A Wasserstein-like Distance on Vector Fields

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

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

vorgelegt von
Vincent Sommer

aus Kassel
Göttingen, 22.12.22

Betreuungsausschuss
Prof. Dr. Max Wardetzky, Diskrete Differential-Geometrie, Institut für Numerische und Angewandte Mathematik

Prof. Dr. Thomas Schick, Topologie und Geometrie, Mathematisches Institut

Prof. Dr. Bernhard Schmitzer, Optimal Transport, Institut für Informatik

Mitglieder der Prüfungskommission
Referent/in: Prof. Dr. Max Wardetzky, Diskrete Differential-Geometrie, Institut für Numerische und Angewandte Mathematik

Korreferent/in: Prof. Dr. Thomas Schick, Topologie und Geometrie, Mathematisches Institut

Weitere Mitglieder der Prüfungskommission:
Prof. Dr. Bernhard Schmitzer, Optimal Transport, Institut für Informatik
Prof. Dr. Axel Munk, Mathematische Statistik, Institut für Mathematische Stochastik
Prof. Dr. Gerlind Plonka-Hoch, Mathematische Signal- und Bildverarbeitung, Institut für Numerische und Angewandte Mathematik
Prof. Dr. Christoph Lehrenfeld„ Institut für Numerische und Angewandte Mathematik

Tag der mündlichen Prüfung: 14.2.2023.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel verfasst habe.

Göttingen, den 22.12.2022 .....

Acknowledgements
I want to thank my advisor Max Wardetzky for his great support in mathematical questions as well as for his methodic and human advice, in particular during the corona virus pandemic. He always had a helping hand and always cared for my problems.
Also, I want to thank Thomas Schick and Bernhard Schmitzer for being part of my advisory committee and for the many ideas they gave me in our talks and discussions.
Moreover, I want to thank Albert Chern and Keenan Crane for their time in our fruitful discussions, which were of great help and sparked the main ideas of this whole thesis. I greatly value their contribution.
Another person to whom I am most grateful and who was of invaluable help to me when discussing questions of numerics and programming was Henrik Schuhmacher.
Last but not least, I want to thank my wife Dorothea, who supported me during happy as well as during difficult times.

Contents

1 Introduction

The main idea of this work originated from various discussions with my advisor Max Wardetzky and his friend Keenan Crane. Coming from a differential geometry background and recently also having started a journey into Wasserstein geometry, in particular being fascinated Villani's work [41], we were initially led by a basic comparison of differential forms and measures. Comparing these various notions for integration, we started wondering whether a Wasserstein-like way to transport sections on Riemannian manifolds such as forms and vector field could be constructed. Thus, we started to investigate in this direction and quickly came up with our own ideas. However, instead of transporting any kind of object of differential geometry, we quickly restricted ourselves to the case of vector fields on surfaces since that already proved to have enough complexity.
A nice feature of optimal transport and the Wasserstein distance is that it has such a beautiful intuition behind it. This intuition has often shown to have an actual meaning that goes beyond being a purely academic quality. In fact, most applications that rely on optimal transport do not just experiment with it as one of many possible distances, but can quite clearly explain why it is the *correct* metric to use. In particular, applications in image morphing come to mind.
A remarkable property of optimal transport is its compatibility with smooth structures to a surprising degree. In fact, Otto showed that the space of probability measures equipped with the Wasserstein-metric can be seen as a smooth manifold itself in many aspects [25]. Similar to the task of discretization, one can ask which aspects of the classical optimal transport should be translated to the new setting. In other words, one needs an anchor point for how to start. Our main idea was to define a distance by using the Wasserstein distance on a bundle over a surface. This is a novel concept for distances on vector fields. We considered two approaches for how the bundle over surfaces can be used to yield new distances for vector fields. One particular approach gave rise to a definition that seemed very interesting to us and has an intuitive relation to parallel transport.
Having experience in discretization, another natural thing to consider was discretizations for the Wasserstein space. Reading up on optimal transport and Wasserstein geometry, we thought about discretising the differential equations involved, in particular the Benamou-Brenier formulation for the Wasserstein distance. This became a side topic for us and led to interesting equations. Unfortunately, we later found out that a similar discretization for the Wasserstein distance on graphs already existed [21] and, even worse, the main formulas derived for triangulated surfaces were also already given by Lavenant et al. [18]. We still want to present our own derivation for these equations, as we give some of the operators in more detail than the mentioned works and the considerations that follow in subsection 3.7 are indeed novel.
Finally, we thought about a possible application for the above mentioned distance and how to generate some experimental images computationally. Unfortunately, the computational implementation is not yet finished. Still, we included a last chapter in which we discuss the possible routes of numerical treatment and supplement some of the necessary formulas.

Outline of the Dissertation In section 2 we start with a short introduction to equip the reader with the necessary basics on optimal transport to understand the following parts. We first give the basic definitions and later focus on the relation of the Wasserstein distance to Riemannian manifolds. This section is strongly inspired by [2] and sometimes also refers to [41]. Coming from a differential geometry background, we do not give a comparable introduction to differential geometry, but instead refer the reader to standard literature (e.g. [19] and [20]).
Section 3 covers our treatment of the sideproject, in which we discretise the continuity equation and ultimately present the discrete Wasserstein space as a Riemannian manifold. As mentioned before, these formulas have already been derived by Lavenant et al. in [18]. However, our exposition is more detailed and the final part 3.7 of this section contains novel ideas regarding the boundary of the probability measure simplex.
Section 4 constitutes the main part of our contribution. In this part, we explain our idea of how to use the Wasserstein space over a bundle over a given surface to come up with interesting distances for vector fields. We go through different approaches, supporting each with possible intuitions. Moreover, we give two toy examples that recover distances for signed measured which partly can also be found in the literature [22]. Finally, we focus on one particular distance for vector fields that relates nicely to parallel transport and collect properties for it.
Lastly, section 5 contains an idea for an application. We discuss several possibilities for numerical treatment. For one treatment, we also work out the relevant formulas.

Related Work For discretizations of the Benamou-Brenier formula on triangulated surfaces, we already mentioned the work of Lavenant et al. [18]. They also use this approach to solve the dynamical Wasserstein problem. For graphs, a similar approach has been done by Li [21]. His approach is similar, but in his work any averaging of a density function is performed over edges instead of triangles. On a more general level, the approach we took is comparable to the approach of information geometry. For this, we refer the reader to the standard introduction [1].
Many other versions of optimal transport have been considered before. Ning and Georgiou [24] used the dual formulation of the Wasserstein-1-distance to define a distance for matrix-valued measures. Similar approaches for the primal Wasserstein-1-distance have been done to define distances for vector-valued measure ([14] and [3]). A regularized Wasserstein distance for matrix-valued functions has been considered in [33]. Ryu et al. [36] use yet another equivalent optimization problem for the Wasserstein-1-distance, which is sometimes called the *Beckmann*-problem (e.g. in [40] referring to a paper by Beckmann [5]) in order to give a Wasserstein distance for vector- and matrix-valued measures.
Brenier and Vorotnikov [9] introduced a distance for positive-definite matrix-valued measures using a generalisation of the Benamou-Brenier formula for such measures. In order to define another distance for vector-valued measures, Chen, Georgiou and Tannenbaum [12] considered a set of measures indexed by the vertices of a graph and came up with a modified Benamou-Brenier formula for this setting. Earlier work [11] by Georgiou and Tannenbaum

used the Lindblad equation from quantum mechanics to give a Wasserstein distance for positive-definite matrices with trace 1. These last two approaches are similar; Chen et al. later gave efficient algorithms for both approaches [13].
Optimal transport for differential forms was considered by Dacorogna et al. [16] and Brenier and Duan [8]. In both works, their definitions were again motivated by the Benamou-Brenier formulation for the Wasserstein distance.
For numerical aspects of optimal transport, we refer to the text [34] and for methods on the dynamical Wasserstein problem to [32]. For applications of the Wasserstein distance to image morphing, we refer the reader to [43], [39].

2 An Introduction to Optimal Transport

For self-containedness we give a brief section on optimal transport in which we summarize the basics needed in the following chapters. Most of the following little introduction is heavily inspired by the text [2] but some of it is also taken from [41] and we will indicate whenever that is the case. For a more detailed introduction see those two texts.
We start with the purely measure-theoretic definition of optimal transport.

2.1 Definitions

First, we need a cost function. In our case this will always be a distance in a metric space. In general much less is needed, in fact not even full continuity but only lower semi-continuity is needed. For more details on this, the reader may consult [41]. For our purposes, however, we will impose continuity of the cost function c. Also, we will need the notion of a Polish space which is a separable metric space and a common denizen in any topic involving topology and measure theory.

Definition 1. *Let X be a Polish space and μ and ν be two probability measures on X. Moreover, let $c : X \times X \to \mathbb{R}$ be a continuous cost function on X. Then the optimal transport cost between μ and ν is defined as*

$$C(\mu, \nu) := \inf_{\rho \in \Pi(\mu,\nu)} \int_{X \times X} c(x, y) \mathrm{d}\rho,$$

where the infimum is taken over the space $\Pi(\mu, \nu)$ which is the space of probability measures on $X \times X$ whose marginals are μ and ν, respectively. More precisely, any $\rho \in \Pi(\mu, \nu)$ will satisfy $(\mathrm{proj}_1)_{\#}\rho = \mu$ and $(\mathrm{proj}_2)_{\#}\rho = \nu$, where proj_1 and proj_2 denote the projections on the first and second factor of $X \times X$, respectively, and $(\cdot)_{\#}$ as usual denotes the push-forward of a measure.

Probability measures on $X \times X$ whose marginals are μ and ν are called *transport plans* and if an minimizer of the above problem exists it is called an *optimal transport plan.*
The conditions on c guarantee that a minimizer for the above infimum always exists. More precisely it can be shown that as long as c is lower semi-continuous and bounded from below, a minimizer exists.
We want to continue by listing some relevant theorems regarding the optimal transport cost C without giving any proofs.
For the above minimization problem there does exist a dual formulation known as the Kantorovich duality. We will state it as a proposition.

Proposition 1. *For the optimal transport problem there is the following dual formulation*

$$C(\mu, \nu) = \sup_{\phi, \psi} \int_X \phi \mathrm{d}\mu + \int_X \psi \mathrm{d}\nu,$$

where the supremum is taken over all $\phi \in L^1(\mu), \psi \in L^1(\nu)$ that satisfy

$$\phi(x) + \psi(y) \leq c(x, y).$$

Sometimes this is also formulated with a minus sign between the integrals and in the condition on ϕ and ψ (as we will see below for the 1-Wasserstein distance).
We can now give the definition of the Wasserstein distance.

Definition 2. *Consider a polish space (X, d). For $p \geq 1$ the p-Wasserstein-distance is defined as*

$$W_p(\mu, \nu) := \left(\inf_{\rho} \int_{X \times X} d^p(x, y) \mathrm{d}\rho(x, y) \right)^{1/p}$$

Moreover, we define the space $\mathcal{P}_p(X)$ as the space of measures with finite p-moment, i.e. $\mu \in \mathcal{P}_p(X)$ is equivalent to $\int_X d^p(x_0, x) \mathrm{d}\mu(x) < \infty$ (this definition does not depend on the choice of x_0). As the name suggests the Wasserstein distance is indeed a distance on $\mathcal{P}_p(X)$ for any $p \geq 1$ (again we omit the proof) and we call $(\mathcal{P}_p(X), W_p)$ the Wasserstein space. Note that for compact X, the space $\mathcal{P}_p(X)$ is just the space of all probability measures $\mathcal{P}(X)$.
For $p = 1$ the dual formulation takes on a special form. It can be shown that the minimum of the optimal transport problem is attained for $\psi = -\phi$ and for ϕ (and thus ψ) being a 1-Lipschitz function. This is summarised in the next proposition.

Proposition 2. *The Wasserstein 1-distance satisfies the following relation.*

$$W_1(\mu, \nu) := \sup_{\phi} \int_X \phi \mathrm{d}(\mu - \nu),$$

where the supremum is taken over all 1-Lipschitz functions ϕ.

2.2 Inheritance

If we start from a metric space (X, d) and consider its Wasserstein space $(\mathcal{P}_p(X), W_p)$, then $(\mathcal{P}_p(X), W_p)$ often inherits various regularity properties of (X, d). For example, one can show that if (X, d) is a Polish space, so is $(\mathcal{P}_p(X), W_p)$ as well. We will go through two such situations in detail and alongside cover some of the most important theorems that will also be used in this work. For simplicity we will restrict the following to the coverage of only the Wasserstein-2-distance and occasionally make remarks about the Wasserstein-1-distance or the general Wasserstein-p-case.
Geodesic Spaces If one wants to discuss arguments on shortest paths connecting two points and parametrisations of curves on metric spaces in a similar manner as in a Riemannian manifold, one needs to introduce those concepts first which is a non-trivial task. A classic way to do this is the notion of *constant speed geodesics*. To a geometer, however, this name might be somewhat confusing and in their language translates better to 'constant speed shortest path'.

Definition 3. *Consider a metric space (X, d) and two points $p, q \in X$. A curve $\gamma : [0, 1] \to X$ with $\gamma(0) = p, \gamma(1) = q$ is called a* constant speed geodesics *between p and q if it satisfies the following condition*

$$d(\gamma(t), \gamma(s)) = |t - s| d(\gamma(0), \gamma(1)) \quad \textit{for any } t, s \in [0, 1].$$

The space (X, d) itself is called a geodesic space *if for any $p, q \in X$ there exists at least one constant speed geodesic connecting them.*

To explain some intuition behind the Wasserstein distances, we will discuss briefly how constant speed geodesics can look like in the Wasserstein space. For this consider a geodesic space (X, d) and two points $p, q \in X$ as well as a constant speed geodesic γ between them. We now want to try to connect the Dirac measures δ_p and δ_q in the the Wasserstein-2-space. We define the measure curve $\mu_t := \delta_{\gamma(t)}$ and see that this is a constant speed geodesic since $W^2(\delta_x, \delta_y) = d(x, y)$ for any $x, y \in X$. In fact, all possible geodesics connecting δ_p and δ_q must be of this kind or combinations originating from different constant speed geodesics between p and q.
Let us now consider another curve that is not a constant speed geodesic, namely the linear interpolation $\nu_t := (1-t)\delta_p + t\delta_q$. We can check that $W_2(\nu_t, \nu_s) = \sqrt{|t-s|}d(p, q)$ since the optimal transport plan between $(1-t)\delta_p + t\delta_q$ and $(1-s)\delta_p + s\delta_q$ is clearly the one that fixes the common mass of both measures at p and q and only moves the difference of $|t-s|$ from p to q (or the other way) and therefore we get for the distance $(|t-s|d^2(p, q))^{1/2}$. Thus, this does not yield a constant speed geodesic. In fact by the very definition of the length of a curve in a metric space this would be a curve of infinite length, though still continuous.
We see that one can think of the Wasserstein distance as an extension of the usual concept of distance to a setting where we allow for fuzzy points, i.e. a setting where we potentially do not have complete knowledge of the points' location but we still have a distance. However, in case we do have the knowledge of the location of the point (i.e. the measure is given to us as a Dirac measure), then we just recover the same distance from the underlying base space.
Finally, we want to note that the above comparison of a curve moving a point on the base space (displacement interpolation) versus a curve resulting from linear interpolation actually works for any $p > 1$ and one would get the distance $W_p(\nu_t, \nu_s) = |t-s|^{1/p}d(p, q)$ in the case of linear interpolation. For $p = 1$, however, the linear interpolated curve is a constant speed geodesic as well. So in that case the geodesics can also do some kind of teleporting and the above intuition does not fully apply. That is why in this work we usually prefer to work with the Wasserstein-2-distance over the Wasserstein-1-distance.
Numerically, the problem of finding a Wasserstein geodesic is referred to as the dynamical Wasserstein problem, whereas finding an optimal transport plan is referred to as the static Wasserstein problem.
Next, we can ask the question whether the Wasserstein space over a geodesic space is also a geodesic space and, as indicated above, this is indeed the case.

