemma.

Lemma 6.58. *Let $(\mu_t)_{t \in [0,1]}$ be any curve in $W_p(X)$. Since in general it holds $D_p([\mu_t], [\mu_s]) \leq W_p(\mu_t, \mu_s)$, for every $t \neq s \in [0, 1]$ there exist $C^{ts} \in \mathbb{R}_{>0}$ such that $C^{ts}D_p([\mu_t], [\mu_s]) = W_p(\mu_t, \mu_s)$ whenever $[\mu_t] \neq [\mu_s]$. \square*

Definition 6.59 (Quotient coefficients). We call the numbers C^{ts} from Lemma 6.58 *quotient coefficients* of the curve $(\mu_t)_{t \in [0,1]}$.

Proposition 6.60. *Let $(\mu_t)_{t \in [0,1]} \subset W_p(X)$ be a geodesic. Then $([\mu_t])_{t \in [0,1]}$ is a geodesic, too, if and only if $[\mu_t] \neq [\mu_s] \forall t, s \in [0, 1]$ and the quotient coefficients C^{ts} satisfy $C^{ts} = C^{01} \forall t \neq s \in [0, 1]$.*

Recall that in Definition 2.29 we demanded $\gamma_0 \neq \gamma_1$ for all metric geodesics $(\gamma_t)_{t \in [0,1]}$.

Proof. If $([\mu_t])_{t \in [0,1]}$ is a geodesic, then by definition $[\mu_t] \neq [\mu_s] \forall t \neq s \in [0, 1]$. Further, it is

$$\begin{aligned} C^{ts} D_p([\mu_t], [\mu_s]) &= W_p(\mu_t, \mu_s) = |t - s| W_p(\mu_0, \mu_1) = |t - s| C^{01} D_p([\mu_0], [\mu_1]) \\ &= C^{01} D_p([\mu_t], [\mu_s]) \end{aligned}$$

and therefore, $C^{ts} = C^{01} \forall t \neq s \in [0, 1]$.

Now let $[\mu_t] \neq [\mu_s] \forall t \neq s \in [0, 1]$ and $C^{ts} = C^{01} \forall t \neq s \in [0, 1]$. Then

$$\begin{aligned} D_p([\mu_t], [\mu_s]) &= \frac{1}{C^{ts}} W_p(\mu_t, \mu_s) = \frac{1}{C^{01}} W_p(\mu_t, \mu_s) = \frac{1}{C^{01}} |t - s| W_p(\mu_0, \mu_1) \\ &= |t - s| D_p([\mu_0], [\mu_1]). \end{aligned}$$

□

Example 6.61. Consider the function $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}$, $x \mapsto x$ and the revolution of its graph around the z -axis. With respect to the induced Euclidean metric structure, the symmetry group consists of the rotations around the z -axis. Let $a = (a^x, a^y, a^z)$ be an arbitrary point on the surface of revolution and $b = (b^x, b^y, b^z)$ another point on the surface with $b^y \neq a^y$ and $b^z \neq a^z$. Let δ_{a_t} be the metric geodesic in the Wasserstein space of this surface between δ_a and δ_b . Then $D_p(\delta_{a_t}, \delta_{a_s}) = \sqrt{(a_t^x - a_s^x)^2 + (a_t^z - a_s^z)^2}$, while the respective Wasserstein distances also depend on the y -components. That means, in this case the quotient coefficients are not equal and thus the projected curve is not a geodesic in Shape space, according to Proposition 6.60. However, if one would choose b such that $b^z \neq a^z$ but $b^y = a^y$ and $\text{sign}(b^x) = \text{sign}(a^x)$, the projection of the geodesics between those two points would be a geodesic again in Shape space.

One might wonder whether in case $\mu_1 = g_{\#} \mu_0$, geodesics between μ_0 and μ_1 have the form $\mu_t = g_t \mu_0$, for $g, g_t \in G$. Then, of course, the projection of this curve would only consist of one point, $[\mu_t] = [\mu_0]$. However, this is not true in general.

Example 6.62. Consider the graph of the function $f(x) = x^2$ and let it revolve around the z -axis. On this manifold with the induced Euclidean metric, the isometry group consists of reflection across and rotation around the z -axis. Now consider the Dirac measures δ_a , $a = (a_x, a_y, a_z)$ defined by $a_x = -1$ and $a_y = 1$ and δ_b , $b = (b_x, b_y, b_z)$ defined by $b_x = 1$ and $b_y = 1$. a and b are the reflections of one another. A continuous curve connecting δ_a and δ_b which is generated by isometries, i.e. a curve of the form $\delta_t = g_t \delta_a$, can only go along the semicircle in the hyperplane $y = 1$, connecting a and b . This arc has length π . But going along the graph of $f(x)$ would be shorter: $\int_{-1}^1 \sqrt{1 + 4x^2} dx < \pi$. So $\delta_t = g_t \delta_a$ cannot be a geodesic.

We so far have only asked, when projections of geodesics are geodesics again. Are there, however, other continuous curves which can be projected onto geodesics in Shape space?

Lemma 6.63. *Let $\mu_t : [0, 1] \rightarrow W_p(X)$ be a continuous curve in $W_p(X)$ and $g_t : [0, 1] \rightarrow G$ be a continuous curve in G . Then $g_t \# \mu_t : [0, 1] \rightarrow W_p(X)$ is a continuous curve in $W_p(X)$.*

Proof. Let $(a_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 1]$ that converges to $a \in [0, 1]$. Then (g_{a_n}, μ_{a_n}) is a net in $G \times W_p(X)$ which converges to (g_a, μ_a) , since g_t and μ_t are continuous. We know from Lemma 6.21 that the map $G \times W_p(X) \rightarrow W_p(X)$, $(g, \mu) \mapsto g \# \mu$ is continuous in our setting. From this we can deduce that $g_{a_n} \# \mu_{a_n}$ converges to $g_a \# \mu_a$, which proves our claim. \square

Corollary 6.64. *Let X be such that $W_p(X)$ is a non-branching geodesic space and such that the path-connected component of the identity element Id_G of G is not trivial. Then there exists a continuous curve $\tilde{\mu}_t : [0, 1] \rightarrow W_p(X)$ which is not a geodesic but its projection onto Shape space is.*

Recall from Proposition 2.40 that $W_p(X)$ is a non-branching geodesic space as soon as (X, d) is, for $p \in (1, \infty)$.

Proof. Since the path-connected component A of Id_G is not trivial, there is an element $\mu_1 \in W_p(X)$ such that for $g \in A$, $g\mu_1 \neq \mu_1$ (for example $\mu_1 = \delta_x$ with $g(x) \neq x$). Now choose $\mu_0 \in W_p(X)$ such that $D_p([\mu_0], [\mu_1]) = W_p(\mu_0, \mu_1)$ (this is possible due to Theorem 6.28). Since $W_p(X)$ is geodesic, there exists a geodesic μ_t between μ_0 and μ_1 and because of Proposition 6.57, $[\mu_t]$ is a geodesic, too.

Now let $\gamma_t : [0, 1] \rightarrow G$ be a continuous curve from Id_G to g , i.e. such that $\gamma_0 = Id_G$ and $\gamma_1 = g$. Let $g_t : [0, 1] \rightarrow G$ be such that $g_t = Id_G$ for $0 \leq t \leq 1/2$ and $g_t = \gamma_{2t-1}$ for $1/2 \leq t \leq 1$. From Lemma 6.63 we know that $g_t \# \mu_t$ is continuous, but it cannot be a geodesic since $W_p(X)$ is a non-branching space and we chose that $g\mu_1 \neq \mu_1$. \square

Above, the procedure we used was the pointwise transformation of a geodesic by isometries such that it branches continuously eventually. Another way to construct a curve that is projected onto a geodesic but which itself is none is detailed in the following. It will in particular allow to drop the assumption on the isometry group.

Lemma 6.65. *Let X be such that $W_p(X)$ is a non-branching geodesic space. In $W_p(X)$ let μ_t be a geodesic such that there is a $g \in G$ with $g\mu_1 \neq \mu_1$. Then, if $F : [0, 1] \rightarrow [0, 1]$ is a continuous function with $F(t) = 1$ for all $t \leq t_0 < 1$ but $F(t) \neq 1$ for all $t > t_0$, the mixing $\tilde{\mu}_t := (F(t) + (1 - F(t))g) \# \mu_t$ is not a geodesic, but its projection onto Shape space is.*

Proof. As long as $F(t) = 1$, $\tilde{\mu}_t = \mu_t$, but $\tilde{\mu}_t$ branches at $t = t_0$ and can therefore not be a geodesic anymore. But it is $[\tilde{\mu}_t] = F(t)[\mu_t] + (1 - F(t))[\mu_t] = [\mu_t]$ and therefore a geodesic by assumption. \square

Next to questions regarding the correspondence of curves in Wasserstein space to geodesics in Shape space, it is also interesting to notice whether in Shape space there is always a geodesic between two given points.