Theorem 1. *If (X, d) is a Polish and geodesic space, then $(\mathcal{P}_2(X), W_2)$ is also a geodesic space.*

The following theorem will again provide us with a bit more intuition for these constant speed geodesics in Wasserstein space.

Theorem 2. *Assume that (X, d) is a Polish and geodesic space and consider the Polish and geodesic space $(\mathcal{P}_2(X), W_2)$. Then the following are equivalent:*

- *The curve μ_t is a constant speed geodesic between μ_0 and μ_1.*
- *There is a measure $\tilde{\gamma}$ on the space of geodesics such that $(e_0, e_1)_{\#}\tilde{\gamma}$ is an optimal transport plan between μ_0 and μ_1 and $\mu_t = (e_t)_{\#}\tilde{\gamma}$.*

Here e_t denotes the evaluation map: $e_t(\gamma) = \gamma(t)$.

This theorem offers change of perspective. When considering a curve in the space of measures, we might have something like a wandering cloud in mind. The second formulation though makes us think of the measure as consisting of multiple particles, each of which is travelling along an optimal path (i.e. a shortest geodesic) and the totality of them is transported in an optimal way. The only thing that needs to be adjusted is at the initial moment when each element is assigned its determination. That this is done correctly, is insured in the condition that $(e_0, e_1)_{\#}\tilde{\gamma}$ is an optimal transport plan between μ_0 and μ_1. As an example consider the four vertices $(0,0), (0,1), (1,0), (1,1)$ of a square in Euclidean space and consider the measures $\mu_0 = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(0,1)}$ and $\mu_1 = \frac{1}{2}\delta_{(1,0)} + \frac{1}{2}\delta_{(1,1)}$. Clearly an optimal path between them is given by the curve $\mu_t = \frac{1}{2}\delta_{(t,0)} + \frac{1}{2}\delta_{(t,1)}$ and visually we can see the point $(0,0)$ move into the point $(1,0)$ along the curve $\gamma(t) = (t, 0)$ and the point $(0,1)$ move into the point $(1,1)$ along the curve $\eta(t) = (t, 1)$. However, an equivalent view is given by considering the measure on the space of geodesics $\tilde{\gamma} = \frac{1}{2}\delta_{\gamma} + \frac{1}{2}\delta_{\eta}$ which satisfies the properties of the above theorem. In contrast now we might also consider the constant speed geodesics $\alpha(t) = (t, t)$ and $\beta(t) = (t, 1-t)$ which move the point $(0,0)$ to $(1,1)$ and $(0,1)$ to $(1,0)$, respectively. For them, $\tilde{\alpha} = \frac{1}{2}\delta_{\alpha} + \frac{1}{2}\delta_{\beta}$ indeed satisfies that the marginals $(e_0)_{\#}\alpha = \boldsymbol{\mu_0}$ and $(e_1)_{\#}\alpha = \boldsymbol{\mu_1}$. However, $(e_0, e_1)_{\#}\alpha$ is not an optimal transport plan since it is the transport plan that transports $(0,0)$ to $(1,1)$ and $(0,1)$ to $(1,0)$.

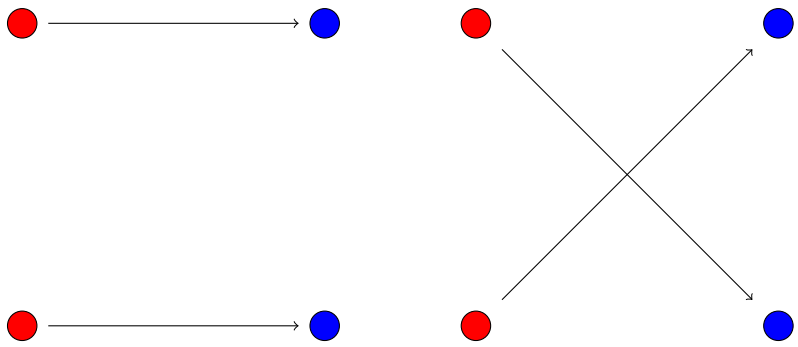

Figure 1: A measure with equal weights on the two left nodes gets transported to a measure with equal weights on the two right nodes. The first is a constant speed geodesic while the second is only locally a constant speed geodesic.

Smooth Manifolds Let us briefly revisit the previous theorem and compare the general non-deterministic setting with the deterministic setting of a Dirac measure. When not

dealing with measures we just have a geodesic which has for each time a velocity $\dot{\gamma}$. In the measure case instead we have a whole measure on the set of geodesics. As it turns out there is still a notion of velocity (for any time), though in this case this will be a vector field. However, there is a question about the regularity of the measure curve that we have to address first. Therefore, let us start by giving the definition of an absolutely continuous curve.

Definition 4. *A curve γ in a metric space (X, d) is called an* absolutely continuous curve *if there exists a function $f \in L^1(0,1)$ such that for all $0 \leq t < s \leq 1$ it holds that*

$$d(\gamma(t), \gamma(s)) \leq \int_t^s f(x)\mathrm{d}x.$$

Equivalently, γ is absolutely continuous *if the* metric derivative

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

exists almost everywhere and is a L^1-function.

From now on let us work on a Riemannian manifold (M, g) instead of just a metric space (X, d). It turns out that for absolutely continuous curves in $(\mathcal{P}_2(M), W_2)$, a velocity vector field exists and the link between the curve and the vector field is given by the so-called *continuity equation*, which is the topic of the following theorem.

Theorem 3. *Consider a curve $\mu_t \subset \mathcal{P}_2(M)$. Then the following are equivalent:*

(i) For the curve μ_t there exists a Borel family of vector fields v_t such that $\int_0^1 |v_t|_{L^2(\mu_t)}\mathrm{d}t < \infty$ and μ_t and v_t satisfy the continuity equation

$$\frac{\partial}{\partial t}\mu_t + \nabla \cdot (v_t \mu_t) = 0$$

in the sense of distributions, i.e. for any smooth and compactly supported test function ϕ, we have

$$\frac{\partial}{\partial t} \int \phi \mathrm{d}\mu_t = \int \langle \nabla \phi, v_t \rangle \mathrm{d}\mu_t.$$

(ii) μ_t is absolutely continuous in the metric sense described above.

The continuity equation plays a central role in the study of Wasserstein distances and it will also be revisited multiple times in this work.
The following theorem covers the famous Benamou-Brienier-formula. It shows that a certain methodology from metric spaces can still be applied to this setting. Moreover, it motivates the definition of the notion of a tangent space for the Wasserstein space.

Theorem 4. *Consider two measures μ^0, μ^1 on the Riemannian manifold M. Then, the Wasserstein distance between them can be expressed via the infimum of measure curves μ_t connecting them and satisfying the continuity equation:*

$$W_2(\mu^0, \mu^1) = \inf \left\{ \int_0^1 \|v_t\|_{L^2(\mu_t)} \mathrm{d}t \quad | \quad \frac{\partial}{\partial t}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \right\}$$

where the infimum is taken over all pairs (μ_t, v_t) which satisfy the continuity equation.

This theorem is intuitive in a way since it mirrors a basic situation from classic differential geometry where the distance between two points can be recovered as an infimum over the lengths of all curves connecting them. The length can be represented by integrating the norm of the derivative of the curve in question. Therefore, v_t should be considered as an analogue to the speed of the curve μ_t.
When discussing the length functional in differential geometry, one will usually also discuss the energy functional for a curve which relates to the length functional in the sense that they have the same minimizers. The following version of the Benamou-Brenier formula reflects that and can be found in [41].

Theorem 5. *With the same assumptions as in the previous theorem we also have yet another very similar form for the Wasserstein-distance between μ^0 and μ^1:*

$$W_2(\mu^0, \mu^1)^2 = \inf \left\{ \int_0^1 \|v_t\|^2_{L^2(\mu_t)} \mathrm{d}t \quad | \quad \frac{\partial}{\partial t}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \right\}$$

This formula reflects a formula from classical differential geometry, namely

$$d^2(x, y) = \inf \int |\dot{\gamma}|^2 \mathrm{d}t.$$

In fact, there is an even deeper intuitive connection: Recall, that a Wasserstein geodesic can also be represented as a measure on the space of geodesics on M, where intuitively each geodesic tells us a path along which a particle is transported. The vector fields v_t at a point p can then be thought of as the velocity of the curve (or particle) moving through p at time t and the Flux at that point and time would correspond to the infinitesimal quantity $v_t(p)\mathrm{d}\mu_t(p)$. The reinterpretation of the curve in the measure space as a measure on the geodesics then corresponds formally to 'sorting' this integral by the curves and thus obtain for the term in the infimum the alternative form

$$\int_0^1 \int_M |v_t|^2 \mathrm{d}\mu_t \mathrm{d}t = \int_\gamma \mathcal{E}(\gamma) \mathrm{d}\mu(\gamma),$$

where $\mathcal{E}(\gamma) = \int_0^1 |\dot{\gamma}|^2 \mathrm{d}t$ denotes the usual curve Energy without the factor $\frac{1}{2}$. This yields an alternative formula

$$W_2(\mu, \nu)^2 = \inf_{\tilde{\gamma}} \{\mathbb{E}_{\tilde{\gamma}} \mathcal{E}(\gamma)\},$$

where the infimum is taken over all measures $\tilde{\gamma}$ on the space of geodesics in M with $(e_0)_\#\tilde{\gamma} = \mu$ and $(e_1)_\#\tilde{\gamma} = \nu$, where e_0 and e_1 denote the evaluation functions at times 0 and 1. This is actually true and made precise in [41].
We want to emphasize the intuitive meaning of this: The Wasserstein distance is the distance for which a constant speed geodesic will transport particles along paths that minimize the energy $\mathcal{E}$ which in turn depends only on the metric of M.
These last two theorems strongly indicate that we can construct a tangent space to $\mathcal{P}_2(M)$, that the tangent vectors will be the vector fields v_t and that on this tangent space we can use the $L^2(\mu_t)$-product as a metric. However, there is one caveat, since the vector fields v_t involved in the continuity equation of μ_t are not unique. Instead, we can introduce μ_t-divergence-free vector fields as vector fields w_t satisfying

$$\nabla \cdot (w_t \mu_t) = 0$$

in the sense of distributions, i.e.

$$\int \nabla\phi \cdot w_t \mathrm{d}\mu_t = 0$$

for any smooth and compactly supported function ϕ.
When adding a μ_t-divergence-free vector field w_t to v_t, the vector fields $v_t + w_t$ will also satisfy the continuity equation for μ_t. Thus, it turns out that the best concept for a tangent space to $P(M)$ consists of those vector fields v_t, which are orthogonal to the μ_t-divergence-free ones. This, on the other hand can also be described as the image of the gradient or more precisely as the $L^2(\mu_t)$-completion of that. Let us summarize the definition.

Definition 5. *For a compact Riemannian manifold M consider its space of probability measures $\mathcal{P}(M)$. Then for any $\mu \in \mathcal{P}(M)$ we may define the* tangent space *of $\mathcal{P}(M)$ at μ by*

$$\begin{aligned} T_\mu\mathcal{P}(M) &= \{v \in L^2(\mu) : \int \langle v, w\rangle \mathrm{d}\mu = 0 \text{ for all } w \text{ with } \nabla \cdot (w_t\mu_t) = 0\} \\ &= \overline{\{\nabla\phi : \phi \in C_c^\infty(M)\}}^{L^2(\mu)}. \end{aligned}$$

In this definition $L^2(\mu)$ denotes the space of square-integrable vector fields (with respect to the measure μ), i.e. vector fields v satisfying $\int |v|^2 \mathrm{d}\mu < \infty$.

This concept of a tangent space allows for some interesting computations. We want to present one particular computation due to Felix Otto. However, this is only a formal computation and we also consider a simplified setting with a compact manifold.
Let μ_t denote a curve in the Wasserstein space. Moreover, assume that this curve is given by a density at all times

$$\mu_t = \rho_t \mathrm{dvol}$$

where dvol denotes the Riemannian volume measure on the Riemannian manifold M. Then, for such measures the *entropy* can be defined as

$$\mathrm{E}(\mu) = \mathrm{E}(\rho) := \int_M \rho \log(\rho) \mathrm{dvol}.$$

This is a functional on the space of measures. In analogy with classical differential geometry, we can try to define a gradient for $\mathrm{E}(\rho)$ as a vector field $\nabla_\rho \mathrm{E}$ satisfying

$$\langle \nabla_\rho \mathrm{E}, v_t \rangle_\rho = \frac{\partial}{\partial t}_{|t=0} \mathrm{E}(\rho_t)$$

for any curve ρ_t such that (ρ_t, v_t) satisfies the continuity equation and $\rho_0 = \rho$ if such a vector field exists. As usual, $\langle v, w \rangle_\rho = \int \langle v, w \rangle \rho \mathrm{dvol}$.
Interestingly we can actually compute the above gradient:

$$\begin{aligned}
\int_M \langle \nabla_\rho \mathrm{E}, v_t \rangle \rho \mathrm{dvol} = \langle \nabla_\rho \mathrm{E}, v_t \rangle_\rho &= \frac{\partial}{\partial t}_{|t=0} \mathrm{E}(\rho_t) \\
&= \frac{\partial}{\partial t}_{|t=0} \int_M \rho \log(\rho) \mathrm{dvol} \\
&= \int_M \dot{\rho}_t \log(\rho_0) \mathrm{dvol} + \int_M \dot{\rho}_t \mathrm{dvol} \\
&= \int_M \dot{\rho}_t \log(\rho_0) \mathrm{dvol} + \underbrace{\frac{\partial}{\partial t}_{|t=0} \int_M \rho_t \mathrm{dvol}}_{=0} \\
&= \frac{\partial}{\partial t}_{|t=0} \int_M \rho_t \log(\rho_0) \mathrm{dvol}.
\end{aligned}$$

This last term now is just the first term from the continuity equation for the test function $\phi = \log(\rho)$. Thus, the computation continues

$$\frac{\partial}{\partial t}_{|t=0} \int_M \rho_t \log(\rho_0) \mathrm{dvol} = \int_M \langle \nabla \log(\rho), v_0 \rangle \rho_0 \mathrm{dvol}$$

and comparing this to the term we started with yields

$$\nabla_\rho \mathrm{E} = \nabla \log(\rho) = \frac{1}{\rho} \nabla \rho.$$

Now, Otto considered a gradient flow for the entropy. More precisely, he considered a measure curve μ_t satisfying

$$\dot{\mu}_t = \nabla_{\mu_t} \mathrm{E}.$$

Note that a gradient flow is just the smooth version of gradient descent or in other words: if we discretize the time parameter in the gradient flow equation, we recover the gradient descent method.

Now let us compute what gradient flow means in terms of densities. Above equation tells us that μ_t and the vector field $\nabla_{\mu_t} \mathrm{E} = \frac{1}{\rho}\nabla\rho$ satisfy the continuity equation. Thus, we have

$$\frac{\partial}{\partial s}_{|s=t} \int_M \phi\rho_s \mathrm{dvol} - \int_M \langle \nabla\phi, \frac{1}{\rho_t}\nabla\rho_t\rangle \rho_t \mathrm{dvol} = 0.$$

This last equation can be reformulated to

$$\begin{aligned}\int_M \phi \cdot \dot{\rho}_t \mathrm{dvol} &= \int_M \langle \nabla\phi, \nabla\rho_t\rangle \mathrm{dvol} \\ &= -\int_M \phi \cdot \mathrm{div}(\nabla\rho_t)\mathrm{dvol} \\ &= -\int_M \phi\Delta\rho_t \mathrm{dvol}\end{aligned}$$

for any test function $\phi \in C_c^\infty$. Thus, we may conclude that the density for any measure curve that follows a gradient flow for the entropy functional will satisfy the heat equation $\dot{\rho}_t + \Delta\rho_t = 0$. A remarkable result.

3 A Piecewise Linear Version of Dynamical Optimal Transport

In this section we introduce an interesting way to study a piecewise linear version of Wasserstein spaces. As mentioned in the introduction the metric we derive was already given by Lavenant et al. [18] but the contents of subsection 3.7 are novel. A very similar thing had been done by Li [21] for graphs.
The core of this discretization will be the Benamou-Brenier formulation for the Wasserstein distance. We start by setting up some standard notation for piecewise linear analysis before giving a discrete version of the continuity equation and the Benamou-Brenier formulation. Finally, we construct a discrete Wasserstein distance from that.

3.1 Notation

In this part we introduce the basic notation for the piecewise linear setting. Let M be a triangle mesh (V, E, F) with vertices V, edges E and triangles (or faces) F. Consider M to be a triangulation of a compact and oriented surface without boundary. In the following, we will often sum over all vertices, edges or triangles which are part of or contain a certain vertex, edge or triangle. We call all of these instances *adjacency* and use the notation $v \in ad(e), v \in ad(f), e \in ad(f)$ whenever a vertex or edge is part of an edge or triangle and also the notation $f \in ad(e), f \in ad(v), e \in ad(v)$ whenever a triangle or edge contains an edge or vertex.
For simplicity, we consider our triangle mesh as embedded into $\mathbb{R}^3$ (equipped with the standard metric) in a way such that all edges are straight lines. Then we inherit a flat Riemannian metric on each triangle and in particular for every triangle f we can compute its area $area(f)$ and also its angles. Moreover, we can integrate functions over the triangles and integration of a function on the whole simplex is given by integrating over each triangle individually and then summing over all triangles.
Also, we denote the functions from a finite set S to the reals by $\mathcal{F}(S)$. The functions on the vertices for example are denoted by the term $\mathcal{F}(V)$ and on them there is a standard basis given by $\{\phi_v : v \in V\}$ where the $\phi_v(u) := \delta_{v,u}$ for all $v, u \in V$ are called basic functions in the literature (e.g. [7]).
In a discretization a natural thing would be to have the measures live on the triangles and continuous functions being discretized as living on vertices. However, for the discretization of the continuity equation it will be more useful to have measures live on vertices as well. Therefore, we think of densities instead of measures, i.e. density functions as well as usual functions are always elements of $\mathcal{F}(V)$. As usual in piecewise linear settings, we assign to functions on vertices piecewise linear functions via linear interpolation on each triangle. We will use the same notation for the function on vertices, the vertex-indexed vector or the piecewise linear function since it should be clear from the context which one is meant. Since the mesh is embedded in $\mathbb{R}^3$ we can integrate these (piecewise linear) functions over the mesh as described above. In particular, the (discrete) integral of the density ρ is then

given by

$$\sum_{f\in F}\frac{1}{3}\sum_{v\in ad(f)}\rho(v)area(f)=\sum_{v\in V}\rho(v)\frac{1}{3}\sum_{f\in ad(v)}area(f).$$

Thus, we can define $A(v):=\frac{1}{3}\sum_{f\in ad(v)}area(f)$ and then the above integration can also be understood as a sum over vertices $\sum_{v\in V}\rho(v)A(v)$.
We then define the space of probability densities as the space

$$\mathcal{P}(M):=\{\rho\in\mathcal{F}(V)|\rho(v)\geq 0 \text{ for all } v\in V \text{ and } \sum_{v\in V}\rho(v)A(v)=1\},$$

where the second condition is just the discrete analogue of $\int\rho=1$.
Another quantity that will show up several times in the upcoming computations is $A_\rho(f):=\frac{1}{3}\sum_{v\in ad(f)}\rho(v)$. This comes from sorting above sums by the triangles and represents a triangle-based (averaged) version of ρ.

3.2 Discrete Continuity Equation

Next, we consider the continuity equation $\frac{\partial}{\partial t}\mu_t+\nabla\cdot(v_t\mu_t)=0$. Since we work with absolutely continuous measures or more precisely, their densities, we get the following form for the continuity equation:

$$\frac{\partial}{\partial t}\rho_t+\nabla\cdot(\rho_t v_t)=0.$$

Now, we discretize as follows: Let $\rho_t\in\mathcal{F}(V)$ be a time-dependent density (i.e. an element of $\mathcal{P}(M)$ for any time $t\in[0,1]$), $\phi\in\mathcal{F}(V)$ a test-function also living on vertices and X be a function assigning to each face a vector tangent to that face (i.e. a vector field that is constant per face). As described above, functions in $\mathcal{F}(V)$ (that is ρ_t and ϕ) shall be understood here as piecewise linear functions on the whole mesh M by linear extension on each face. Then, we can integrate:

$$\frac{\partial}{\partial t}\int\phi\rho_t+\int\phi div(\rho_t X)=0. \tag{1}$$

On the left hand side we encounter the term $\int\phi\rho_t$. Recall that this is literally an integral over the mesh where ϕ and ρ_t are linear functions on each triangle.
If we adopt a vectorial notation for ϕ and ρ_t (i.e. ϕ and ρ_t are interpreted as vectors indexed by the set of vertices instead of piecewise linear functions), then in terms of the input vectors $\phi\in\mathbb{R}^V$ and $\rho_t\in\mathbb{R}^V$ this integral is just a bilinear form and can be written as $\phi^tM\rho_t$ for a so-called mass-matrix. The mass-matrix is a well-known object in discrete differential geometry and numerics and is explicitly given by

$$M_{uv}=\begin{cases}\sum_{f\in ad(u)}area(f)/6, & u=v,\\ \sum_{f\in ad(uv)}area(f)/12, & u\sim v,\\ 0, & \text{else}\end{cases}$$

for any pair of vertices $u, v \in V$ (see for example [7]). Note, that this matrix just represents the L^2-inner product on the simplex restricted to piecewise linear functions and is therefore invertible.
However, for the right hand side of (1) we need to come up with a discretization of the term $div(\rho X)$, which for a discrete function and a discrete vector field shall yield a discrete function.

A Discrete ρ-divergence

In order to make sense of the divergence term in (1) we use a weak formulation approach.

$$\int < \nabla\phi, X > \rho = \int \phi div(\rho X). \tag{2}$$

Here, the integration is performed, as described above, on each triangle individually and then summed up over the triangles. In particular taking the scalar product only needs to be understood on each triangle individually and the gradient is the usual euclidean gradient on any (flat) triangle applied to the piecewise linear function ϕ, yielding a constant vector per triangle.
Also note that this time ϕ is the name of the test-function. The approach shall yield the definition of the operator $div(\rho\cdot)$: vector fields $\to \mathcal{F}(V)$.
Recall that $div(\rho X)$ should be a function on vertices. Then the right hand side of (2) is $\phi^t M div(\rho X)$ in vector-matrix notation, where M is the above mentioned mass matrix.
On the left hand side we have the term $< \nabla\phi, X >$ which is constant per face. In order to compute the gradient of ϕ note that for a basic function ϕ_u one has $\nabla\phi_u = e^\perp/(2area(f))$ on a face $f \in ad(u)$, where $e^\perp$ is the vector pointing from u to its opposing edge e in the triangle f which is orthogonal to e and also of the same length as e. Therefore, on a triangle f we have

$$\nabla\phi = \sum_{u \in ad(f)} \phi(u) e^\perp_{u,f}/2area(f) = \frac{1}{2area(f)} \sum_{u \in ad(f)} \phi(u) e^\perp_{u,f}$$

where $e^\perp_{u,f}$ in accordance with the above notation refers to the vector pointing from u to its opposing edge e in the triangle f which is orthogonal to e and also of the same length as e. Thus, the integral over one triangle f is

$$\begin{aligned}\int_f < \nabla\phi, X > \rho =& \frac{1}{2area(f)} \sum_{u \in ad(f)} \phi(u) < e^\perp_{u,f}, X > \cdot \frac{1}{3} area(f) \sum_{v \in ad(f)} \rho(v) \\ =& A_\rho(f) \sum_{u \in ad(f)} \phi(u) \frac{1}{2} < e^\perp_{u,f}, X > .\end{aligned}$$

Therefore, for the integral over the whole mesh we get

$$\begin{aligned}\int < \nabla\phi, X > \rho =& \sum_f A_\rho(f) \sum_{u \in ad(f)} \phi(u) \frac{1}{2} < e^\perp_{u,f}, X(f) > \\ =& \sum_u \phi(u) \sum_{f \in ad(u)} A_\rho(f) \frac{1}{2} < e^\perp_{u,f}, X(f) > \\ =& \phi^t \tilde{L} X,\end{aligned}$$

where

$$\tilde{L}_\rho(u,f) := \begin{cases} A_\rho(f) \frac{1}{2} e^\perp_{u,f}, & \text{for } u \in ad(f) \\ 0, & \text{else} \end{cases}$$

defines a matrix $\tilde{L}_\rho := A_\rho(f) \frac{1}{2} e^\perp_{u,f} \in (\mathbb{R}^3)^{V \times F}$ indexed by the vertices and the faces of the mesh whose entries at indices (u, f) are vectors tangent to the face f. A multiplication of this matrix with a discrete vector field $X \in (\mathbb{R}^3)^F$, which is given by a vector indexed by the triangles and which has vectors tangent to the triangles as entries, is defined by $\tilde{L}_\rho \cdot X = \sum_{f \in F} A_\rho(f) \frac{1}{2} e^\perp_{u,f} \cdot X(f)$.
Now, comparing the discrete notions for the left hand side and the right hand side yield the discrete divergence operator as $div(\rho X)(u) = (M^{-1} \tilde{L}_\rho X)_u$. In other word the divergence can be identified with the matrix $M^{-1} \tilde{L}_\rho$.
A nice remark is that the matrix $\tilde{L}$ can be written as

$$\tilde{L} = E \cdot \mathrm{Diag}(A_\rho(f)),$$

where $E(u, f) = \frac{1}{2} e^\perp_{u,f}$ is the matrix that corresponds to the usual divergence in the literature for piecewise linear Laplacians and finite elements. Multiplying a vector field with the matrix $\mathrm{Diag}(A_\rho(f))$ can be seen as a discrete analogue of multiplying a smooth vector field with a smooth function ρ. Thus, the matrix $\tilde{L}$ compares nicely to its smooth version $div(\rho \cdot)$ in the sense that multiplying it to a vector field X consists of an analogue of a multiplication with ρ (the matrix $\mathrm{Diag}(A_\rho(f))$) and then an analogue to the application of the divergence (the matrix E).

3.3 ρ-divergence on im(∇)

As we will see later, it is important to understand the ρ-divergence in the case that X is a gradient thus yielding a discrete version of a ρ-Laplacian. If X is the gradient of a piecewise linear function ψ (living on the vertices) then $\int < \nabla\phi, X > \rho = \int < \nabla\phi, \nabla\psi > \rho$ can be further computed.
A well-known formula from linear finite elements for $< \nabla\phi, \nabla\psi >$ on a triangle f is given by

$$< \nabla\phi, \nabla\psi >= \frac{1}{2area(f)} \sum_{e \in ad(f)} \cot(\alpha_{e,f}) \delta\phi(e) \delta\psi(e),$$

where $\alpha_{e,f}$ is the angle of the triangle f at the vertex opposite to the edge e in the triangle f. For the definition of $\delta\phi$, which is a function on the edges of the mesh one usually introduces a so-called orientation on the edges first, so that then $\delta\phi(uv) = \phi(u) - \phi(v)$ or $\delta\phi(uv) = \phi(v) - \phi(u)$ depending on the orientation chosen on the edge uv. However, in this formula, the chosen orientation does not matter since the term $\delta\phi(uv)\delta\psi(uv) = (\phi(u)-\phi(v))(\psi(u)-\psi(v)) = (\phi(v)-\phi(u))(\psi(v)-\psi(u))$ does not depend on the orientation of the edge uv. Therefore, the integral over a triangle f is given by

$$\begin{aligned}\int_f < \nabla\phi, \nabla\psi > \rho =& \frac{1}{2area(f)} \sum_{e\in ad(f)} \cot(\alpha_{e,f})\delta\phi(e)\delta\psi(e) \cdot \frac{1}{3} area(f) \sum_{v\in ad(f)} \rho(v) \\ =& \sum_{e\in ad(f)} \frac{1}{2} \cot(\alpha_{e,f})\delta\phi(e)\delta\psi(e) \cdot A_\rho(f).\end{aligned}$$

So, for the total integral we get

$$\begin{aligned}\int < \nabla\phi, X > \rho =& \sum_f \Big(\sum_{e\in ad(f)} \frac{1}{2} \cot(\alpha_{e,f})\delta\phi(e)\delta\psi(e) \cdot A_\rho(f)\Big) \\ =& \sum_{u\sim v} \frac{1}{2}\Big(\sum_{f\in ad(uv)} \cot(\alpha_{uv,f}) A_\rho(f)\Big)(\phi(u) - \phi(v))(\psi(u) - \psi(v)) \\ =& \phi^t \Delta_\rho \psi,\end{aligned}$$

where $\Delta_\rho(u,u) = \sum_{v:v\sim u} \omega_{uv}$ and $\Delta_\rho(u,v) = -\omega_{uv}$ if $u \sim v$, where

$$\omega_{uv} = \frac{1}{2}\Big(\sum_{f\in ad(uv)} \cot(\alpha_{uv,f}) A_\rho(f)\Big).$$

Again, by comparing right hand sides and left hand sides we thus find that the discrete ρ-divergence of the gradient of a function ψ is given by $div(\rho\nabla\psi) = M^{-1}\Delta_\rho\psi$. Clearly, Δ_ρ is a sort of discrete distorted Laplace (or more precisely $M^{-1}\Delta_\rho$ is a discrete Laplace operator which is self-adjoint with respect to the scalar product M and whose induced bilinear and symmetric form is given by the matrix Δ_ρ). Smooth versions of such Laplacians in the context of Wasserstein geometry were already considered in the literature [41].