Corollary 6.66. *If $W_p(X)$ is geodesic, then $\mathcal{S}_p(X)$ is geodesic, too.*

Proof. We argue similarly as in the proof of 6.64. Let $\bar{\mu}, \bar{\nu} \in \mathcal{S}_p(X)$ and $\mu \in W_p(X)$ such that $[\mu] = \bar{\mu}$. Then according to Theorem 6.28, there exists a $\nu \in W_p(X)$ such that $[\nu] = \bar{\nu}$ and $W_p(\mu, \nu) = D_p([\mu], [\nu]) = D_p(\bar{\mu}, \bar{\nu})$. Since we assumed $W_p(X)$ to be geodesic, there exists a geodesic μ_t in $W_p(X)$ with $\mu_0 = \mu$ and $\mu_1 = \nu$. Now, according to Proposition 6.57, $[\mu_t]$ is a geodesic in $\mathcal{S}_p(X)$ connecting $\bar{\mu}$ and $\bar{\nu}$. \square

Remember from Theorem 2.34 that $W_2(X)$ is geodesic whenever (X, d) is geodesic.

The geodesic we have found in the proof of 6.66 between $\bar{\mu}$ and $\bar{\nu}$ need not be unique. A priori it is not clear whether geodesics between other elements of the equivalence classes $\bar{\mu}$ and $\bar{\nu}$ which have Wasserstein distance $D_p(\bar{\mu}, \bar{\nu})$, are projected onto the same curve in Shape space. And, quite in general, it can still be that there are geodesics in Shape space which cannot be written as the projection of a geodesic, or any other continuous curve, in Wasserstein space at all.

Chapter 7

Infinitesimal dynamics on Shape space

As outlined in Chapter 5, we have defined Shape space to be able to talk about the dynamics of shapes in quantum theory. The idea was to make quantum dynamics independent of positioning in (absolute) space. In this chapter, we finally want to find out, to which extend solutions of the Madelung equations, and in this sense solutions of the Schrödinger equation, can be regarded as the motion of shapes. Concretely, we want to capture, how much information is lost, when projecting a solution μ_t to its image curve $[\mu_t]$ in Shape space and whether it is possible to find an intrinsic description of this dynamics in Shape space.¹

As in Section 2.4, let (M, h) be a smooth, connected, complete Riemannian manifold throughout this chapter.

7.1 Isometric actions on $TW(M)$

In Chapter 6 we have already extensively made use of the fact that $ISO(M)$ induces an isometric action also on $W_p(M)$ via pushforward of measures. Using the definition of differentiability of maps between Wasserstein spaces introduced in Section 3, we are going to see now that $ISO(M)$ also acts isometrically on $TW(M)$.

Recall the following corollary from Section 3:

¹Let us point out that our aim here is different from the aim of investigating symmetry groups \tilde{G} for the Madelung equations (or the Schrödinger equation) and from finding solutions that are invariant under the action of such a symmetry group and which therefore correspond to functions on the quotient space M/\tilde{G} . The reader who is interested in these kind of things should consult for example [Olv00].

Corollary 7.1. *Let $g \in ISO(M)$. The corresponding pushforward of measures $g_{\#} : W(M) \rightarrow W(M)$ is differentiable with $dg_{\#\mu}(v) = dg(v)$, up to a negligible set.*

Remark 7.2. In the expression $dg_{\#\mu}(v) = dg(v)$, the left hand side displays the differential of the map $g_{\#}$ at point $(\mu, v) \in T_{\mu}W(M)$. The right hand side displays the vector $dg(v) \in T_{g_{\#\mu}}W(M)$, where $dg_x : T_xM \rightarrow T_{g(x)}M$ is the tangential map of the diffeomorphism $g \in ISO(M)$.

Lemma 7.3. *Let G be a subgroup of $ISO(M)$. The map*

$$\begin{aligned} G \times TW(M) &\rightarrow TW(M) \\ (g, (\mu, v)) &\mapsto (g\mu, dg(v)) \end{aligned}$$

defines an induced action of G on the tangent bundle of $W(M)$, where we regard dg as a differential of $g_{\#}$.

Proof. It is $(id, (\mu, v)) \mapsto (\mu, v)$ and $((g_1g_2)\mu, d(g_1g_2)(v)) = (g_1(g_2\mu), dg_1dg_2(v))$, according to Corollary 3.25. \square

Note that even though $dg_{\#\mu}(v) = dg(v)$ is a pointwise well-defined map between tangent bundles of Wasserstein spaces (it is in particular independent of μ and $dg(v)$ is a well defined map between tangent bundles of manifolds), it is not unique as a differential of the map $g_{\#}$. In this proof we fixed dg as one possible differential of $g_{\#}$. If we would not do that, i.e. if we would not fix this or fix no differential at all, results would only hold up to a negligible set which would not grant an action of G on $TW(M)$.

Remark 7.4. Instead of G we could have equivalently let $G_{\#} = \{g_{\#} \mid g \in G\}$ act on $TW(M)$ (see Corollary 6.14), with the same definition of the action. It's just a matter of the perspective which group to choose for this.

Definition 7.5. Let $F : W(N) \rightarrow W(M)$ be differentiable, dF a fixed differential of F , $\mu \in W(N)$ and $H_{F(\mu)}$ the formal Riemannian metric tensor on $W(M)$ at point $F(\mu) \in W(M)$ (see Definition 2.59). Then, for $v, w \in T_{\mu}W(M)$, the *pullback* $(F^*H)_{\mu}$ of $H_{F(\mu)}$ is defined as

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

Remark 7.6. If we would not fix a differential of F , the pullback would not be uniquely defined. Unfortunately, one cannot simply say that otherwise the pullback is only defined up to a negligible set, since these are subsets of $TW(M)$, whereas F^*H is defined pointwise on $T_{\mu}W(M) \times T_{\mu}W(M)$.

Definition 7.7. Analogously to the finite dimensional case, we call a bijective differentiable map $F : W(M) \rightarrow W(M)$ with differentiable inverse a *formal Riemannian isometry*, in case there is a representative of dF such that for all $\mu \in W(M)$ $(F^*H)_{\mu}(v, w) = H_{\mu}(v, w)$ for all $(v, w) \in T_{\mu}W(M) \times T_{\mu}W(M)$.

We want to call such a representative as in Definition 7.7 *suitable*.

Lemma 7.8. *Let $g \in ISO(M)$, then $g_{\#} : W(M) \rightarrow W(M)$ is a formal Riemannian isometry.*

Proof. Let $dg_{\#} = dg$ be the differential found in Lemma 7.1. We have to show that it is suitable. For this, let $\mu \in W(M)$ and $v, w \in T_{\mu}W(M)$. Then,

$$\begin{aligned} (g_{\#}^*H)_{\mu}(v, w) &= \int_M h_y(dg(v), dg(w)) \, dg\mu(y) = \int_M h_{g^{-1}(y)}(v, w) \, dg\mu(y) \\ &= \int_M h_x(v, w) \, d\mu(x) = H_{\mu}(v, w). \end{aligned}$$

□

Recall that in Example 3.7 we have already seen that

$$\|(dg_{\#})_{\mu}\| = \operatorname{ess\,sup}_{x \in M}^{\mu} \|dg_x\| = 1.$$

Lemma 7.9. *F is a formal Riemannian isometry in the sense of Definition 7.7 iff there is a representative of dF such that for every $\mu \in W(M)$ $dF_{\mu} : T_{\mu}W(M) \rightarrow T_{F(\mu)}W(M)$ is a metric isometry in the sense of 6.1 with respect to the metrics induced by the L^2 -norms.*

Proof. This is a direct consequence of the definition of the formal Riemannian metric tensor as the L^2 -scalar product. Let us first assume that F is a formal Riemannian isometry and let dF be a suitable representative. Then for every $\mu \in W(M)$ and $v \in T_{\mu}W(M)$,

$$\|dF(v)\|_{L^2(F(\mu))}^2 = H_{F(\mu)}(dF(v), dF(v)) = H_{\mu}(v, v) = \|v\|_{L^2(\mu)}^2.$$