3.4 Properties of the ρ-divergence

In this part we want to analyse the image and the kernel of the operator $div(\rho\cdot)$. To ease the analysis let us for a moment impose the additional condition that the density ρ has has only positive entries $\rho(v) > 0$ for all v. Now, let us again have a look at the weak formulation (2) of the ρ-divergence $\int < \nabla\phi, X > \rho = \int \phi div(\rho X)$. The left hand side becomes zero if we plug in a constant function for ϕ. In particular, this means that $div(\rho X) \perp \mathbb{1}$ for any vector field X, where $\mathbb{1}$ denotes the constant one-function on vertices $\mathbb{1}(v) = 1$ for all $v \in V$.

On the other hand, if ϕ is not constant, there is a face f on which $\nabla\phi$ is not zero. By choosing X appropriately we can make the left hand side non-zero. In particular, $\int \phi div(\rho X)$ is zero for all X if and only if ϕ is constant. Thus, we can conclude that $(im(div(\rho\cdot)))^\perp = \mathbb{R}\mathbb{1}$ (and therefore $im(div(\rho\cdot)) = (\mathbb{R}\mathbb{1})^\perp$) where the orthogonality comes from the L^2-inner product $\int \phi \cdot div(\rho X)$ which is given by the mass-matrix M when we think of ϕ and $div(\rho X)$ as vertex-indexed vectors.
To make this set of arguments more precise, let us fix two scalar products. On functions $\mathcal{F}(V)$ we use the scalar product $< f, g >= \int fg$. On vector fields we use the scalar product $< X, Y >_\rho = \int < X, Y > \rho$ which is indeed a scalar product with our assumption from above ($\rho(v) > 0$ for all v). Then by (2), $div(\rho\cdot)$ is the adjoint of the usual gradient ∇. Therefore,

$$(\ker div(\rho\cdot))^\perp = im(\nabla).$$

However, the image of the gradient operator can be parametrized by the functions that are orthogonal to the constants. This is due to the fact that ∇ is a linear map and thus its image is isomorphic to its domain modulo its kernel. With our assumptions the kernel just consists of constant functions and using our scalar product such a quotient is isomorphic to the orthogonal complement of the kernel. Thus, we have $im(\nabla) \cong (\ker \nabla)^\perp$.
On the other hand, since the operators are adjoint we have that

$$im(div(\rho\cdot)) = (\ker \nabla)^\perp$$

and of course $(\ker \nabla)^\perp = (\mathbb{R}\mathbb{1})^\perp$. Thus, after restricting the image to $im(div(\rho\cdot))$ we get the following chain of maps:

$$(\ker \nabla)^\perp \overset{\nabla}{\cong} im(\nabla) \hookrightarrow vector fields \overset{div(\rho\cdot)}{\to} (\ker \nabla)^\perp.$$

Since

$$0 \to \ker div(\rho\cdot) \hookrightarrow vector fields \overset{div(\rho\cdot)}{\to} (\ker \nabla)^\perp \to 0$$

is an exact sequence and by the above we know that $(\ker div(\rho\cdot))^\perp = im(\nabla)$ we have that

$$(\ker \nabla)^\perp \overset{\nabla}{\cong} im(\nabla) \overset{div(\rho\cdot)}{\to} (\ker \nabla)^\perp$$

is actually a sequence of bijections/identification. Thus, under the right restrictions and identifications we can view $div(\rho\cdot)$ as a map $div(\rho\cdot) : (\mathbb{R}\mathbb{1})^\perp \to (\mathbb{R}\mathbb{1})^\perp$, which is a bijective.

3.5 Definition of the Metric

Recall the Benamou-Brenier formula from the introduction to the Wasserstein space. The following discussion there showed that people were able to provide a notion of a tangent space for the Wasserstein space as well as a Riemannian metric for it. Using the discrete

operators we just introduced one can do the same on triangulated surfaces and thus define a Wasserstein-Riemannian metric in the following way:
Note that the space $\mathcal{P}(M)$ is a simplex of dimension $|V|-1$ and therefore a Riemannian manifold in a natural way. A curve μ_t in $\mathcal{P}(M)$ needs to satisfy $\int \frac{\partial}{\partial t}\mu_t \cdot 1 = 0$. For a density curve ρ_t, satisfying $\rho_0 = \rho$ for some density ρ, this means that

$$\frac{\partial}{\partial t}\sum_{v\in V}\rho_t(v)A(v) = \sum_{v\in V}\dot{\rho}_t(v)A(v) = 0.$$

Clearly, the vector $\dot{\rho}_t$ is tangent to the simplex $\mathcal{P}(M)$. This vector, however, is not a discrete analogue to the vector fields considered in section 2. Thus, let us adopt an adhoc-notion and call all vectors $U \in \mathcal{F}(V)$ satisfying $\sum_{v\in V} U(v)A(v) = 0$ *simplex tangent vectors*. Note that this condition is equivalent to U being orthogonal to $\mathbb{1}$ as a function.
The discrete analogue for the tangent vectors to the tangent space of the Wasserstein space as described in section 2 are not tangent vectors $\dot{\rho}_t$ to the simplex. Instead, we use the discrete continuity equation $\int \phi\dot{\rho}_t + \int \phi div(\rho_t X) = 0$ to associate a discrete vector field X to $\dot{\rho}_t$ which is orthogonal to the ρ-divergence-free vector fields. In analogy with section 2 the discrete tangent space to the Wasserstein space is defined as the set of those vector fields equipped with the $L^2(\rho)$-metric. To distinguish these from the simplex tangent vectors, let us call the vector fields X, that are orthogonal to all ρ-divergence-free discrete vector fields, *Wasserstein tangent vectors*.
In summary: the discrete Wasserstein tangent space at a density ρ is given by Wasserstein tangent vectors, i.e. the discrete vector fields that are orthogonal to all ρ-divergence-free vector fields. The metric on them is given by the $L^2(\rho)$-metric $\int < X, Y > \rho$ for two Wasserstein tangent vectors X, Y.

3.6 Computing the Metric

Instead of giving a Wasserstein distance directly for our setting, we can try to compute this Riemannian metric in the coordinates given by the simplex tangent vectors. More precisely: For two given simplex tangent vectors U and V, what is the Wasserstein-Riemannian metric of their associated Wasserstein tangent vectors?
In accordance with the the non-discrete case we first have to solve the continuity equation $\dot{\rho}_t = div(\rho X)$ with an X that is orthogonal to the ρ-divergence free vector fields (recall the definition of the tangent space for the Wasserstein space). More precisely, we need to find the $div(\rho\cdot)$-preimages X and Y of U and W that are orthogonal to the $div(\rho\cdot)$-free vector fields and then compute their $L^2(\rho)$-product $\int < X, Y > \rho$.
As we saw in a previous subsection, the orthogonal complement of the kernel of $div(\rho\cdot)$ is $im(\nabla)$. Moreover, U and W are orthogonal to $\mathbb{1}$. Thus, the $div(\rho\cdot)$-preimages X and Y (that re orthogonal to all $div(\rho\cdot)$-free vector fields) exist and are uniquely defined.
Alternatively, the above process is equivalent to solving the following optimization problem: minimize $\int < X, Y > \rho$ subject to $div(\rho X) = U$, $div(\rho Y) = W$. This second formulation comes directly from a variational approach:

Proof. Let X', Y' be such that $div(\rho(X + tX')) = U, div(\rho(Y + tY')) = W$ for all $t \in (-\epsilon, \epsilon)$ (i.e. $div(\rho X') = div(\rho Y') = 0$), then

$$0 \overset{!}{=} \frac{\partial}{\partial t} \int < X + tX', Y + tY' > \rho = \int < X, Y' > \rho + \int < X', Y > \rho$$

and this is zero for all ρ-divergence free X', Y' if and only if X and Y are orthogonal to all ρ-divergence free vector fields. □

Here, however, it will be easier to work directly with a vector field orthogonal to all $div(\rho\cdot)$-free vector fields instead of considering this minimization problem.
Thus, in order to compute the Riemannian metric, first we have to understand what it means to take the preimages with respect to $div(\rho\cdot)$.

Inverting $div(\rho\cdot)$ We have to invert the operator $div(\rho\cdot)$ to compute $X = (div(\rho\cdot))^{-1}(U)$ and $Y = (div(\rho\cdot))^{-1}(W)$. However, being orthogonal to all ρ-divergence free vector fields for X and Y means to be in $(\ker div(\rho\cdot))^\perp$, which in the last subsection we found to be $im(\nabla)$. For $X, Y \in im(\nabla)$ and $U, W \in (\ker \nabla)^\perp$ on the other hand, we see that we can compute X and Y by inverting the map

$$(\ker \nabla)^\perp \overset{\nabla}{\cong} im(\nabla) \overset{div(\rho\cdot)}{\to} (\ker \nabla)^\perp$$

(which above we proved to be a bijection) and then taking the gradient again. That map $div(\rho\nabla\cdot)$, however, is just the distorted Laplacian $M^{-1}\Delta_\rho$ that we computed earlier!
So, to sum things up, we now have to compute $\nabla((M^{-1}\Delta_\rho)^{-1}U)$ (and $\nabla((M^{-1}\Delta_\rho)^{-1}W)$, respectively), where the distorted Laplacian is restricted in its image and its domain to $(\ker \nabla)^\perp$.

Inverting the distorted Laplacian Since the distorted Laplacian of a discrete function $\phi \in \mathcal{F}(V)$ is given by the multiplication with $M^{-1}\Delta_\rho$, the inverse of the simplex tangent vector U is given by

$$\Delta_\rho^{-1} M \cdot U \in (\mathbb{R}\mathbb{1})^\perp \subset \mathcal{F}(V).$$

Therefore, the Riemannian metric at ρ for the two simplex tangent vectors U and W is given by

$$(U, W) := \int < X, Y > \rho,$$

where $X = \nabla\Delta_\rho^{-1} M \cdot U$ and $Y = \nabla\Delta_\rho^{-1} M \cdot W$. Thus, we get

$$\begin{aligned}(U,W)_\rho =& \int < X, Y > \rho \\ =& \int < \nabla \Delta_\rho^{-1} MU, \nabla \Delta_\rho^{-1} MW > \rho \\ =& (\Delta_\rho^{-1} MU)^t M div(\rho \cdot)(\nabla \Delta_\rho^{-1} MW > \rho) \\ =& U^t M \Delta_\rho^{-1} M M^{-1} \Delta_\rho \Delta_\rho^{-1} MW \\ =& U^t M \Delta_\rho^{-1} MW.\end{aligned}$$

Therefore, instead of the definition from the earlier subsection, we now have a much more compact expression for the definition.

Proposition 3. *For two simplex vector fields U and W in $(\mathbb{R}\mathbb{1})^\perp$ the Riemannian metric of the Wasserstein space at ρ is given by*

$$(U,W)_\rho = U^t M \Delta_\rho^{-1} MW.$$

Note that the right hand side can be interpreted as $U^t M (M^{-1}\Delta_\rho)^{-1} W = \int U \cdot (M^{-1}\Delta_\rho)^{-1} W$. Thus, in a way we can say that the Wasserstein metric is given by the inverse of the distorted Laplacian.

The Smooth Picture Formally, the fact that the Wasserstein metric is given by the inverse of the Laplacian, can also be seen in the smooth picture: Define the distorted Laplacian $\Delta_\rho f := div(\rho \nabla f)$ as for example in [41]. Then, let

$$(u,w)_\rho := \int < X, Y > \rho,$$

where $X = div(\rho \cdot)^{-1} u$ and $Y = div(\rho \cdot)^{-1} w$ are ρ-divergence-free preimages of simplex tangent functions u, w (i.e. they satisfy $\int_M u = \int_M w = 0$).
Then assume that $X = \nabla f$ and $Y = \nabla g$ for function f, g satisfying $f = \Delta_\rho^{-1} u$ and $g = \Delta_\rho^{-1} w$. Thus, we can formally calculate:

$$\begin{aligned}(u,w) =& \int_M < \nabla \Delta_\rho^{-1} u, \nabla \Delta_\rho^{-1} w > \rho \\ =& \int_M u \Delta_\rho^{-1} w.\end{aligned}$$

3.7 The Wasserstein Space as a Riemannian Manifold

The Probability measure simplex equipped with this Wasserstein metric is a high-dimensional Riemannian manifold and an interesting subject in itself. We want to present three operators/computations.

Laplacian Recall the formula for the Laplace operator on a Riemannian manifold in coordinates:

$$\Delta f = \frac{1}{\sqrt{|g|}}\partial_i(\sqrt{|g|}g^{ij}\partial_j f)$$

where g^{ij} are entries of the inverse metric tensor and $|g|$ refers to the determinant of the metric tensor.
In our situation, the metric at ρ is given by $M\Delta_\rho^{-1}M$. Thus, let us denote the inverse metric tensor by $G_\rho^{inv} := M^{-1}\Delta_\rho M^{-1}$. Let us also use the notation $|A|$ for the determinant of a matrix A. Then, at a point ρ our Laplace is given by

$$\begin{aligned}\sqrt{|G_\rho^{inv}|}\sum_i \partial_i\Big(\frac{1}{\sqrt{|G_\rho^{inv}|}}\sum_j (G_\rho^{inv})_{ij}\partial_j f\Big) &= \sqrt{\frac{|\Delta_\rho|}{|M|^2}}\sum_i \partial_i\Big(\sqrt{\frac{|M|^2}{|\Delta_\rho|}}\sum_j (G_\rho^{inv})_{ij}\partial_j f\Big)\\ &= \sqrt{|\Delta_\rho|}\sum_i \partial_i\Big(\sqrt{\frac{1}{|\Delta_\rho|}}\sum_j (M^{-1}\Delta_\rho M^{-1})_{ij}\partial_j f\Big).\end{aligned}$$

Alternatively, this can also be written in divergence form:

$$\Delta f = \sqrt{|G_\rho^{inv}|}div(\frac{1}{\sqrt{|G_\rho^{inv}|}}\cdot G_\rho^{inv}\nabla f).$$

What is remarkable here is that in order to compute this, we only need the inverse of the mass-matrix M which does not depend on ρ. In particular, no inversion of Δ_ρ is necessary. We briefly considered using this to conduct a heat method (see [15], a method that uses heat flow to compute geodesics) on this space and thus have a new way to compute dynamical optimal transport. However, it quickly turned out that for solving the heat equation, the dimension of the space is prohibitively large. A work-around could be to use Monte-Carlo methods (as in [37]) but the author Keenan Crane noted that even this approach will probably be very slow.

Where is Δ_ρ invertible? The set $\mathcal{P}(M)$ consists of functions $V \to \mathbb{R}$ such that $\rho(v) \geq 0$ and $\sum_{v\in V}\rho(v)A(v) = 1$. In other words: $\mathcal{P}(M)$ is a simplex.
The metric is defined by the inverse of Δ_ρ and we now want to ask what happens if ρ hits the boundary of the simplex, on which the metric is not necessarily defined any longer. If ρ is on the boundary (i.e. if $\rho(v) = 0$ for one or more vertices v) then the metric exists if and only if Δ_ρ is invertible (with the usual restriction to $(\mathbb{R}1)^\perp$).
Let us first consider the case where $\rho(v) = 0$ for exactly one vertex v. Having a look at (1) shows us that the reasoning of the *Properties of the rho-divergence*-subsection still applies so that the metric still exists at this part of the boundary. Therefore, on the inside of the facets of the simplex, the metric is still defined. In order to make Δ_ρ non-invertible, we

need ρ to vanish at at least 3 vertices so that A_ρ becomes zero on a triangle. However, we will see that even then the distorted Laplacian is usually invertible.
We want to find a condition that characterises when Δ_ρ is invertible. For this we introduce some notation: Let us call a triangle f a *zero-triangle* if $A_\rho(f) = 0$, i.e. ρ vanishes for all three vertices of the triangle. Also, recall the concept of the dual graph which is a the graph that has the triangles as its vertices and two triangles are adjacent if they share an edge. Moreover, the vertices of the original graph become faces in the dual graph which, however, are not necessarily triangles anymore. Now, we can formulate the following result.