Conversely, let F be such that there is a representative of dF such that dF_{μ} is an L^2 -isometry for every $\mu \in W(M)$. It in particular preserves the corresponding L^2 -scalar products. For $\mu \in W(M)$ and $v, w \in T_{\mu}W(M)$, then,

$$\begin{aligned} (F^*H)_{\mu}(v, w) &= H_{F(\mu)}(dF_{\mu}(v), dF_{\mu}(w)) = \int h(dF_{\mu}(v), dF_{\mu}(w)) \, d(F(\mu)) \\ &= \int h(v, w) \, d\mu = H_{\mu}(v, w). \end{aligned}$$

Thus, dF is a suitable representative and F is a formal Riemannian isometry. □

Proposition 7.10. *Every formal Riemannian isometry is an isometry in the metric sense of its Wasserstein space.*

Proof. Let F be a formal Riemannian isometry. Since by definition F is bijective with differentiable inverse, every a.c. couple (μ_t, v_t) can be represented as the image of another a.c. couple $(\tilde{\mu}_t, \tilde{v}_t)$. Just choose $\tilde{\mu}_t := F^{-1}(\mu_t)$ and $\tilde{v}_t := dF^{-1}(v_t)$. Then, $\mu_t = F(\tilde{\mu}_t)$ and, using Corollary 3.25, $v_t = dF(\tilde{v}_t)$ almost everywhere. Conversely, every image of an a.c. couple, in the above sense, is an

a.c. couple. Let dF be a suitable representative. For $\mu, \nu \in W(M)$ and μ_t a.c. connecting them, we then have according to 2.57:

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

□

It would be interesting to find out, whether the converse implication of Proposition 7.10 is true as well, as it is in finite dimensions.

We are going to need the following Lemma to legitimize our definition for the tangent space on $\mathcal{S}(M)$ in the next section.

Lemma 7.11. *Let $g \in ISO(M)$, then $T_{g\#\mu}W(M) = dg(T_\mu W(M))$ for all $\mu \in W(M)$. Here, we again regard dg as a, fixed, differential of $g\#$.*

Proof. The inclusion “ \supset ” follows from dg being a differential of $g\#$ and the other inclusion follows from g^{-1} being differentiable, too. □

Of course, this lemma holds true for every bijective differentiable map F whose inverse map is differentiable, too.

7.2 Tangent space on $\mathcal{S}(M)$

Before we deal with the construction of tangent spaces on $\mathcal{S}(M)$, let us have a look at the finite dimensional case, where the situation is already well explored: In case a finite dimensional Lie group G acts smoothly, freely and properly on a finite dimensional smooth manifold M , the quotient space M/G is a topological manifold of dimension $\dim M/G = \dim M - \dim G$. The quotient space furthermore has a unique smooth structure with respect to which the quotient map $\pi : M \rightarrow M/G$ is a smooth submersion (see [Lee01]). The orbits $G.x$, $x \in M$, then, can proven to be embedded submanifolds of M and the kernel of the linear and surjective map $d\pi_x : T_x M \rightarrow T_{\pi(x)}(M/G)$ is precisely $T_x(G.x)$ (for these statements see again [Lee01], Lemma 5.29 and Theorem 7.5). This yields the following isomorphism of linear spaces:

$$(7.1) \quad T_{[x]}(M/G) \cong T_x M / T_x(G.x).$$

This means, qualitatively, in $T_{[x]}(M/G)$ all directions showing “orbitwards” are modded out.

Of course, we cannot apply this statement to our case - $W(M)$ is an infinite dimensional, only *formal*, Riemannian manifold and also our G -action is not free

in general². However, we can mod out *by hand* those directions of $T_\mu W(M)$ that point towards the orbit of μ .

Recall that the group of isometries $ISO(M)$ of a finite dimensional Riemannian manifold M is a Lie group, due to a theorem by Myers and Steenrod, [MS39]. And in case $M = \mathbb{R}^n$, $ISO(M) = O(n) \ltimes \mathbb{R}^n$, where $O(n)$ is the orthogonal group of \mathbb{R}^n and \ltimes denotes the semidirect product of groups. We will find the above mentioned orbitward directions on Wasserstein space to be induced by the Lie algebra $\mathfrak{iso}(n)$ of $ISO(\mathbb{R}^n)$. One can show that $\mathfrak{iso}(n) = \mathfrak{so}(n) \ltimes \mathbb{R}^n$, where \ltimes denotes the semidirect sum of algebras. (See e.g. [Fec06] for a reference on this result and for a definition of the semidirect sum. However, we will be more explicit, below.) By $\mathfrak{so}(n) := \{M \in \mathbb{R}^{n \times n} \mid M^t = -M\}$ we denote the vector space of skew-symmetric matrices with the matrix-commutator $[M, N] := M \cdot N - N \cdot M$. It is $\dim \mathfrak{iso}(n) = \dim ISO(\mathbb{R}^n) = \frac{n(n+1)}{2}$.

We assume that the reader is familiar with the basic notions of Lie theory.³ However, we would like to recall the definition of fundamental vector fields, as they will be important for us.

Let \mathfrak{g} be the Lie algebra of a Lie group G , $X \in \mathfrak{g}$ a left-invariant vector field and $\gamma_X(t)$ the maximal integral curve of X starting at the identity element, i.e. with $\gamma_X(0) = e$. Then $\varphi_X(t)$ is defined for all values $t \in \mathbb{R}$ and the map $\varphi_X : \mathbb{R} \rightarrow G$ is a homomorphism of Lie groups, i.e. $\gamma_X(0) = e$, $\varphi_X(s+t) = \varphi_X(s) \cdot \varphi_X(t)$ and $\varphi_{sX}(t) = \varphi_X(s \cdot t)$ for all $s, t \in \mathbb{R}$.

Definition 7.12 (Exponential map). The map

$$\begin{aligned} \exp : \mathfrak{g} &\longrightarrow G \\ X &\longmapsto \varphi_X(1) \end{aligned}$$

is called the *exponential map* of the Lie group G .

The curve $\mathbb{R} \ni t \mapsto \exp(tX) = \varphi_{tX}(1) = \varphi_X(t)$, then, is the maximal integral curve of X through e and the curve $\mathbb{R} \ni t \mapsto g \cdot \exp(tX)$ the maximal integral curve of X through $g \in G$.

We now assume that the Lie group G acts on a smooth manifold M . Then, every element of the Lie algebra \mathfrak{g} of G defines a certain vector field on M , the so called *fundamental vector field*.

Definition 7.13 (Fundamental vector field). Let G act on M on the left and let further be $X \in \mathfrak{g}$. The vector field $\tilde{X} \in \Gamma^\infty(M)$ on M defined by

$$\tilde{X}(x) := \frac{d}{dt} (\exp(-tX) \cdot x) \Big|_{t=0}$$

²Just consider a normal distribution on \mathbb{R}^3 centered around the origin. It is invariant under rotations, as we have seen in Section 4.2.

³For more background knowledge on Lie theory consult for example [Bau09] (in German) or [Lee01].

is called the *fundamental vector field* corresponding to X .⁴

The map $\mathfrak{g} \rightarrow \Gamma^\infty(M)$, $X \mapsto \tilde{X}$ is linear and $[\widetilde{[X, Y]}] = [\tilde{X}, \tilde{Y}]$, where the latter is the commutator for vector fields, i.e. the Lie bracket. In particular, the set of all fundamental vector fields form a Lie-subalgebra of the Lie-algebra of all smooth vector fields on M and is Lie-algebra homomorphic to \mathfrak{g} .

Is G the isometry group of a complete Riemannian manifold (M, h) , the Lie algebra of fundamental vector fields corresponding to \mathfrak{g} is precisely the Lie algebra of Killing vector fields⁵, the so called *Killing algebra* on (M, h) . In case $M = \mathbb{R}^n$ and h is the Euclidean metric tensor, one can calculate the general Killing vector field ([Fec06]), and thus the general fundamental vector field, to be of the form

$$(7.2) \quad \xi^{(A,a)} = \frac{1}{2}(Ah)^{ij}M_{ij} + a^iP_i,$$

where $A \in \mathfrak{so}(n)$ and $a \in \mathbb{R}^n$. Here, we have used the Einstein summation convention. The vector fields

$$(7.3) \quad M_{ij} = -M_{ji} = x_i\partial_j - x_j\partial_i \quad \text{and} \quad P_i = \partial_i$$

constitute a basis of the Killing algebra.

Recall from Section 2.4, in particular Theorem 2.47, that the tangent space on $W(M)$ is constructed using the weak continuity equation. A family of vector fields v_t is considered to be tangent along an absolutely continuous curve μ_t if the weak continuity equation holds and if $\|v_t\|_{L^2(\mu_t)}$ is minimal among all the possible solutions. Now, to identify all those tangent vectors that show in the direction of the orbit, let us consider the pushforward of a measure μ by the flow of the left-invariant vector fields on $G = ISO(M)$.

Proposition 7.14. *Let $X \in \mathfrak{g}$ and \tilde{X} be the corresponding fundamental vector field on M . Further, let $\mu \in W(M)$ and $\mu_t := \exp(-tX)_\# \mu$, $0 \leq t \leq 1$. Then, the tuple (μ_t, \tilde{X}) is a solution of the weak continuity equation (2.6), i.e.*

$$\int_0^1 \int_M \left(\frac{\partial}{\partial t} \varphi(x, t) + h(\nabla \varphi(x, t), \tilde{X}(x)) \right) d(\exp(-tX)_\# \mu)(x) dt = 0.$$

Proof. The integral curve $\exp(-tX)$ on G defines the flow $\exp(-tX) : M \rightarrow M$ on M which is a diffeomorphism for every t and $\exp(0) = Id$. It is in particular the flow of the fundamental vector field \tilde{X} , so that the statement of the proposition follows from Theorem 2.42. \square

Let us note that because the tangent map of the orbit map $g_x : G \rightarrow G.x$, $g \mapsto g.x$ at point e , i.e. $(dg_x)_e : T_e G \rightarrow T_x G.x$, is surjective, every curve

⁴There is a similar definition for right-actions of G . But here, we will only be interested in left-actions.

⁵A vector field X is a Killing vector field of the Riemannian metric h , in case $\mathcal{L}_X h = 0$, where \mathcal{L} denotes the Lie derivative.

that is initially tangential to $G.x$ infinitesimally coincides with the integral line of a fundamental vector field starting at x . This is why in our case we will settle with those kinds of curves considered in Proposition 7.14. By doing this, we naturally only consider the path-connected component of Id_G with respect to which directions it is going to be modded out by.

The next question to be asked is whether μ_t from Proposition 7.14 is indeed absolutely continuous in $W(M)$. According to Theorem 2.47 this is so, in case $\int_0^1 \|\tilde{X}\|_{L^2(\mu_t)} dt < \infty$.

We will give a positive answer for this only in case (M, h) is the Euclidean space. However, we conjecture that the same holds true also for compact manifolds and Riemannian manifolds with nonnegative curvature. In the hyperbolic case a restriction to the subspace of compactly supported measures seems necessary.

Proposition 7.15. *Let \tilde{X} be a fundamental vector field corresponding to an element $X \in \mathfrak{iso}(n)$ and let $\mu \in W(\mathbb{R}^n)$. Then*

$$\|\tilde{X}\|_{L^2(\mu)} < \infty.$$

Proof. First we point out that since $\mu \in W(M)$, it is $\int_{\mathbb{R}^n} \|x\|^2 d\mu(x) < \infty$. Because of this, also $\int_{\mathbb{R}^n} x_i d\mu(x) < \infty$ and $\int_{\mathbb{R}^n} x_i x_j d\mu(x) < \infty$, where x_i denotes the i -th component of x .

In (7.2) we have seen that every fundamental vector field \tilde{X} on \mathbb{R}^n is of the form $\tilde{X} = \frac{1}{2}(Ah)^{ij} M_{ij} + a^i P_i$. So, calculating the Euclidean norm of \tilde{X} , we obtain for $a = (a_1, \dots, a_n)$,

$$\|\tilde{X}\|^2 = \|\frac{1}{2}A^{ij}M_{ij}\|^2 + \|a\|^2 + \langle A^{ij}M_{ij}, a^i P_i \rangle.$$

The integral over the second term is finite. The third term is a linear combination of the components x_i , whose integral is again finite. The first term consists of a linear combination of terms of the form $x_i^2 + x_j^2$ plus a linear combination of terms of the form $x_i x_j$. The integral over both of these terms is also finite, which means that in total $\int_{\mathbb{R}^n} \|\tilde{X}\|^2 d\mu = \|\tilde{X}\|_{L^2(\mu)}^2 < \infty$. \square

Lemma 7.16. *Let G be the Lie group of isometries of a Riemannian manifold M and \mathfrak{g} its Lie algebra. Further, let \tilde{X} be the fundamental vector field corresponding to $X \in \mathfrak{g}$. If $\mu \in \mathcal{P}(M)$ and $\mu_t := (\exp(-tX))_{\#}\mu$, then, for all t ,*

$$\|\tilde{X}\|_{L^2(\mu_t)}^2 = \|\tilde{X}\|_{L^2(\mu)}^2.$$

Proof.

$$\begin{aligned} \|\tilde{X}\|_{L^2(\mu_t)}^2 &= \int_{\mathbb{R}^n} \|\tilde{X}_x\|^2 d((\exp(-tX))_{\#}\mu)(x) = \int_{\mathbb{R}^n} \|\tilde{X}_{\exp(-tX)y}\|^2 d\mu(y) \\ &= \int_{\mathbb{R}^n} \|d(\exp(-tX))\tilde{X}_y\|^2 d\mu(y) = \int_{\mathbb{R}^n} \|\tilde{X}_y\|^2 d\mu(y) \\ &= \|\tilde{X}\|_{L^2(\mu)}^2. \end{aligned}$$

□

Corollary 7.17. *Let $X \in \mathfrak{iso}(n)$, \tilde{X} the corresponding fundamental vector field, $\mu \in W(M)$ and $\mu_t := (\exp(-tX))_{\#}\mu$. Then*

$$\int_0^1 \|\tilde{X}\|_{L^2(\mu_t)} dt < \infty.$$

Proof. Combining Lemma 7.16 and Proposition 7.15, we get

$$\int_0^1 \|\tilde{X}\|_{L^2(\mu_t)} dt = \int_0^1 \|\tilde{X}\|_{L^2(\mu)} dt = \|\tilde{X}\|_{L^2(\mu)} < \infty.$$

□

But although \tilde{X} is an element of $L^2(\mu_t)$ for all t and fulfills the regularity condition of Corollary 7.17, it is not necessarily the minimal vector field along μ_t . (Consider for example for $n = 2$, $\mu_t := \exp(-tX)_{\#}\mu$ where μ has a normal distribution and $\exp(-tX)$ consists of rotations around the origin. Then $\mu_t = \mu$ for all t and the minimal vector field is 0.) But as we have seen in Proposition 3.16, we can apply the orthogonal linear projection $P^\mu : L^2(\mu) \rightarrow T_\mu W(\mathbb{R}^n)$, so that $P^{\mu_t}(\tilde{X}) \in T_{\mu_t} W(\mathbb{R}^n)$ and the continuity equation still holds. This is why we define

$$U_\mu := \{P^\mu(\tilde{X}) \mid \tilde{X} \text{ is a fundamental vector field}\} \subset L^2(\mu).$$

U_μ is a complete subspace of $T_\mu W(\mathbb{R}^n)$ and contains all the vectors pointing towards the orbit of μ . With this definition, we can finally propose a notion for $T_{[\mu]}\mathcal{S}(\mathbb{R}^n)$.

Definition 7.18 (Tangent space on $\mathcal{S}(\mathbb{R}^n)$). We define the tangent space on $\mathcal{S}(\mathbb{R}^n)$ at point $[\mu]$ to be the set

$$T_{[\mu]}\mathcal{S}(\mathbb{R}^n) := T_\mu W(\mathbb{R}^n) / U_\mu.$$

Proposition 7.19. *It is*

$$T_\mu W(\mathbb{R}^n) / U_\mu \cong T_{g\mu} W(\mathbb{R}^n) / U_{g\mu}.$$

In this sense, Definition 7.18 is independent of the choice of the representative.

Proof. We know from Lemma 7.11 that $dg : T_\mu W(\mathbb{R}^n) \rightarrow T_{g\mu} W(\mathbb{R}^n)$ is an isomorphism for every $g \in G$. It remains to show that dg also maps U_μ isomorphically onto $U_{g\mu}$. To show this, let us recall the known fact that if \tilde{X} is a fundamental vector field, then so is $dg(\tilde{X})$ (see for example [Bau09], Theorem 1.25). Since for $v = v^\top \oplus v^\perp \in T_\mu W(\mathbb{R}^n) \oplus T_\mu^\perp W(\mathbb{R}^n)$, $dg(v) = dg(v^\top) \oplus dg(v^\perp)$ with $dg(v^\top) \in T_{g\mu} W(\mathbb{R}^n)$ and $dg(v^\perp) \in T_{g\mu}^\perp W(\mathbb{R}^n)$, it is $dg \circ P^\mu = P^{g\mu} \circ dg$. So, if $v \in U_\mu$, then there is a fundamental vector field \tilde{X} such that $v = P^\mu(\tilde{X})$ and $dg(v) = dg(P^\mu(\tilde{X})) = P^{g\mu}(dg(\tilde{X}))$ and the statement holds true. □

Example 7.20.

$$T_{[\delta]}\mathcal{S}(\mathbb{R}^3) = \mathbb{R}^3 / U_\mu \cong \{0\}.$$

One can interpret this in a sense that there is no non-trivial classical one-particle motion in Shape space (which quantifies what we have mentioned in Chapter 5 about the motion of classical particles).