Proposition 4. *Assume that we have a triangulation for which all the interior angles of the triangle are acute, i.e. strictly less than $\pi/2$. With this assumption we can give a sufficient and necessary condition for when Δ_ρ is invertible:*
Δ_ρ restricted to $(\mathbb{R}\mathbb{1})^\perp$ is not invertible if and only if in the dual graph there exists a cycle of zero triangles, which splits the dual graph into two non-empty parts.
Expressed more algebraically: Δ_ρ restricted to $(\mathbb{R}\mathbb{1})^\perp$ is not invertible if and only if there exists a cocycle of zero-faces in the dual graph cohomologous to zero.

Proof. Assume that in the dual graph there exists a cycle of zero triangles and denote the set of enclosed vertices by S. Fix two numbers $a \neq b \in \mathbb{R}$ and set a function f to be a on the vertices of S and b everywhere else. Then for any edge from S to its complement the two triangles adjacent to it will both be zero-triangles. Thus, for those edges the term $\frac{1}{2}\Big(\sum_{f\in ad(e)} \cot(\alpha_{e,f})A_\rho(f)\Big)$ is zero and we can see that

$$\phi^t \Delta_\rho f = \sum_{u\sim v} \frac{1}{2}\Big(\sum_{f\in ad(uv)} \cot(\alpha_{uv,f})A_\rho(f)\Big)(\phi(u)-\phi(v))(\psi(u)-\psi(v)) = 0$$

for any function ϕ. In particular, Δ_ρ has elements in its kernel besides the constants and is therefore not invertible.
On the other hand, let us now assume that there is a function f such that $\Delta_\rho f = 0$ and f is taking at least two different values. Choose an arbitrary vertex v and let S be the connected component of v in $\{u \in V | f(u) = f(v)\} = f^{-1}(f(v))$. Now, consider all the edges emanating from S. Obviously, $\omega_e = \frac{1}{2}\Big(\sum_{f\in ad(e)} \cot(\alpha_{e,f})A_\rho(f)\Big)$ has to be zero for all those edges. By our assumption on the angles, the cot-terms are positive. Thus, A_ρ has to be zero for all faces adjacent to those edges. Therefore, the boundary of S in the dual graph is a cycle of zero-faces. □

The Volume Form We see that the issues arise precisely for ρs with a certain combinatorial property. What happens to the volume form if we push ρ to the parts of the boundary of the simplex that have this property?
Recall the volume form of a Riemannian manifold in coordinates:

$$dvol = \sqrt{|g|}dx^1 \wedge \cdots \wedge dx^n. \tag{3}$$

In our case, the scalar factor is given by $\sqrt{|M\Delta_\rho^{-1}M|} = |M|/\sqrt{|\Delta_\rho|}$. For a better understanding of this term let us have a look at the eigenvalues of Δ_ρ. We sort the eigenvalues of the matrix Δ_ρ: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{max}$.
The largest eigenvalue can be bounded by absolute row sums using the explicit formula for the weights $\frac{1}{2}\Big(\sum_{f\in ad(e)} \cot(\alpha_{e,f})A_\rho(f)\Big)$. Since $A_\rho(f)\frac{1}{3}\sum_{v\in ad(f)}\rho(v) \leq \max_{v\in V}\{\frac{1}{A(v)}\}$ (as $\sum_{v\in V}\rho(v)A(v) = 1$), we get a bound C for $\lambda_{max} := \lambda_{|V|-1}$ that only depends on the mesh (and not on ρ).
Therefore, for the inverses of the eigenvalues, we have $\frac{1}{\lambda_1} \geq \frac{1}{\lambda_2} \geq \cdots \geq \frac{1}{\lambda_{max}} \geq 1/C > 0$. Thus, we can bound $|M|/\sqrt{|\Delta_\rho|}$ from above and from below only in terms of $\frac{1}{\lambda_1}$ and we see that when ρ gets close to the area where Δ_ρ is not invertible, λ_1 goes to zero and thus the scalar factor from the coordinate expression of the volume form grows to infinity.

4 A Wasserstein-like Distance on Vector Fields

The aim of this section is to present two novel distances for vector fields on manifolds that are inspired by the Wasserstein distance. The second of the two is a bit more intuitive and seems to allow more interesting relations to other fields. Moreover, the first construction will usually not even be a distance in many cases due to a lack of the triangle inequality but due to the similarity in the construction we still want to present both concepts. Before discussing our approach let us consider some simple thoughts in order to gain intuition.
A first thing to notice is that it is not trivial to copy the algebraic formulation of optimal transport. The first question that one might ask is how to think of vector fields in a way that is similar to measures. Our mental picture for this was something like a fuzzy vector fields. Intuitively speaking, at every point the direction of the vector field could potentially involve some notion of non-determinism. The question would then be what a transport plan shall be. Trying to even write something meaningful down for this proved to be awkwardly difficult. In fact, most Wasserstein-like distances for vector-valued measures or matrices in the literature involve some alternative formulation for the Waaserstein distance (e.g. the Benamou-Brenier formula for dynamical optimal transport or dual formulations).
We went for a different, more geometrical approach. Our idea is to conduct the optimal transport on some bundle over the base space instead of the base space itself. For this, the objects to transport need to be translated in some way to objects on a bundle and there are different ideas for how this could be done.
The approach that seemed most interesting to us was an approach that interprets vector fields as measures on the unit circle bundle over the surface, since this approach allowed for an intuitive understanding and has a certain connection to parallel transport. Before we get to that, however, we want to discuss another approach where the optimal transport is performed on measures on a bundle which project to given vector fields in a certain way.
We will keep most of the survey to a very restricted setting where we transport vector fields on a compact, oriented 2-dimensional Riemannian manifold (without boundary).

4.1 Projection-Approaches

The concept we introduce here is rather general and can applied to multiple spaces. In particular, we will recover a distance function that was already considered in other places. However, unless $p = 1$, this approach will usually not yield a distance.

We will forget about our resolution of focussing on vector fields on surfaces for the moment and describe a very general, but also somewhat formal setting. Let us denote a compact manifold by M and let B be a metric fibre bundle over M (equipped with the distance d^B) such that there exists a surjective map

$$\mathrm{proj} : \mathcal{P}(B) \to \Xi(M),$$

where $\mathcal{P}(B)$ is the space of probability measures on B (or, as in our first example, measures with a given total mass) and $\Xi(M)$ is a space of measure-theoretical objects that we want

to transport and proj is induced by some algebraic operation on the fibre. Then, we can define a semi-distance on $\Xi(M)$ in the following way:

Definition 6. *For two objects α, β in $\Xi(M)$ we define their semi-distance as*

$$W_p^M(\alpha, \beta) := \inf_{\tilde{\alpha}, \tilde{\beta}} W_p^B(\tilde{\alpha}, \tilde{\beta}),$$

where the infimum is over all $\tilde{\alpha}$ and all $\tilde{\beta}$ in $\mathcal{P}(B)$ that satisfy $\mathrm{proj}(\tilde{\alpha}) = \alpha$ *and* $\mathrm{proj}(\tilde{\beta}) = \beta$ *and W_p^B denotes the Wasserstein p-distance on B.*

In this generality it is of course by no means clear why this should be a distance (in fact it usually is not) and what properties it has. Even the term 'semi-distance' is chosen somewhat hopeful since the property that $W_p^M(\alpha, \beta) \neq 0$ for $\alpha \neq \beta$ still requires a proof for any concrete application of the concept. We will, however, not bother with semi-distances but only discuss whether the objects at hand are distances or not. Thus, in the following we use the term 'semi-distance' as an empty expression only there to have a name by which we call the above function. Let us consider two examples, where we give definitions for semi-distances for signed measures and for vector fields, respectively.

4.1.1 A Toy Example

To make the above considerations more precise, let us consider a toy example. For this let M be an oriented, compact Riemannian manifold and $B = M \times \{\pm 1\}$ a trivial bundle equipped with the obvious metric. Let W_p^B denote the Wasserstein distance on B as usual. Note that B is disconnected and therefore optimal transport on B is conducted by conducting optimal transport on the connected components $M \times \{+1\}$ and $M \times \{-1\}$ independently. Taking into account the power p, the Wasserstein distance of two measures $\tilde{\mu}, \tilde{\nu} \in \mathcal{M}(B)$ on B can then be written as

$$W_p^B(\tilde{\mu}, \tilde{\nu}) = \left(W_p^{M\times\{+1\}}(\tilde{\mu}_{|M\times\{+1\}}, \tilde{\nu}_{|M\times\{+1\}})^p + W_p^{M\times\{-1\}}(\tilde{\mu}_{|M\times\{-1\}}, \tilde{\nu}_{|M\times\{-1\}})^p \right)^{1/p},$$

where $\tilde{\mu}_{|M\times\{+1\}}$ denotes the restriction of the measure $\tilde{\mu}$ to the set $M \times \{+1\}$ and the total masses of the restrictions of $\tilde{\mu}, \tilde{\nu}$ need to be finite and satisfy $|\tilde{\mu}_{|M\times\{+1\}}| = |\tilde{\nu}_{|M\times\{+1\}}|$ and $|\tilde{\mu}_{|M\times\{-1\}}| = |\tilde{\nu}_{|M\times\{-1\}}|$ (such that the Wasserstein distance exists).
Now we use for $\Xi(M) = \mathcal{S}_0(M) := \{\mu \in \mathcal{S}(M) : \int_M \mathrm{d}\mu = 0\}$ the set of signed measures with zero total mass. For our projection we notice that there is a map from functions $\tilde{f}$ on B to functions on M via $m(\tilde{f})(x) = \tilde{f}(x, +1) - \tilde{f}(x, -1)$. This map easily translates to a map mapping measures on B to signed measures on M via push-forward, defined by

$$\mathrm{proj}(\tilde{\mu}) = \pi_\#(\tilde{\mu}_{|M\times\{+1\}}) - \pi_\#(\tilde{\mu}_{|M\times\{-1\}}),$$

or equivalently, via the property $\int_M f \mathrm{d}\mathrm{proj}(\tilde{\mu}) = \int_{M\times\{+1\}} f(x)\mathrm{d}\tilde{\mu}(x, +1) - \int_{M\times\{-1\}} f(x)\mathrm{d}\tilde{\mu}(x, -1)$.
In other words, we interpret $\tilde{\mu}$ restricted to $M \times \{+1\}$ and $\tilde{\mu}$ restricted to $M \times \{-1\}$ as

two measures on M and just take their difference.
We can now consider the above minimization problem

$$W_p^{\mathcal{S}_0}(\mu,\nu) := \inf_{\tilde{\mu},\tilde{\nu}} W_p^{M\times\{\pm 1\}}(\tilde{\mu},\tilde{\nu}),$$

where the infimum is taken over all measures $\tilde{\mu},\tilde{\nu} \in \mathcal{M}(B)$ of finite total mass which project to the signed measures μ,ν, i.e. $\mathrm{proj}(\tilde{\mu}) = \mu$ and $\mathrm{proj}(\tilde{\nu}) = \nu$, and satisfy that the total mass of $\tilde{\mu}_{|M\times\{+1\}}$ is the same as the total mass of $\tilde{\nu}_{|M\times\{+1\}}$ and similarly for $M\times\{-1\}$ (so that the Wasserstein distance is defined on $M\times\{+1\}$ and $M\times\{-1\}$).
Intuitively speaking, using this projection means allowing the signed measures μ to have positive *and* negative mass at the same time at a given point in M (whenever $\tilde{\mu}_{|M\times\{+1\}}$ and $\tilde{\mu}_{|M\times\{-1\}}$ both have mass there).
Recall that any signed measure μ can be written in a unique way as $\mu = \mu_+ - \mu_-$ for measures μ_+ and μ_- of minimal total mass. Since $\tilde{\mu}_{|M\times\{+1\}}$ and $\tilde{\mu}_{|M\times\{-1\}}$ are both (non-negative) measures, it is clear that $\tilde{\mu}_{|M\times\{+1\}} = \mu_+ + \gamma$ and $\tilde{\mu}_{|M\times\{-1\}} = \mu_- + \gamma$ for some measure $\gamma \in \mathcal{M}(M)$ after identifying $M\times\{+1\}$ and M (or $M\times\{-1\}$ and M, respectively). Thus, the above semi-distance can also be written as

$$W_p^{\mathcal{S}_0}(\mu,\nu) = \inf_{\gamma_1,\gamma_2} \left(W_p^M(\mu_+ + \gamma_1, \nu_+ + \gamma_2)^p + W_p^M(\mu_- + \gamma_1, \nu_- + \gamma_2)^p\right)^{1/p},$$

where the infimum is taken over all measures $\gamma_1,\gamma_2 \in \mathcal{M}(M)$ satisfying $|\mu_+|+|\gamma_1| = |\nu_+|+|\gamma_2|$ (or equivalently $|\mu_-| + |\gamma_1| = |\nu_-| + |\gamma_2|$). Here the absolute values denote the total mass of the measures (i.e. $|\mu_+| := \int_M \mathrm{d}\mu_+$). For $p = 1$ we can show that this expression can be simplified a bit.