Remark 7.21. Definition 7.18 is meaningful for every $\mu \in W(M)$ on a general Riemannian manifold, as long as $\int_0^1 \|\tilde{X}\|_{L^2(\mu_t)} dt < \infty$ for every μ_t of the form $\mu_t = (\exp(-tX))_\# \mu$. In particular, Proposition 7.19 does not use any properties specific to \mathbb{R}^n . Just substitute M for \mathbb{R}^n .

7.3 Quantum dynamics on $\mathcal{S}(\mathbb{R}^n)$

To finalize the program that we have outlined in Chapter 5 we now want to find out, how far solutions of the free Schrödinger equation can naturally be regarded as curves in Shape space.

To begin with, let thus $\psi(x, t)$ be a solution of the free Schrödinger equation (i.e. of equation (4.1) with $V = 0$), $\rho(x, t) = |\psi(x, t)|^2$ and $d\mu_t = \rho(x, t)d\lambda$, while λ being the Lebesgue measure on \mathbb{R}^3 .

Proposition 7.22. *There is no $g \in ISO(\mathbb{R}^3)$ such that $\mu_t = g\mu_s$ for all possible times $s \neq t$. In particular, $\mu_t \neq \mu_s$ for all $t \neq s$.*

In particular, this means that $[\mu_t] \neq [\mu_s]$ for all times $t \neq s$, so that a passage from the curve of probability measures on the background space \mathbb{R}^3 to a curve of shapes (without explicit background) is possible in a meaningful and well-defined way.

Proof of Proposition 7.22. To prove our claim we use formula (4.7), i.e. the fact that free wave packets spread over time. Let $t_0 = 0$ and $t > s$. Assume $\mu_t = g\mu_s$. Then according to Lemma 4.11, $\rho(x, t) = \rho(g^{-1}(x), t)$.

$$\begin{aligned} \sigma_x^2(t) &= \int_{\mathbb{R}^3} (x - \langle x(t) \rangle)^2 d\mu_t = \int_{\mathbb{R}^3} \left(x - \int_{\mathbb{R}^3} x d\mu_t \right)^2 d\mu_t \\ &= \int_{\mathbb{R}^3} \left(x - \int_{\mathbb{R}^3} x \rho(g^{-1}(x), s) d\lambda(x) \right)^2 \rho(g^{-1}(x), s) d\lambda(x) \\ &= \int_{\mathbb{R}^3} \left(g(x) - \int_{\mathbb{R}^3} g(x) d\mu_s(x) \right)^2 d\mu_s(x) \\ &= \int_{\mathbb{R}^3} \|g(x)\|^2 d\mu_s(x) - \left(\int_{\mathbb{R}^3} g(x) d\mu_s(x) \right)^2 \end{aligned}$$

The last step is a general rule for variances.

It is $ISO(\mathbb{R}^3) = O(\mathbb{R}^3) \ltimes \mathbb{R}^3$, where $O(\mathbb{R}^3)$ is the orthogonal group of \mathbb{R}^3 . So, suppose first that $g \in O(\mathbb{R}^3)$. Then,

$$\begin{aligned} & \int_{\mathbb{R}^3} \|g(x)\|^2 d\mu_s(x) - \left(\int_{\mathbb{R}^3} g(x) d\mu_s(x) \right)^2 \\ &= \int_{\mathbb{R}^3} \|x\|^2 d\mu_s(x) - \left(g \left(\int_{\mathbb{R}^3} x d\mu_s(x) \right) \right)^2 \\ &= \int_{\mathbb{R}^3} \|x\|^2 d\mu_s(x) - \left(\int_{\mathbb{R}^3} x d\mu_s(x) \right)^2. \end{aligned}$$

So $\sigma_x^2(t) = \sigma_x^2(s)$ which cannot be true due to the spreading of wave packets.

Assume now that $g(x) = x + a$, $a \in \mathbb{R}^3$. Then

$$\begin{aligned} & \int_{\mathbb{R}^3} \|g(x)\|^2 d\mu_s(x) - \left(\int_{\mathbb{R}^3} g(x) d\mu_s(x) \right)^2 \\ &= \int_{\mathbb{R}^3} (\|x\|^2 + 2\langle x, a \rangle + \|a\|^2) d\mu_s(x) - \left(\int_{\mathbb{R}^3} x d\mu_s(x) + \int_{\mathbb{R}^3} a d\mu_s(x) \right)^2 \\ &= \int_{\mathbb{R}^3} \|x\|^2 d\mu_s(x) + \int_{\mathbb{R}^3} \langle x, a \rangle d\mu_s(x) + \int_{\mathbb{R}^3} \|a\|^2 d\mu_s(x) \\ &\quad - \left(\int_{\mathbb{R}^3} x d\mu_s(x) \right)^2 - \left(\int_{\mathbb{R}^3} a d\mu_s(x) \right)^2 - 2\langle \int_{\mathbb{R}^3} x d\mu_s(x), \int_{\mathbb{R}^3} a d\mu_s(x) \rangle \\ &= \sigma_x^2(s). \end{aligned}$$

In this case, again, $\sigma_x^2(t) = \sigma_x^2(s)$ and thus a contradiction. \square

Corollary 7.23. *Let $g \in ISO(\mathbb{R}^3)$. Then $\psi(x, t) \neq \psi(g^{-1}(x), s)$ for all t, s .*

Proof. Suppose $\psi(x, t) = \psi(g^{-1}(x), s)$, then also $|\psi(x, t)|^2 = |\psi(g^{-1}(x), s)|^2$, i.e. $\rho(x, t) = \rho(g^{-1}(x), s)$, which does not hold true. \square

Let us now continue to study of the specific solution of the Schrödinger equation that we were already engaged with in Chapter 4. In the following let $\psi(x, t)$ always be of the form (4.10), $\rho(x, t)$ of the form (4.11), $S(x, t)$ the phase of $\psi(x, t)$ and $F_s(\cdot, t)$ the flow map of $\nabla S(x, t)$ (see Lemma 4.9). Let furthermore μ_t be the measure defined by $d\mu_t = \rho(x, t)d\lambda(x)$.

Lemma 7.24. *Let $g \in O(3)$. Then μ_t is invariant under the action of g , i.e. $g\mu_t = \mu_t$.*

Proof.

$$\begin{aligned} \rho(g^{-1}(x), t) &= N^2 \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2} \exp\left(-\left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1} \|g^{-1}(x)\|^2\right) \\ &= N^2 \left(1 + \frac{\hbar^2 t^2}{m^2 l^4}\right)^{-3/2} \exp\left(-\left(l^2 + \frac{\hbar^2 t^2}{m^2 l^2}\right)^{-1} \|x\|^2\right) \\ &= \rho(x, t). \end{aligned}$$

□

In Theorem 4.12 we have seen that $F_s(\cdot, t)$ is the optimal transport map from μ_s to μ_t . Now we will see that the optimal transport of the ranged measure $g\mu_t$ to μ_s is given by the composition of moving $g\mu_t$ back to μ_t and transporting this to μ_s .

Proposition 7.25. *Let $g \in ISO(\mathbb{R}^3)$. Then $F_s(\cdot, t) \circ g^{-1}$ is the optimal transport map from $g\mu_s$ to μ_t for every t, s .*

Proof. As in Theorem 4.12, we argue with the cost function $d^2/2$. (See also Remark 4.13). Because of Lemma 7.24, we only have to deal with isometries of the form $g(x) = x + a$, $a \in \mathbb{R}$. Let g be such a map. Defining $C = \sqrt{\frac{1 + \frac{\hbar^2}{m^2 l^4} t^2}{1 + \frac{\hbar^2}{m^2 l^4} s^2}}$ and $\bar{\varphi}(x) := \frac{1}{2}C\|x - a\|^2$, we see that $F_s(g^{-1}(x), t) = C(x - a) = \nabla \bar{\varphi}$. With this and $\varphi(x) := \frac{1}{2}(\|x\|^2 - C\|x - a\|^2)$, the rest of the proof is analogous to the proof of Theorem 4.12. □

Theorem 7.26. *Let $g(x) = x + a \in ISO(\mathbb{R}^3)$ with $a \in \mathbb{R}^n$. Then*

$$(7.4) \quad W(g\mu_s, \mu_t)^2 = W(\mu_t, \mu_s)^2 + \|a\|^2.$$

Equation (7.4) resembles Pythagoras' formula: Going with the flow of ∇S and translation by a seem to be in this sense *orthogonal* to each other.

Proof. Again, for calculational convenience we use the cost function $d^2/2$, so that in the end we have to multiply by the factor 2. With the definition of C as in the previous proof, we have

$$\begin{aligned} W^{d^2/2}(g\mu_s, \mu_t)^2 &= \frac{1}{2} \int_{\mathbb{R}^3} \|x - F_s(g^{-1}(x), t)\|^2 d(g\mu_s)(x) \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \|g(x) - F_s(x, t)\|^2 d\mu_s(x) \\ &= \frac{1}{2} \int_{\mathbb{R}^3} \|x - a - Cx\|^2 d\mu_s(x). \end{aligned}$$

Defining $\bar{Q} := N^2 \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right)^{-3/2}$ and $\bar{A} := \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right)^{-1}$ transforms this into

$$\frac{\bar{Q}}{2} \int_{\mathbb{R}^3} \|(1 - C)x - a\|^2 \exp(-\bar{A}\|x\|^2) d\lambda(x).$$

Performing a transformation with the diffeomorphism $\Phi(y) = \frac{1}{1-C}(y+a)$ yields

$$\begin{aligned}
& \frac{\bar{Q}}{2} \int_{\mathbb{R}^3} \|y\|^2 \exp\left(-\frac{\bar{A}}{(1-C)^2} \|y+a\|^2\right) \frac{1}{(1-C)^3} d\lambda(x) \\
&= \frac{\bar{Q}}{2(1-C)^3} \int_{\mathbb{R}^3} \|y\|^2 \exp\left(-\frac{\bar{A}}{(1-C)^2} (\|y\|^2 + \|a\|^2 + 2\langle a, y \rangle)\right) d\lambda(x) \\
&= \frac{\bar{Q}}{2(1-C)^3} \exp\left(-\frac{\bar{A}}{(1-C)^2} \|a\|^2\right) \cdot \\
&\quad \int_{\mathbb{R}^3} \|y\|^2 \exp\left(-\frac{\bar{A}}{(1-C)^2} (\|y\|^2 + 2\langle a, y \rangle)\right) d\lambda(x).
\end{aligned}$$

Define $A := 2\frac{\bar{A}}{(1-C)^2}$, $Q := \frac{\bar{Q}}{(1-C)^3} \exp(-\frac{1}{2}A\|a\|^2)$ and $s := -Aa$. We then have

$$\begin{aligned}
& \frac{1}{2}Q \int_{\mathbb{R}^3} \|y\|^2 \exp\left(-\frac{1}{2}A\|y\|^2 + \langle s, y \rangle\right) d\lambda(x) \\
&= \left(\frac{2\pi}{A}\right)^{3/2} \exp\left(\frac{1}{2A}\|s\|^2\right) \left(\frac{3Q}{2A} + \frac{Q}{2A^2}\|s\|^2\right) \\
&= \left(\frac{2\pi}{A}\right)^{3/2} \exp\left(\frac{A}{2}\|a\|^2\right) \left(\frac{3Q}{2A} + \frac{Q}{2}\|a\|^2\right) \\
&= \left(\frac{2\pi}{A}\right)^{3/2} \frac{Q}{2} \exp\left(\frac{A}{2}\|a\|^2\right) \left(\frac{3}{A} + \|a\|^2\right) \\
&= \left(\frac{2\pi}{A}\right)^{3/2} \frac{\bar{Q}}{2(1-C)^3} \left(\frac{3}{A} + \|a\|^2\right) \\
&= \left(\frac{2\pi(1-C)^2}{2A}\right)^{3/2} \frac{\bar{Q}}{2(1-C)^3} \left(\frac{3(1-C)^3}{2A} + \|a\|^2\right) \\
&= \left(\frac{\pi}{\bar{A}}\right)^{3/2} \frac{\bar{Q}}{2} \left(\frac{3(1-C)^2}{2\bar{A}} + \|a\|^2\right).
\end{aligned}$$

Substituting the remaining abbreviations, we get

$$\begin{aligned}
& \pi^{3/2} l^3 \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right)^{3/2} \frac{(l^2 \pi)^{-3/2}}{2} \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right)^{-3/2} \\
&\quad \left(\frac{3}{2} \left(l^2 + \frac{\hbar^2 s^2}{m^2 l^2}\right) \left(1 - \sqrt{\frac{1 + \frac{\hbar^2 t^2}{m^2 l^4}}{1 + \frac{\hbar^2 s^2}{m^2 l^4}}}\right)^2 + \|a\|^2\right) \\
&= \frac{1}{2} \left(\left(\frac{3}{2} l^2 \left(1 + \frac{\hbar^2 s^2}{m^2 l^4}\right) \left(1 - \sqrt{\frac{1 + \frac{\hbar^2 t^2}{m^2 l^4}}{1 + \frac{\hbar^2 s^2}{m^2 l^4}}}\right)^2\right) + \|a\|^2\right) \\
&= \frac{3}{4} l^2 \left(\sqrt{1 + \frac{\hbar^2 s^2}{m^2 l^4}} - \sqrt{1 + \frac{\hbar^2 t^2}{m^2 l^4}}\right)^2 + \frac{1}{2} \|a\|^2 \\
&= W^{d^2/2} (\mu_t, \mu_s)^2 + \frac{1}{2} \|a\|^2.
\end{aligned}$$

The last step follows from Theorem 4.16. □

Evidently, $W(g\mu_s, \mu_t)$ is minimal, in case $a = 0$. So we can formulate the following corollary.

Corollary 7.27. *For all times t, s we have*

$$W_2(\mu_t, \mu_s) = D_2([\mu_t], [\mu_s]).$$

□

So not only is the passage from μ_t to $[\mu_t]$ a meaningful procedure, as we have seen in Proposition 7.22, for this particular solution also the distances between the intermediate times do not change by this: There is no location on space a measure μ_t can be brought to such that cost for transporting it to any other μ_t gets cheaper. Or, in other words, μ_t cannot be brought closer to any μ_s than the location it is already situated. Using Corollary 4.17, this leads to the following important statement.

Theorem 7.28. *Let the curve of measures μ_t be defined by $d\mu_t = \rho(x, t)d\lambda(x)$, where $\rho(x, t)$ is of the form (4.11). Then for $s, t, u \in \mathbb{R}_{\geq 0}$, $s < u < t$ it is*

$$D([\mu_s], [\mu_t]) = D([\mu_s], [\mu_u]) + D([\mu_u], [\mu_t]).$$

So $([\mu_t])$ is a geodesic in $\mathcal{S}(\mathbb{R}^3)$ in the sense of shortest paths.

Coming back to what we have outlined in Chapter 5, we have seen that the curve μ_t can naturally be regarded as motion in Shape space, as a curve that subjects the change of shapes. In particular without resorting to a spatial background structure. With the definitions in Section 7.2, also $\frac{1}{m}\nabla S(x, t)$ can naturally be regarded as tangent to $[\mu_t]$: Since $\mu_s \neq g\mu_t$ for any $t, s \in [0, \infty]$ and $g \in ISO(\mathbb{R}^3)$, ∇S can be attached to $[\mu_t]$ in a well-defined way. And because $\nabla S(x, t) \neq P^{\mu_t}(\tilde{X})$ for any fundamental vector field \tilde{X} , $[\nabla S(x, t)] \neq 0$.

Appendix A

Additional information

A.1 Useful concepts

In this part of the appendix, we briefly recall a few basic concepts which are of importance to us. For a thorough reading on the topological notions and on proper maps we recommend for example [TD09] (available also in English) or [Que01]. More information on Polish spaces and can be found in [Que01].

A.1.1 Topological spaces

Let X be a topological space. For the definition of proper maps we have given in 3.2, we needed to know what locally compact and what Hausdorff spaces are. We recall the definitions here.

Definition A.1. A *neighborhood* U_x of a point $x \in X$ is a subset of X which contains an open subset V of x which in turn contains X , i.e. $x \in V \subset U_x$.

Definition A.2. X is called *locally compact* if every point in X has a compact neighborhood.

Example A.3. \mathbb{R}^n with the euclidean topology is locally compact, \mathbb{Q}^n together with the induced topology is not.