Proposition 5. *For $p = 1$ the above (semi-)distance can be written as*

$$W_1^{\mathcal{S}_0}(\mu,\nu) = W_1^M(\mu_+ + \nu_-, \mu_- + \nu_+).$$

Proof. "$\leq$":
Clearly, in above formula for $W_p^{\mathcal{S}_0}(\mu,\nu)$ we can just choose $\gamma_1 = \nu_-$ and $\gamma_2 = \mu_-$. Indeed, $|\mu_+| + |\gamma_1| = |\mu_+| + |\nu_-| = |\mu_-| + |\nu_+| = |\nu_+| + |\gamma_2|$ is then fulfilled and thus we have

$$\begin{aligned} W_1^{\mathcal{S}_0}(\mu,\nu) &= \inf_{\gamma_1,\gamma_2} \{W_1^M(\mu_+ + \gamma_1, \nu_+ + \gamma_2) + W_1^M(\mu_- + \gamma_1, \nu_- + \gamma_2)\} \\ &\leq W_1^M(\mu_+ + \nu_-, \nu_+ + \mu_-) + \underbrace{W_1^M(\mu_- + \nu_-, \nu_- + \mu_-)}_{=0} \\ &= W_1^M(\mu_+ + \nu_-, \mu_- + \nu_+). \end{aligned}$$

"$\geq$": Let $\gamma_1,\gamma_2 \in \mathcal{P}(M)$ satisfy $|\mu_+| + |\gamma_1| = |\nu_+| + |\gamma_2|$. Recall the dual formulation for the Wasserstein-1-distance by which we have

$$\begin{aligned} &W_1^M(\mu_+ + \gamma_1, \nu_+ + \gamma_2) + W_1^M(\mu_- + \gamma_1, \nu_- + \gamma_2)\} \\ &= \sup_f \{\int_M f \mathrm{d}(\mu_+ + \gamma_1 - \nu_+ - \gamma_2)\} + \sup_g \{\int_M g \mathrm{d}(\nu_- + \gamma_2 - \mu_- - \gamma_1)\}, \end{aligned}$$

where both suprema are taken over all 1-Lipschitz functions f, g and for the letter term we changed the order of the argument $W_1^M(\mu_- + \gamma_1, \nu_- + \gamma_2) = W_1^M(\nu_- + \gamma_2, \mu_- + \gamma_1)$. Clearly, the sum of those two suprema can only become smaller if we just choose $f = g$. So we can bound this from below by

$$\begin{aligned}
&\geq \sup_f \{\int_M f \mathrm{d}(\mu_+ + \gamma_1 - \nu_+ - \gamma_2) + \int_M f \mathrm{d}(\nu_- + \gamma_2 - \mu_- - \gamma_1)\} \\
&= \sup_f \{\int_M f \mathrm{d}(\mu_+ + \nu_- - \mu_- - \nu_+)\} \\
&= W_1^M(\mu_+ + \nu_-, \mu_- + \nu_+).
\end{aligned}$$

Since this inequality holds for any $\gamma_1, \gamma_2 \in \mathcal{P}(M)$ that satisfy $|\mu_+| + |\gamma_1| = |\nu_+| + |\gamma_2|$, it also holds for the infimum. Thus, we get $W_1^{\mathcal{S}_0}(\mu, \nu) \geq W_1^M(\mu_+ + \nu_-, \mu_- + \nu_+)$. □

In [22] Edoardo Mainini defines a very similar set of semi-distances for signed measures. Before discussing those and comparing them to our definition, let us quote an important theorem from his paper:

Theorem 6. *Let $p \geq 1, \alpha \geq 0$, and let μ, ν be two measures with total mass α. Then for any $n \in \mathbb{N}$ there exists a measure σ with total mass $n\alpha$, such that*

$$W_p^M(\mu + \sigma, \nu + \sigma)^p \leq \frac{1}{(n+1)^{p-1}} W_p^M(\mu, \nu)^p.$$

In particular, we can immediately see that for $p > 1$ we get zero for the our semi-distance: just choose $\gamma_1 = \nu_+ + \sigma$ and $\gamma_2 = \mu_+ + \sigma$ such that for $W_p^{\mathcal{S}_0}(\mu, \nu)$ we get $W_p^{\mathcal{S}_0}(\mu, \nu) \leq \inf_\sigma (0 + W_p^M(\mu_- + \nu_+ + \sigma, \nu_- + \nu_+ + \sigma)^p)^{1/p} = 0$.
Therefore, for $p > 1$ a more sensible definition would be to restrict the distance to signed measures with bounded total mass $|\mu| := |\mu_+| + |\mu_-|$. Thus, we might fix some numbers $0 < b_1 < b_2$ and consider the semi-distance on the set of signed measures with total mass less than b_1 given by

$$W_p^{\mathcal{S}_0}(\mu, \nu) := \inf_{\tilde{\mu}, \tilde{\nu}} W_p^{M \times \{\pm 1\}}(\tilde{\mu}, \tilde{\nu}),$$

where the infimum is taken over all measures $\tilde{\mu}, \tilde{\nu}$ with with total mass less than b_2. Alternatively, this is given by

$$W_p^{\mathcal{S}_0}(\mu, \nu) = \inf_{\gamma_1, \gamma_2} \left(W_p^M(\mu_+ + \gamma_1, \nu_+ + \gamma_2)^p + W_p^M(\mu_- + \gamma_1, \nu_- + \gamma_2)^p \right)^{1/p},$$

where γ_1 and γ_2 are as above but additionally satisfy that $|\mu_+| + |\mu_-| + 2|\gamma_1| \leq b_2$ and $|\nu_+| + |\nu_-| + 2|\gamma_2| \leq b_2$.
Mainini [22] defines various semi-distances for signed measures with total mass less than some fixed number b. One of them is given by the formula

$$W_{p,b}^{\text{Mainini}} := W_p^M(\mu_+ + \nu_-, \mu_- + \nu_+).$$

So, for $p = 1$ we recovered one of his semi-distances.
In his paper Mainini also proves that for $p = 1$, the above semi-distance is actually a distance. For all other $p > 1$, he gives an example for which the triangle inequality is not fulfilled. Due to the addition of the γ_1/γ_2-terms, it is not clear whether his example also works for our semi-distances though. In fact, we do not know whether our semi-distances $W_p^{\mathcal{S}_0}(\mu, \nu)$ are distances for $p > 1$, but our expectation is that they are not.

4.1.2 Vector Fields

The same approach can be done when dealing with vector fields: Assume we equipped the tangent space itself with a metric d_{TM}. This defines the Wasserstein distance W_p^{TM}. However, since TM is not compact anymore, the Wasserstein distance is only a distance for measures of finite p-moment, i.e. measures $\tilde{\mu}$ satisfying $\int_{TM} d_{TM}^p(v, v_0)\mathrm{d}\tilde{\mu} < \infty$ and this definition does not depend on the choice of v_0 (see [41]).
By disintegration, any probability measure $\tilde{\mu}$ on TM can be written in the form $\tilde{\mu} = \mu(x) \otimes \nu_x$ where $\mu = \pi_{\#}\tilde{\mu}$ is a probability measure on M and ν_x is a probability measure on T_xM. Since T_xM is a vector space, we can take the mean of ν_x resulting in a vector $\int_{T_xM} v\mathrm{d}\nu_x$ of T_xM if the mean exists. This motivates the following definition: Define $\Xi(M)$ as the space $\mathcal{X}(M) \times_{\mathcal{F}(M)} \mathcal{P}(M)$ where $\mathcal{F}(M)$ denotes the set of measurable functions over M and $\mathcal{X}(M)$ the vector fields over M with a measurable regularity (i.e. vector fields whose components are measurable in any coordinate chart). The notation $\times_{\mathcal{F}(M)}$ means that we identify elements of the forms $(f \cdot X, \mu)$ and $(X, f \cdot \mu)$ with each other (as is common for example for the notation for associated vector bundles from principal bundle theory). More precisely, $\mathcal{X}(M) \times_{\mathcal{F}(M)} \mathcal{P}(M) = \mathcal{X}(M) \times \mathcal{P}(M)/\sim$ where the equivalence relation is defined via $(f \cdot X, \mu) \sim (X, f \cdot \mu)$ whenever μ and $f\mu$ are both probability measures.
This space can be seen as a way to build dual elements for 1-forms that respect the local structure of M and do not require any metric on M. Indeed, given a 1-form α the composition

$$\langle [(X, \mu)], \alpha \rangle := \int_M \alpha(X)\mathrm{d}\mu$$

is well-defined (though from the functional analytical view point $\mathcal{X}(M) \times_{\mathcal{F}(M)} \mathcal{M}(M)$ would probably be the more interesting space).
In order to guarantee that the means $\int_{T_xM} v\mathrm{d}\nu_x$ exist almost everywhere let $\mathcal{P}_{finite}(TM)$ be the space of measures on TM that have finite p-moment as well as finite 1-moment and assume that there exists some norm $\|\cdot\|$ on M and a constant $c > 0$ such that the metric d_{TM} satisfies: for any $x \in M$ and any $v \in T_xM$ we have $\|v\| \leq c \cdot d_{TM}(v, 0)$ where 0 denotes the zero vector in T_xM. Indeed, then we have

$$\int_{T_xM} \|v\|\mathrm{d}\nu_x \leq c \int_{T_xM} d_{TM}(v, 0)\mathrm{d}\nu_x \leq c d_{TM}(0, v_0) + c \int_{T_xM} d_{TM}(v, v_0)\mathrm{d}\nu_x$$

and we know that $\int_{T_xM} d_{TM}(v, v_0)\mathrm{d}\nu_x < \infty$ exists almost everywhere since the 1-moment $\int_M d_{TM}(v, v_0)\mathrm{d}\tilde{\mu} = \int_M \int_{T_xM} d_{TM}(v, v_0)\mathrm{d}\nu_x\mathrm{d}\mu$ is finite.

We define a projection proj : $\mathcal{P}_{finite}(TM) \to \mathcal{X}(M) \times_{\mathcal{F}(M)} \mathcal{P}(M)$ by mapping an element $\tilde{\mu} \in \mathcal{P}_{finite}(TM)$ to $[(X, \mu)]$, where $\mu = \pi_{\#}\tilde{\mu}$ and $X_x := \int_{T_xM} v \mathrm{d}\nu_x$ are constructed from the disintegration of $\tilde{\mu}$.
Again, we can now define a semi-metric on $\Xi(M)$ in the way described above.

Definition 7. *For two elements in* $[(X, \mu)], [(Y, \nu)] \in \mathcal{X}(M) \times_{\mathcal{F}(M)} \mathcal{P}(M)$ *a semi-distance is defined as*

$$W_p^{\Xi}([(X, \mu)], [(Y, \nu)]) := \inf_{\tilde{\mu}, \tilde{\nu}} W_p^{TM}(\tilde{\mu}, \tilde{\nu}),$$

where the infimum is taken over all $\tilde{\mu}, \tilde{\nu}$ *satisfying* $\mathrm{proj}(\tilde{\mu}) = [(X, \mu)]$ *and* $\mathrm{proj}(\tilde{\nu}) = [(Y, \nu)]$.

Intuitively speaking, this semi-distance transports vectors on the tangent bundle but similarly to our toy example we allow the vectors to change instantaneously into a measure on their respective tangent space which has the same mean.
Clearly, this approach is not very fulfilling since it is difficult to interpret this semi-distance and the mere fact that we usually do not get actual distances makes the approach less interesting. One question that seemed interesting though would be whether this could be a distance for $p = 1$. Unfortunately, we could not prove this nor refute it but our conjecture is that the key to answering this question lies in finding a dual formulation, similarly to our toy example.

4.2 Approaches using Embeddings

In the following we will consider approaches using embeddings into the space of probability measures over some bundle. This will turn out to be more fruitful than the previous approach.
In order to gain some intuition, we again start by discussing a toy example. After that we introduce the needed spaces and discuss a certain energy before giving the definition of a Wasserstein-like distance for vector fields. Then, we collect some properties and give intuition for this distance.

4.2.1 A Toy Example

Again, we consider the case of signed measures. As before, let us write the signed measure μ in terms of its positive and negative parts $\mu = \mu_+ - \mu_-$. Define the space

$$\mathcal{S}_{1/2,1/2} := \{\mu \in \mathcal{S}(M) : \int_M \mathrm{d}\mu_+ = \frac{1}{2} = \int_M \mathrm{d}\mu_-\}.$$

Then we again consider the trivial bundle $M \times \{\pm 1\} \to M$. Here we have two sections: $s_+(m) := (m, +1)$ and $s_-(m) := (m, -1)$. Our embedding $i : \mathcal{S}_{1/2,1/2} \to \mathcal{P}(B)$ is now defined as

$$i = (s_+)_{\#}\mu_+ + (s_-)_{\#}\mu_-.$$

In other words, we simply put the measure μ_+ on $M \times \{+1\}$ and the measure μ_- on $M \times \{-1\}$. For two signed measures $\mu, \nu \in \mathcal{S}_{1/2,1/2}$, their distance is then given by

$$D_p(\mu, \nu) = W_p(i(\mu), i(\nu)) = \Big(W_p(\mu_+, \nu_+)^p + W_p(\mu_-, \nu_-)^p\Big)^{1/p}.$$

This is a very simple object as it just transports positive and negative parts independently. The fact that this is a distance is obvious. For $p = 1$ this is yet another distance that was considered by Mainini in [22].
The rest of this section is devoted to vector fields on compact, oriented, 2-dimensional Riemannian manifolds.

4.2.2 Circle Bundles and the Kaluza-Klein Space

In this subsection, we will give a gentle introduction to some of the concepts needed from differential geometry. Again we consider the setting of a compact and oriented Riemannian surface without boundary and denote the Levi-Civita connection by ∇.
A typical way of thinking about vectors is to divide the information into the vector's direction and its length. The directional part can be described by the mathematical notion of the unit circle bundle which is the space that consists of all unit length vectors for a given surface

$$\mathbb{S}M = \{v \in TM : |v| = 1\}.$$

This is then a fibre bundle where each fibre is given by a circle. Note that on each fibre there is a group action given by the group $SO(T_xM)$ which can be canonically identified with the group $\mathbb{S}^1$. In particular, we have a group action on this bundle by the group $\mathbb{S}^1$ which preserves the fibres and acts by rotating the fibres.
We want to consider moving vectors. A slight change of perspective is given by viewing a unit vector U moving along a curve γ as a curve $\tilde{\gamma}^U$ in $\mathbb{S}M$ instead. We know that our moving vector U will be parallel iff $\nabla_{\dot{\gamma}} U = 0$. Now, one can ask what being parallel means for the curve $\tilde{\gamma}^U$ whose derivative would lie in the tangent space to the unit circle bundle. This question is answered by the introduction of the horizontal space. Let us consider a single unit vector v sitting at a point $x \in M$. For any curve γ passing through x we will then get a curve $\tilde{\gamma}$ in $\mathbb{S}M$ by parallel translating the vector v along γ. One can then show that the space

$$H_v := \operatorname{span}\{\dot{\tilde{\gamma}}_{|t=0} : \gamma \text{ a curve passing through } v \text{ with } \gamma(0) = x\} \subset T\mathbb{S}M$$

forms a 2-dimensional linear subspace of the tangent space to the unit circle bundle that is called the *horizontal space* at v and which is isomorphic to TM via the isomorphism $\mathrm{d}\pi_{H_v}$, where $\pi : \mathbb{S}M \to M$ is the bundle projection. Then it becomes clear that in terms of the curve $\tilde{\gamma}^U$, the field U being parallel means that $(\tilde{\gamma}^{U(t)})^{\cdot}$ lies in $H_{U(t)}$ for all times. The natural complement of the horizontal space is the vertical space which is defined as $\ker \mathrm{d}\pi$ and indeed, $T\mathbb{S}M_v = H_v + V_v$.

A supremely important property of the unit circle bundle is its symmetry with respect to the rotations on the respective circles and most meaningful mathematical objects will respect this symmetry meaning that they are invariant (or equivariant) under the action of the group $\mathbb{S}^1$ that acts on this bundle. Of course this is also true for the horizontal spaces: Let s_θ be the rotation by the angle θ, then $(s_\theta)_* H_v = H_{s_{\theta(v)}}$.
Next, we want to turn $\mathbb{S}M$ also into a geometric space by endowing it with a metric. This metric shall be invariant under the group action as well. Thus, the most natural way is to use the metric g pulled back via π on the horizontal space and requiring horizontal space and vertical space to be orthogonal. All that is left is then to also endow the vertical space with an $\mathbb{S}^1$-invariant metric. This can also be achieved in a simple manner: The group $\mathbb{S}^1$ itself can be given a (left- and right-)$\mathbb{S}^1$-invariant metric by identifying it with $\mathbb{R}/2\pi\mathbb{Z}$. Since the vertical space consists of the tangent spaces to the unit circles in $\mathbb{S}M$ and these are isomorphic to $\mathbb{S}^1$ (e.g. by acting), we can endow the vertical space with the metric from $\mathbb{S}^1$. More precisely: Let X be a tangent vector to the tangent space of $\mathbb{S}^1$ at the identity. Define the Killing vector K^X in V_v as the vector originating from X via (the exponential of) the group action

$$K^X = \frac{\partial}{\partial t}_{|t=0} v.\exp(tX).$$

Note that here we identified $\mathbb{S}^1$ with the unit circle in $\mathbb{C}$ and its tangent space at the identity (which now is just 1) with $i\mathbb{R}$. Also, the group action on the unit circle bundle is as usual written as a right action. Then we define

$$\langle K^X, K^Y \rangle_{\mathbb{S}M} = \langle X, Y \rangle_{\mathbb{S}^1}$$

and since the map $\mathbb{S}^1 \to \pi^{-1}(\pi(v))$ is an isomorphism, so is the map $K^\cdot : T_1\mathbb{S}^1 \to T_v\pi^{-1}(\pi(v)) = V_v$. Thus, this defines a metric on the vertical space that is rotation invariant by construction. In total, we get a rotation invariant metric on the whole unit circle bundle $\mathbb{S}M$ which we call the Kaluza Klein metric. The space $\mathbb{S}M$ endowed with this metric we call the Kaluza Klein space. The name Kaluza Klein is used here since we were inspired to use this metric by other work, that used similar metrics on the circle bundle and associated these constructions with the Kaluza Klein space from physics (e.g. [26] in appendix A). However, this exact metric was introduced by Sasaki in [27] and is usually known as Sasaki-metric in the literature. For a more up-to-date survey of the metric we refer the reader to [31] and [28].
Remark: In the literature the horizontal space is defined in a different manner since it is usually used to introduce a new language that allows for an alternative way to define the notions of affine connections and the Levi-Civita connection. Thus, our exposition of using a connection to introduce the horizontal space is not the usual way. We chose this approach since it seemed to be the shortest way to introduce the horizontal space.

4.2.3 Energies of Moving Vectors

Before discussing a distance for vector fields, we investigate a certain Energy for a unit vector moving along the manifold. More precisely, let us consider a unit vector field U

along a curve $\gamma : [0,1] \to M$ and define the energy

$$\mathcal{E}(\gamma, U) := \int_0^1 |\dot{\gamma}|^2 + |\nabla_{\dot{\gamma}} U|^2 \mathrm{d}t$$

where the first term corresponds to the energy associated to the curve while the second term is a measure for how parallel U is along γ.
One might also consider the starting vector $U(0) := U_{\gamma(0)}$ at $p = \gamma(0)$ and parallel transport it along γ to a parallel vector field $P^{\gamma}_{0,t}(U(0))$. That would then yield an angle function that for time t gives the angle between $U(t) := U_{\gamma(t)}$ and $P^{\gamma}_{0,t}(U(0))$

$$\theta(t) := \angle(U(t), P^{\gamma}_{0,t}(U(0))).$$

Alternatively, one might view U altogether as a curve in the unit circle bundle $\mathbb{S}M$ instead. We will denote the associated curve in $\mathbb{S}M$ by $\tilde{U}$ (instead of the previous notation $\tilde{\gamma}^U$). Thus, the splitting of the tangent space of the unit circle bundle into horizontal space and vertical space yields a splitting of the derivative of $\tilde{U}$ into the horizontal part of the derivative $\dot{\tilde{U}}^H$, which has the same norm as $\dot{\gamma}$ in TM (with $\mathbb{S}M$ equipped with the Kaluza-Klein metric) and its vertical part $\dot{\tilde{U}}^V$.
In the following lemma we give two alternative forms for the above energy.

Lemma 1. *With the notation just introduced, we have*

$$\begin{aligned}\mathcal{E}(\gamma, U) &= \int_0^1 |\dot{\gamma}|^2 + |\dot{\theta}|^2 \mathrm{d}t \\ &= \int_0^1 |\dot{\tilde{U}}|^2 \mathrm{d}t.\end{aligned}$$

This lemma and particularly the second formula here is the motivation for the distance defined in the next subsection.

Proof. Since the horizontal space is isometrically isomorphic to the tangent space of M, we have $|\dot{\tilde{U}}^H| = |\mathrm{d}\pi(\dot{\tilde{U}})| = |(\pi \circ \tilde{U})^{\cdot}| = |\dot{\gamma}|$. Thus, the only thing that has to be shown is

$$|\nabla_{\dot{\gamma}} U| = |\dot{\theta}| = |\dot{\tilde{U}}^V|.$$

For the first equation, we write $U(t) = R_{\theta(t)} P^{\gamma}_{0,t}(U(0)))$ with $R_{\theta(t)}$ being the rotation by the angle $\theta(t)$. Introducing the parallel fields $V(t) = P^{\gamma}_{0,t}(U(0)))$ and $V^{\perp}(t) = R_{\pi/2} P^{\gamma}_{0,t}(U(0)))$, we can then write

$$\begin{aligned}\nabla_{\dot{\gamma}} U &= \nabla_{\dot{\gamma}} \Big(\cos(\theta(t)) V(t) + \sin(\theta(t)) V^{\perp}(t) \Big) \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t} \cos(\theta(t)) \right) V(t) + \left(\frac{\mathrm{d}}{\mathrm{d}t} \sin(\theta(t)) \right) V^{\perp}(t) \\ &= \Big(-\sin(\theta(t)) V(t) + \cos(\theta(t)) V^{\perp}(t) \Big) \cdot \dot{\theta}(t).\end{aligned}$$

Taking norms yields the first equation.
The proof of the second equation is similar: Denote the curve in $\mathbb{S}M$ that corresponds to $V(t)$ by $\tilde{V}$, then $\tilde{U}(t) = \tilde{V}(t).\theta(t)$. Thus, by the Leibniz rule we have

$$\dot{\tilde{U}} = \frac{\partial}{\partial t}_{|t=t_0} \tilde{U} = \frac{\partial}{\partial t}_{|t=t_0} \tilde{V}(t).\theta(t_0) + \frac{\partial}{\partial t}_{|t=t_0} \tilde{V}(t_0).\theta(t)$$

whose first part is obviously horizontal while the second part is

$$\frac{\partial}{\partial t}_{|t=t_0} \tilde{V}(t_0).\theta(t) = \frac{\partial}{\partial t}_{|t=t_0} \tilde{V}(t_0).\theta(t_0).\exp\left(\frac{\partial}{\partial s}_{|s=t_0} \theta(s) \cdot t\right),$$

which is vertical with norm $|\frac{\partial}{\partial s}_{|s=t_0}\theta(s)|$. □

4.2.4 A Metric for Vector Fields

We now introduce a distance for smooth vector fields by identifying vector fields on the surface with measures on the circle bundle, which is equipped with the Kaluza-Klein metric. The starting point for our distance is splitting the information that is given by a vector into its direction and its length. Parallelism and measuring how parallel a moving vector is are also concepts that very interest us. Since parallel transport does not affect the length of a vector, it is natural to restrict ourselves to vectors of unit length. However, we also need to associate a mass to a moving vector. Thus, we reinterpret a vector by viewing its normalized version as the unit vector in question, while its length will tell us the infinitesimal mass of that unit vector. For the case of a zero-vector, the corresponding normalized version will not be defined, but since in that case there is also no mass, these vectors are never troublesome. Let us make this more precise.
For a smooth vector field $X \in \mathcal{X}(M)$, we define a section s_X on the circle bundle $\mathbb{S}M$ that is defined whenever X is not zero by

$$s_X(p) := X(p)/|X(p)|.$$

Also, we associate a measure $\mu_X^M \in \mathcal{P}(M)$ to X via

$$\mathrm{d}\mu_X^M(p) = |X(p)|\mathrm{dvol}$$

where dvol denotes the Riemannian volume measure on M.

Definition 8. *We define the* fuzzy vector field $\mu_X \in P(\mathbb{S}M)$ *associated to* X *via*

$$\mu_X := (s_X)_\#(\mu_X^M).$$

Proof. The well-definedness follows from the measure-theoretic fact that for a μ-almost everywhere-defined function f also $f_\#\mu$ is well-defined. Thus, in this case, all there is to check is that s_X is μ_X^M- almost everywhere defined, which is obvious from the definitions. □

We use the term fuzzy vector since this is now a measure. In fact, we may call the space $\mathcal{P}(\mathbb{S}M)$ the space of fuzzy vector fields, which contains the space of smooth vector fields. Note that this measure's support is exactly given by the normalized vector field while its infinitesimal mass is given by the (squared) norm of the vector field. Also, for a smooth vector field, the support of this measure will look surface-like away from the vector field's zeros. Of course, generically, even smooth vector fields will have zeros (compare to the hairy ball theorem) at which the support will have a singular shape.
Intuitively, the construction of the associated fuzzy vector field corresponds to thinking of a vector field as consisting of a collection unit vectors with different masses associated.
Now, we can define a Wasserstein-like distance for vector fields:

Definition 9. *For two vector fields X, Y on a closed, oriented Riemannian surface satisfying $\|X\|_{L^1} = \|Y\|_{L^1}$, we define a distance via*

$$W_2(X,Y) := W_2^{\mathbb{S}M}(\mu_X, \mu_Y) = \left(\inf_{\rho \in \Pi(\mu_X, \mu_Y)} d_{\mathbb{S}M}(x,y)^2 \mathrm{d}\rho(x,y) \right)^{1/2}$$

where $W_2^{\mathbb{S}M}$ denotes the Wasserstein distance on $\mathbb{S}M$ equipped with the Kaluza-Klein metric, i.e. $d_{\mathbb{S}M}$ is the distance on $\mathbb{S}M$ induced from the Kaluza-Klein metric.

Note that at this point the number 2, is rather arbitrary. Of course one could define the Wasserstein-p-distance for vector fields exactly the same way.

Intuition We want to remind the reader of the remark that followed the Benamou-Brenier formula in section 2. There we noted that intuitively, a geodesic for the Wasserstein distance will try to move all particles along path that try to minimize the energy $\int_0^1 |\dot{\tilde{U}}|^2 \mathrm{d}t$. Together with the lemma from the previous subsection regarding curve energies, this yields that our distance is exactly the metric for which geodesics will try to move all vectors simultaneously as parallel as possible and along paths as short as possible! However, that dynamical transportation problem now takes place in the space of fuzzy vector fields. In general, there is no guarantee that when transporting a smooth vector field to another smooth vector field, the in-between states are still vector fields but generally we only get fuzzy vector fields.

4.2.5 Variational Analysis

We now want to revisit the above energy

$$\mathcal{E}(\gamma, U) := \int_0^1 |\dot{\gamma}|^2 + |\nabla_{\dot{\gamma}} U|^2 \mathrm{d}t$$

which shall be minimized over all smooth curves γ and all smooth unit vector fields U along γ. We want to perform variational analysis on this energy in order to gather more information on the nature of its minimizers. For this, we will vary the curve γ as well as the vector field U, for which we look for a critical point subject to $|U| = 1$ and fixed start and end points $U(0) = U^{(0)}$ and $U(1) = U^{(1)}$. By a variation with respect to U we

understand a curve (γ, U_s) that depends smoothly on a parameter $s \in (-\epsilon, \epsilon)$ and which satisfies $U_0 = U$ as well as $U_s(0) = U^{(0)}$ and $U_s(1) = U^{(1)}$. The derivative of $\mathcal{E}(\gamma, U_s)$ with respect to s at time $s = 0$ only depends on $V = \frac{\partial}{\partial s}_{|s=0} U_s$ (which at any point $p \in M$ is an element of $TT_pM \cong T_pM$ and can thus be understood as a vector field along γ as well) and $\frac{\partial}{\partial s}_{|s=0} \mathcal{E}(\gamma, U_s)$ will be denoted by $D\mathcal{E}(\gamma, U)(0, V)$. On the other hand, under a variation with respect to γ we understand a curve (γ_s, U_s) where U_s is parallel in a sense that will be made more precise below and $\frac{\partial}{\partial s}_{|s=0} \mathcal{E}(\gamma_s, U_s)$ will be denoted as $D\mathcal{E}(\gamma, U)(u, 0)$, where u is the vector field $\frac{\partial}{\partial s}_{|s=0} \gamma_s$ along γ. For the reader that is less familiar with variational analysis we refer to the first and second variation formulas of length (i.e in [20]). In particular the variation with respect to γ uses tricks similar to the ones from the computation of the second variation formula.
The energy $\mathcal{E} = \int |\dot\gamma|^2 + |\nabla_{\dot\gamma} U|^2 \mathrm{d}t$ splits into the two energies $\int |\dot\gamma|^2 \mathrm{d}t$ and $\int |\nabla_{\dot\gamma} U|^2 \mathrm{d}t$. Therefore, the computation can be split twofold: Considering both of these energies separately as well as varying with respect to γ and U. The variation of the first energy with respect to U is clearly zero (since U does not even appear in that term), while the variation with respect to γ is a classical, very well known variational problem (the first variation formula for the energy) which has the variation

$$-2 \int \langle \nabla_{\dot\gamma} \dot\gamma, u \rangle \mathrm{d}t$$

where u denotes the variational vector field along γ.
More interesting are the variations of the latter energy. Varying the latter energy with respect to U should intuitively yield that U turns with constant speed along γ. Let us make this more precise. Denote the variation of U by a vector field V along γ, i.e. the variation is given via $U_\epsilon = U + \epsilon V$. Since there is the constraint that $|U| = 1$ along γ, we can assume that $\langle U, V \rangle = 0$. Then for the variation at (γ, U) with respect to V, we get

$$\begin{aligned} D\mathcal{E}(\gamma, U)(0, V) &= \int 2 \langle \nabla_{\dot\gamma} U, \nabla_{\dot\gamma} V \rangle \mathrm{d}t \\ &= -2 \int \langle \nabla_{\dot\gamma} \nabla_{\dot\gamma} U, V \rangle \mathrm{d}t, \end{aligned}$$

where the zero in the argument means that no variation with respect to γ is taken. Here, the boundary term originating from $\int \frac{\partial}{\partial t} \langle \nabla_{\dot\gamma} U, V \rangle \mathrm{d}t = \langle \nabla_{\dot\gamma} U, V(1) \rangle - \langle \nabla_{\dot\gamma} U, V(0) \rangle$ vanishes as $V(0) = 0$ and $V(1) = 0$ due to the constraint on U at 0 and 1. From the above, we can now conclude that $\langle \nabla_{\dot\gamma} \nabla_{\dot\gamma} U, V \rangle = 0$ at a critical point. However, since $\langle U, V \rangle = 0$ and the tangent space is only two dimensional, this implies that

$$\nabla_\gamma \nabla_{\dot\gamma} U = \lambda(t) U$$

for some function λ along γ. In particular, then $\nabla_{\dot\gamma} \nabla_{\dot\gamma} U$ is collinear to U and thus orthogonal to $\nabla_{\dot\gamma} U$ (again since $|U| = 1$).