Definition A.4. X is called a *Hausdorff* space if for every two points $x, y \in X$ there exist neighborhoods U_x of x and U_y of y with $U_x \cap U_y = \emptyset$.

Since on general topological spaces there is no notion of distance between elements, continuous functions between topological spaces cannot be characterized via convergence of sequences, other than continuous functions between metric spaces. However, there is a generalization of this concept, namely the convergence of so called nets, which we want to discuss briefly. We used this in the proof of Lemma 6.21, where we proved the continuity of the group action of G on $\mathcal{P}_p(X)$, provided G is a topological group with continuous group action on X .

Definition A.5. A set I is called *directed* if there is a relation \triangleleft on I with the properties

1. $i \triangleleft i \ \forall i \in I$,
2. $i_1 \triangleleft i_2 \ \wedge \ i_2 \triangleleft i_3 \Rightarrow i_1 \triangleleft i_3$,
3. $\forall i_1, i_2 \in I \ \exists i_3 \in I : i_1 \triangleleft i_3 \ \wedge \ i_2 \triangleleft i_3$.

Definition A.6. Let I be a directed set and X a set. A *net* is defined as a map $x : I \rightarrow X$ and often denoted by $(x_i)_{i \in I}$, in analogy to sequences.

Apparently, every sequence is a net with $I = \mathbb{N}$.

Definition A.7. Let X be a topological space. A net $(x_i)_{i \in I}$ *converges* towards $x \in X$, denoted by $x_i \rightarrow x$, if for each neighborhood U_x of x there is an $i_0 \in I$ such that $x_i \in U$ for all $i_0 \triangleleft i$.

In Hausdorff spaces, limits of nets are unique [Wil04].

Theorem A.8. Let $f : X \rightarrow Y$ be a map between topological spaces. One can show, [Que01], that f is continuous in $x \in X$ iff from $x_i \rightarrow x$ it follows $f(x_i) \rightarrow f(x)$ for every net $(x_i)_{i \in I}$ in X .

In Section 6.3 we provided the isometry group of a Riemannian manifold with the compact-open topology to ensure that the Shape distance is really metric distance. We here recall the definition of this topology. By $C(X, Y)$ we mean the set of continuous functions from X to Y .

Definition A.9. A collection \mathcal{B} of open subsets of a topological space X is called a *basis* for the topology, in case every open subset of X can be written as an arbitrary union of elements of \mathcal{B} .

Definition A.10 (Compact-open topology). Let X, Y be topological spaces and $H \subset C(X, Y)$. For $A \subset X$ and $B \subset Y$ let

$$[A, B] := \{f \in H \mid f(A) \subset B\} \subset H.$$

The set

$$\mathcal{B}_{Sub} := \{[C, O] \subset H \mid C \subset X \text{ compact, } O \subset Y \text{ open}\}$$

defines the subbasis of the *compact-open topology* on H . This means that

$$\mathcal{B} := \left\{ \bigcap_{i=1}^n [C_i, O_i] \mid C_i \subset X \text{ compact, } O_i \subset Y \text{ open, } 1 \leq i \leq n, n \in \mathbb{N} \right\}$$

is a basis for the compact-open topology on H .

This means, with respect to the compact-open topology, every open subset of H is an arbitrary union of finite intersections of elements in \mathcal{B}_{Sub} .

A.1.2 Proper maps

The help of proper maps is needed in this thesis at two occasions. First, to find examples of maps between Wasserstein spaces that are absolutely continuous (Definition 3.2 and Theorem 3.3) and secondly, in the form of proper actions, to find conditions for which the Shape distance is actually a metric distance (Definition 6.24 and Corollary 6.29). Here, we give a more general definition of a proper map than we have given in 3.2 and relate these two definitions with each other. For a proof of the theorems given below, see [TD09].

Let X and Y be topological spaces.

Definition A.11. A continuous map $f : X \rightarrow Y$ is called *proper*, if one of the following equivalent conditions are fulfilled:

1. f is closed and every preimage $f^{-1}(y)$, $y \in Y$ is compact
2. For every space T , $f \times id : X \times T \rightarrow Y \times T$ is closed

The equivalence of these two conditions is due to a property of compactness which does not use open covers:

Theorem A.12. *If the projection $X \times T \rightarrow T$ is closed for every T then X is compact.*

Now, one can show the following two theorems, the latter of which is taken as the definition for proper maps in section 3, Definition 3.2:

Theorem A.13. *Is $f : X \rightarrow Y$ proper and $K \subset Y$ compact, then $f^{-1}(K)$ is compact.*

Theorem A.14. *Let f be a continuous map from a Hausdorff space X to a locally compact Hausdorff space Y . Then f is proper iff for every compact set $K \subset Y$ the preimage is compact. In case f is proper, then X is locally compact.*

A.1.3 Disintegration theorem

To be able to prove Theorem 3.3, we rely on the following statement (see [AGS08]).

Theorem A.15. *Let X and Y be Radon spaces. Furthermore let $\mu \in \mathcal{P}(X)$ and $f : X \rightarrow Y$ be a measurable map. Then there exists a $f_{\#}\mu$ -almost everywhere uniquely determined family of probability measures $\{\mu_y\}_{y \in Y}$ on X such that*

- for every measurable set $A \subset X$ the map $y \mapsto \mu_y(A)$ is measurable,
- $\mu_y(X \setminus f^{-1}(y)) = 0$ for $f_{\#}\mu$ -almost every $y \in Y$,
- for every measurable function $g : X \rightarrow [0, \infty]$ it is

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

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

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

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

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

A.2 Remarks on differentiable structures on probability measures

A.2.1 Otto's approach

The intuition that on \mathbb{R}^n the set of probability-densities ρ allows for a formal infinite dimensional Riemannian structure based on the Wasserstein distance is due to Felix Otto. In his seminal paper [Ott01] he applied this to study the porous medium equation, interpreted as a Wasserstein-gradient flow. While the concepts developed there were still on a heuristic level, later other authors have set out to make notions precise and calculations rigorous (see [Vil08] for an exhaustive bibliography). In this treatise, we are mainly concerned with the approach taken by Ambrosio, Gigli and Savaré, which started with [AGS05]. However, to enable the reader to compare literature we briefly want to mention Ottos original ideas. He explicitly noted to be “deliberately sloppy about the differential structure of the manifold”. To start with, he thought of the tangent space at a density ρ as $\{s : \mathbb{R}^n \rightarrow \mathbb{R} \mid \int s d\lambda = 0\}$. Which makes sense since an infinitesimal variation of ρ should not change its total mass. Having the continuity equation $\partial_t \rho + \nabla \cdot (\rho \nabla p) = 0$ in mind, in a next step Otto identified this space, up to an additive constant, with all the maps $p : \mathbb{R}^n \rightarrow \mathbb{R}$ for which $-\nabla \cdot (\rho \nabla p) = s$ holds, s having zero Lebesgue mass. Then the Riemannian metric tensor is defined by

$$h_\rho(s_1, s_2) := \int \langle \nabla p_1, \nabla p_2 \rangle \rho d\lambda,$$

where $s_i = \nabla \cdot (\rho \nabla p_i)$. Equivalently, the tangent space at ρ can directly be seen as $\{(-\nabla \cdot (\rho v))\}$, v ranging through all the gradients. Accordingly, the metric tensor looks like $h_\rho(-\nabla(v\rho), -\nabla(w\rho)) := \int \langle v, w \rangle \rho d\lambda$.

A.2.2 John Lott’s geometric calculations on Wasserstein space

In [Lot07], John Lott studied geometric properties of the space

$$\mathcal{P}_2^\infty(M) := \{\mu \in \mathcal{P}_2(M) \mid \mu = \mu(x)dvol_M, \mu(x) > 0, \mu \in \mathcal{C}^\infty(M)\}^1,$$

where the manifold M is supposed to be smooth, connected and closed. Instead of equipping $\mathcal{P}_2^\infty(M)$ with the Wasserstein (subspace-) topology, Lott claims the existence of an (infinite dimensional) smooth structure in the sense of [KM97]. Coordinates of a measure μ are then given by $(F_\phi(\mu))_{\phi \in \mathcal{C}^\infty(M)}$, where $F_\phi(\mu) := \int_M \phi d\mu$. Although being a priori topologically distinct from Wasserstein space, Lott claims the smooth tangent space at $\mathcal{P}_2^\infty(M)$ to correspond to the definition Otto gave. Assuming this is true, every ac curve in $\mathcal{P}_2^\infty(M)$ is almost everywhere differentiable. Within this setup, Lott showed that for every smooth curve, its length with respect to the Wasserstein distance equals its Riemannian length, using Otto’s Riemannian metric.

A.2.3 Geometric tangent space

The definition of the tangent space at point μ we gave in 2.5 has the unsatisfying feature that $\dim \mathcal{T}_\delta W(M) = \dim M$, whereas often, $\dim \mathcal{T}_\mu W(M) = \infty$. In [Gig08], Gigli defined what he called the *geometric tangent space*. The basic idea is that it consists of all constant speed geodesics which start at the point μ where the tangent space is supposed to be located at, such that it can be thought of *space of directions*. It can be turned into a complete and separable metric space and it can be shown that it is generally larger than the regular, analytic, tangent space. It allows for an injective isometry of the latter, mapping $\nabla\varphi$ to the constant speed geodesic $(Id \times t\nabla\varphi)_\# \mu$. However, in case μ is a regular measure, this map is an isomorphism. This situation becomes comprehensible by recalling that not every geodesic is induced by a map in the above manner (but which is indeed the case as soon as the starting point is a regular measure). Of course, every geodesic is an absolutely continuous curve and its tangent vectors are well-defined for every $t \in (0, 1)$. But at time $t = 0$ it can be the case that mass needs to be split and this ”direction” cannot be captured by a tangent vector in the analytical tangent space. These are precisely the directions which are added in the geometric tangent space.

Gigli used his enlarged tangent space in [Gig08] to define a subdifferential calculus for geodesically convex functionals that is able to reproduce classical results from convex analysis. We, on the other hand, are interested in defining a total derivative for mappings between Wasserstein spaces. For this, it seems more convenient to use the analytic tangent space. But also due to the relationships we want to draw with Quantum mechanics, we are prone to the study of differentiability properties along ac curves.

¹Here we use the same symbol for the measure and its density with respect to the Riemannian volume.

Bibliography

- [AC08] Luigi Ambrosio and Gianluca Crippa. Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields. *Transport Equations and Multi-D Hyperbolic Conservation Laws, Lecture Notes of the Unione Matematica Italiana*, 5:3–4, 2008.
- [AG09] Luigi Ambrosio and Nicola Gigli. *A user’s guide to optimal transport*. web draft, 2009.
- [AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser, 2005.
- [AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows*. Birkhäuser, 2008.
- [Aub82] Thierry Aubin. *Nonlinear analysis on manifolds. Monge-Ampere equations*, volume 252. Springer, 1982.
- [Bar01] Julian Barbour. *The end of time: The next revolution in physics*. Oxford University Press, 2001.
- [Bau09] Helga Baum. *Eichfeldtheorie*. Springer, 2009.
- [BB82] Julian B. Barbour and Bruno Bertotti. Mach’s principle and the structure of dynamical theories. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 382(1783):295–306, 1982.
- [BB99] Jean-David Benamou and Yann Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. *Contemporary Mathematics*, 226:1–12, 1999.
- [BBCK92] Iwo Białynicki-Birula, Marek Cieplak, and Jerzy Kaminski. *Theory of Quanta*. Oxford University Press, 1992.
- [Bes87] Arthur L. Besse. *Einstein Manifolds*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, 1 edition, 1987.

- [Bor26a] Max Born. Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 37:863–867, 1926.
- [Bor26b] Max Born. Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 38(11-12):803–827, 1926.
- [Bre87] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.
- [Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.
- [DvW28] David Dantzig and Bartel Leendert von Waerden. Über metrische homogene Räume. *Abhandlungen aus dem Mathematischen Seminar an der Universität Hamburg*, 6:374–376, 1928.
- [Fec06] Marián Fecko. *Differential geometry and Lie groups for physicists*. Cambridge University Press, 2006.
- [Gig08] Nicola Gigli. *On the geometry of the space of measures in \mathbb{R}^d endowed with the quadratic optimal transportation distance*. PhD thesis, Scuola Normale Superiore, Pisa, 2008.
- [Gig12] Nicola Gigli. *Second Order Analysis on $(P_2(M), W_2)$* . American Mathematical Society, 2012.
- [Hör90] Lars Hörmander. *The Analysis of Linear Partial Differential Operators I*. Grundlehren Der Mathematischen Wissenschaften. Springer, 2 edition, 1990.
- [Kan58] Leonid Kantorovitch. On the translocation of masses. *Management Science*, 5(1):1–4, oct 1958.
- [Kil94] Tero Kilpeläinen. Weighted sobolev spaces and capacity. *Ann. Acad. Sci. Fenn. Ser. AI Math*, 19(1):95–113, 1994.
- [KM97] Andreas Kriegl and Peter Michor. *The Convenient Setting of Global Analysis*. American Mathematical Society, sep 1997.
- [Kuf85] Alois Kufner. *Weighted sobolev spaces*, volume 31. John Wiley & Sons Incorporated, 1985.
- [Lee01] John M. Lee. *Introduction to Smooth manifolds*. Springer, 2001.
- [Lot07] John Lott. Some geometric calculations on wasserstein space. *Communications in Mathematical Physics*, 277(2):423–437, nov 2007.

- [Lüc02] Wolfgang Lück. *L2-invariants: Theory and applications to geometry and K-theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge A Series of Modern Surveys in Mathematics v. 44. Springer, 1 edition, 2002.
- [Mac83] Ernst Mach. *Die Mechanik in ihrer Entwicklung*. Brockhaus, 1883.
- [Mad27] Erwin Madelung. Quantentheorie in hydrodynamischer Form. *Zeitschrift für Physik*, 40(3-4):322–326, mar 1927.
- [Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.
- [MS39] Sumner Byron Myers and Norman Earl Steenrod. The group of isometries of a riemannian manifold. *The Annals of Mathematics*, 40(2):400, apr 1939.
- [NA13] Alyssa Ney and David Z Albert. *The wave function: Essays on the metaphysics of quantum mechanics*. Oxford University Press, 2013.
- [Olv00] Peter J. Olver. *Applications of Lie groups to differential equations*, volume 107. Springer, 2000.
- [Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, jan 2001.
- [Que01] Boto von Querenburg. *Mengentheoretische Topologie*. Springer, 2001.
- [RH64] Willi Rinow and Heinz Hopf. Über den Begriff der vollständigen differentialgeometrischen Fläche. In *Selecta Heinz Hopf*, pages 64–79. Springer, 1964.
- [Rov96] Carlo Rovelli. Relational quantum mechanics. *International Journal of Theoretical Physics*, 35(8):1637–1678, 1996.
- [Sch26a] Erwin Schrödinger. Quantisierung als Eigenwertproblem I. *Annalen der Physik*, 384(4):361–376, 1926.
- [Sch26b] Erwin Schrödinger. Quantisierung als Eigenwertproblem II. *Annalen der Physik*, 384(6):489–527, 1926.
- [Sch26c] Erwin Schrödinger. Quantisierung als Eigenwertproblem III. *Annalen der Physik*, 385(13):437–490, 1926.
- [Sch26d] Erwin Schrödinger. Quantisierung als Eigenwertproblem IV. *Annalen der Physik*, 386(18):109–139, 1926.
- [Sch69] Horst Schubert. *Topologie*. B.G. Teubner, Stuttgart, second edition, 1969.

- [Sch08] Jürgen Schlecht. Die Isometriegruppe einer Riemannschen Mannigfaltigkeit. Master's thesis, Eberhard Karls Universität Tübingen, 03 2008.
- [SK63] Katsumi Nomizu Shoshichi Kobayashi. *Foundations of Differential Geometry (Wiley Classics Library) (Volume 1)*. Wiley-Interscience, 1963.
- [Sta01] Reinhard Starkl. *Materie-Feld-Struktur: Repetitorium der Theoretischen Physik*. Springer, 2001.
- [TD09] Tammo Tom Dieck. Mengentheoretische Topologie. *Skript, www.uni-math.gwdg.de/tammo/sos*, 2009.
- [Vil03] Cédric Villani. *Topics in Optimal Transportation (Graduate Studies in Mathematics, Vol. 58)*. American Mathematical Society, 2003.
- [Vil08] Cédric Villani. *Optimal Transport: Old and New (Grundlehren der mathematischen Wissenschaften)*. Springer, 2008.
- [vR12] Max-K von Renesse. An optimal transport view of schrödinger's equation. *Canadian mathematical bulletin*, 55(4):858–869, 2012.
- [Wal94] Timothy C Wallstrom. Inequivalence between the schrödinger equation and the madelung hydrodynamic equations. *Physical Review A*, 49(3):1613, 1994.
- [Wil04] Stephen Willard. *General Topology*. Dover, 2004.
- [Zha] David Zhang. Gaussian integrals. <http://david-k-zhang.com/notes/gaussian-integrals.html>.