In fact, we can even compute $\nabla_{\dot\gamma}\nabla_{\dot\gamma}U$: We have $\langle\nabla_{\dot\gamma}\nabla_{\dot\gamma}U,U\rangle = \nabla_{\dot\gamma}\underbrace{\langle\nabla_{\dot\gamma}U,U\rangle}_{=0} - \langle\nabla_{\dot\gamma}U,\nabla_{\dot\gamma}U\rangle = -\langle\nabla_{\dot\gamma}U,\nabla_{\dot\gamma}U\rangle$ and so (again by $|U|=1$), we conclude that

$$\nabla_{\dot\gamma}\nabla_{\dot\gamma}U = -\langle\nabla_{\dot\gamma}U,\nabla_{\dot\gamma}U\rangle U = -|\nabla_{\dot\gamma}U|^2U = \lambda(t)U,$$

for $\lambda(t) = -|\nabla_{\dot\gamma}U|^2$.
Thus, it follows that

$$\nabla_{\dot\gamma}\langle\nabla_{\dot\gamma}U,\nabla_{\dot\gamma}U\rangle = 2\langle\nabla_{\dot\gamma}\nabla_{\dot\gamma}U,\nabla_{\dot\gamma}U\rangle = 0,$$

which implies our original claim that $|\nabla_{\dot\gamma}U|$ is constant (and so is λ) and therefore U turns with constant speed along γ.
A little bit more involved is the variation of the second energy with respect to γ. Denote the variational vector field along γ by u, then a standard notation is introduced by defining a function on two parameters on $[0,1]\times(\epsilon,\epsilon)$:

$$\Gamma(t,s) = \exp_{\gamma(t)}(su(t)).$$

Since the energy $\mathcal{E}$ is independent under reparametrisation, we may assume that u and $\dot\gamma$ are not collinear.
In that case Γ yields a local coordinate system that satisfies $\frac{\partial}{\partial s}_{|s=0}\Gamma(t,0) = u(t)$ and $\Gamma(t,0)=\gamma(t)$. It will be the basis of our analysis and it defines the derivatives $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial s}$. Next, we also define a vector field $U(t,s)$ by setting $U(t,0)$ to $U(t)$ and defining

$$U(t,s) = P^\Gamma_{0,s}U(t),$$

where $P^\Gamma_{0,s}$ denotes the parallel transport along the path $s\mapsto\Gamma(t,s)$. Then, for our variation we get

$$\begin{aligned}
D\mathcal{E}(\gamma,U)(u,0) &= \int 2\langle\nabla_{\dot\gamma}U,\frac{\nabla}{\partial s}\frac{\nabla}{\partial t}U(t,s)_{|s=0}\rangle\mathrm{d}t\\
&= 2\int\langle\nabla_{\dot\gamma}U,\frac{\nabla}{\partial t}\underbrace{\frac{\nabla}{\partial s}U(t,s)_{|s=0}}_{=0}\rangle\mathrm{d}t\\
&\quad + 2\int\langle\nabla_{\dot\gamma}U,\mathrm{Riem}(u(t),\dot\gamma(t))U(t)\rangle\mathrm{d}t\\
&= 2\int\mathcal{R}(u(t),\dot\gamma(t),U,\nabla_{\dot\gamma}U)\mathrm{d}t
\end{aligned}$$

where $\mathcal{R}(u,v,w,z) := \langle\mathrm{Riem}(u,v)w,z\rangle$ and Riem denotes the Riemann curvature tensor. Also note that $\frac{\nabla}{\partial s}U(t,s)$ vanishes since $U(t,s)$ was defined to be parallel in the $s-$coordinate. Recall the symmetries $\mathcal{R}(u,v,w,z) = \mathcal{R}(w,z,u,v)$, $\mathcal{R}(u,v,w,z) = -\mathcal{R}(u,v,z,w)$, and $\mathcal{R}(u,v,w,z) = -\mathcal{R}(v,u,w,z)$. In particular, we have

$$\begin{aligned}
\mathcal{R}(u(t),\dot\gamma(t),U,\nabla_{\dot\gamma}U) &= \mathcal{R}(\dot\gamma(t),u(t),\nabla_{\dot\gamma}U,U)\\
&= \mathcal{R}(\nabla_{\dot\gamma}U,U,\dot\gamma(t),u(t))
\end{aligned}$$

Since we are working with two-dimensional manifolds, we know that the Riemannian curvature tensor is just given by

$$R(u,v)w = K\langle u, Jv\rangle Jw,$$

where J denotes the rotation by 90 degrees as usual.
Finally, putting together the variations with respect to U for the two energies $\int |\dot\gamma|^2 \mathrm{d}t$ and $\int |\nabla_{\dot\gamma} U|^2 \mathrm{d}t$ we get for the variation of $\mathcal{E}$:

$$0 = 2\int -\langle \nabla_{\dot\gamma}\dot\gamma, u\rangle + K\langle \nabla_{\dot\gamma} U, JU\rangle \langle J\dot\gamma, u\rangle \mathrm{d}t.$$

Therefore, we can conclude that a minimizer of the above energy satisfies

$$\nabla_{\dot\gamma}\dot\gamma = K\langle \nabla_{\dot\gamma} U, JU\rangle J\dot\gamma.$$

In particular, this implies that γ is a constant speed curve since we know that

$$\begin{aligned}\nabla_{\dot\gamma}\langle\dot\gamma,\dot\gamma\rangle &= 2\langle\nabla_{\dot\gamma}\dot\gamma,\dot\gamma\rangle \\ &= 2K\langle\nabla_{\dot\gamma}U, JU\rangle \underbrace{\langle J\dot\gamma,\dot\gamma\rangle}_{=0} = 0.\end{aligned}$$

Note that for the derivative of the term $\langle \nabla_{\dot\gamma} U, JU\rangle$, we have

$$\nabla_{\dot\gamma}\langle\nabla_{\dot\gamma}U, JU\rangle = \langle\nabla_{\dot\gamma}\nabla_{\dot\gamma}U, JU\rangle - \underbrace{\langle\nabla_{\dot\gamma}U, \nabla_{\dot\gamma}JU\rangle}_{=\langle\nabla_{\dot\gamma}U, J\nabla_{\dot\gamma}U\rangle=0} = 0$$

since $\nabla_{\dot\gamma}\nabla_{\dot\gamma}U$ is collinear with U as we have seen above. Therefore, the term $\langle\nabla_{\dot\gamma}U, JU\rangle$ is constant as well. More precisely we can compute it to

$$|\langle\nabla_{\dot\gamma}U, JU\rangle| = |\nabla_{\dot\gamma}U| = const$$

since $|U| = 1$ and so $\nabla_{\dot\gamma}U$ is orthogonal to U and thus collinear with JU. Due to continuity, $\langle\nabla_{\dot\gamma}U, JU\rangle$ must be constant as well.
Since we also showed that λ can be computed to $-|\nabla_{\dot\gamma}U|^2$, we then get

$$\begin{aligned}\nabla_{\dot\gamma}\nabla_{\dot\gamma}U &= const_1 \cdot U \\ \nabla_{\dot\gamma}\dot\gamma &= K \cdot const_2 \cdot J\dot\gamma\end{aligned}$$

where $const_1 = -|\nabla_{\dot\gamma}U|^2$ and $const_2 = \pm|\nabla_{\dot\gamma}U|$.
By replacing $\dot\gamma$ with $\tau = \frac{\dot\gamma}{|\dot\gamma|}$, we can get a statement on the geodesic curvature of γ. Let us summarize these results in the following theorem.

Theorem 7. *Let (γ, U) be a critical point to the energy $\mathcal{E}$. Moreover, let $\tau := \frac{\dot\gamma}{|\dot\gamma|}$, then we have*

$$\begin{aligned}\nabla_\tau\tau &= \frac{1}{|\dot\gamma|^3} K \cdot const_2 \cdot J\tau \qquad \text{and} \\ \nabla_\tau\nabla_\tau U &= const \cdot U,\end{aligned}$$

for some constants const and $const_2$, where $const_2 = \langle\nabla_{\dot\gamma}U, JU\rangle$ which has absolute value $|\nabla_{\dot\gamma}U|$ (i.e. the turning velocity). In particular if the surface has constant Gauss-curvature, then the geodesic curvature κ_γ of γ is constant.

Proof. We showed that a critical point (γ, U) to the energy satisfies that γ is constant speed. In particular $|\dot{\gamma}|$ is constant and thus

$$\begin{aligned}\nabla_\tau \tau &= \nabla_{\frac{\dot{\gamma}}{|\dot{\gamma}|}} \frac{\dot{\gamma}}{|\dot{\gamma}|} \\ &= \frac{1}{|\dot{\gamma}|^2} \nabla_{\dot{\gamma}} \dot{\gamma} \\ &= \frac{1}{|\dot{\gamma}|^2} K \cdot const_2 \cdot J\dot{\gamma} \\ &= \frac{1}{|\dot{\gamma}|^3} K \cdot const_2 \cdot J\tau\end{aligned}$$

and thus

$$\begin{aligned}\kappa_\gamma &= \langle \nabla_\tau \tau, J\tau \rangle \\ &= \langle \frac{1}{|\dot{\gamma}|^2} K \cdot const_2 \cdot J\dot{\gamma}, J\frac{\dot{\gamma}}{|\dot{\gamma}|} \rangle \\ &= \frac{1}{|\dot{\gamma}|^3} K \cdot const_2 \cdot \langle J\dot{\gamma}, J\dot{\gamma} \rangle \\ &= \frac{1}{|\dot{\gamma}|} K \cdot const_2\end{aligned}$$

and, by assumption and as shown above, all of these terms are constant. Thus, the geodesic curvature of γ is constant. □

4.2.6 Relation to Elastic Rods

In the theory of elastic rods, one usually considers a unit-speed space curve $\gamma : [0,1] \to \mathbb{R}^3$ equipped with a frame (t, n, b), where $\dot{\gamma} = t$ and the frame (t, n, b) forms an orthonormal basis for any time $t \in [0,1]$. Physically, this describes a thick rod with centerline γ and a material cross section spanned by n and b. Clearly, the frame is already determined by t (the tangent) and n (the normal). For such a framed curve one considers two energies: The so-called bending energy $\int_0^1 |\dot{t}|^2 \mathrm{d}t$, which describes how strongly the physical curve is bending and the so-called torsion energy $\int_0^1 |\nabla_t n|^2 \mathrm{d}t$, which describes how strongly the curve is twisting. A compound energy

$$\alpha \int_0^1 |\dot{t}|^2 \mathrm{d}t + \beta \int_0^1 |\nabla_t n|^2 \mathrm{d}t$$

is then usually the subject of further study (see for example [6] or [4]).
We want to reformulate this slightly: First, note that since γ is unit-speed, we know that $t : [0,1] \to \mathbb{S}^2$ and if we know the starting point of γ, we can easily recover γ by integrating the curve t. So we might as well formulate everything in terms of t.
Now, the frame is determined by t and n. Thus, knowing t, we additionally have to give n

which for every time is a unit vector orthogonal to t. In other words, we can understand (t, n) as an element of the unit circle bundle of the sphere! Then for $a = b = 1$, the energy

$$\int_0^1 |\dot{t}|^2 \mathrm{d}t + \int_0^1 |\nabla_{\dot{t}} n|^2 \mathrm{d}t$$

is exactly the energy that we considered in the previous subsections for the manifold $M = \mathbb{S}^2$. In other words, the energy considered above is a generalization of the bending-torsion energy for elastic rods to a setting where the unit tangent t may live in a different manifold than $\mathbb{S}^2$.
Recall, that the Wasserstein geodesic between two Dirac measures at points x and y of a Riemannian manifold is given by a moving Dirac measure which moves along a geodesic between x and y. In particular, any algorithm that can solve the dynamical optimal transport problem for fuzzy vector fields will also be able to minimize the elastic-rod-energy. Note, however, that for elastic rods there is the additional constraint that the endpoints of the curve γ are fixed, which is not guaranteed by simply fixing the start- and endvectors.

4.2.7 A Dual Formulation

In this part, we want to consider the dual formulation for the Wasserstein distance on $\mathbb{S}M$ equipped with the Kaluza-Klein metric. Before we do that, we note that there is also a version for optimal transport for which the two measures in question may live on different spaces. In that case, we have two Polish spaces X, Y and a cost function $c : X \times Y \to \mathbb{R}$. Then the transport cost between two probability measures $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ is defined by

$$C(\mu, \nu) := \inf_{\rho} \int_{X \times Y} c(x, y) \mathrm{d}\rho,$$

where the infimum is taken over all ρ that have μ and ν as their marginals, i.e. $(\mathrm{proj}_1)_{\#}\rho = \mu$ and $(\mathrm{proj}_2)_{\#}\rho = \nu$.
Then, with the right assumptions, we again have a dual formulation:

Proposition 6. *Assume that the cost function $c : X \times Y \to \mathbb{R}$ is continuous and bounded from below and that there exist two functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$ such that*

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

Then, we have

$$C(\mu, \nu) = \sup_{f,g} \int_X f \mathrm{d}\mu + \int_Y g \mathrm{d}\nu,$$

where the infimum is taken over all functions $f \in L^1(\mu), g \in L^1(\nu)$ that satisfy

$$f(x) + g(y) \leq c(x, y) \text{ for all } x \in X, y \in Y.$$

This is well-known and can, for example, be found in [2].
Now, consider the Wasserstein distance on $\mathbb{S}M$ for two smooth vector fields X and Y with $\|X\|_{L^1} = \|Y\|_{L^1}$:

$$W_2(\mu_X, \mu_Y) = \left(\inf_{\rho \in \Pi(\mu_X, \mu_Y)} d^2_{\mathbb{S}M}(v, w) \mathrm{d}\rho(v, w) \right)^{1/2}.$$

The infimum is an optimal transport problem, which we can restrict to the support of μ_X and μ_Y, respectively, i.e. to the images of the sections s_X and s_Y. One easily checks that these spaces then fulfil the assumption of the proposition above. Therefore, we get

$$\begin{aligned} W_2(\mu_X, \mu_Y)^2 &= \sup_{\tilde{f} m \tilde{g}} \int_{im(s_X)} \tilde{f} \mathrm{d}\mu_X + \int_{im(s_Y)} \tilde{g} \mathrm{d}\mu_Y \\ &= \sup_{\tilde{f} m \tilde{g}} \int_{im(s_X)} \tilde{f} \mathrm{d}(s_X)_\#(|X|\mathrm{dvol}) + \int_{im(s_Y)} \tilde{g} \mathrm{d}(s_Y)_\#(|Y|\mathrm{dvol}) \\ &= \sup_{\tilde{f} m \tilde{g}} \int_M \tilde{f} \circ s_X \cdot |X| \mathrm{dvol} + \int_M \tilde{g} \circ s_Y \cdot |Y| \mathrm{dvol}, \end{aligned}$$

where the supremum is taken over all $\tilde{f}, \tilde{g}$ satisfying $\tilde{f}(v) + \tilde{g}(w) \leq d^2_{\mathbb{S}M}(v, w)$. Now we can define $f := \tilde{f} \circ s_X$ and $G = \tilde{g} \circ s_Y$ yielding

$$W_2(\mu_X, \mu_Y)^2 = \sup_{f,g} \int f \cdot |X| \mathrm{dvol} + \int g \cdot |Y| \mathrm{dvol},$$

where the supremum is taken over all functions f, g satisfying $f(x) + g(y) \leq d^2_{\mathbb{S}M}(s_X(x), s_Y(y))$. This is clearly yet another optimal transport cost, this time between the measures $|X|\mathrm{dvol}$ and $|Y|\mathrm{dvol}$, but for a different cost function $c(x, y) = d^2_{\mathbb{S}M}(s_X(x), s_Y(y))$.
This formula could be interesting for numerical computations since it shows that for smooth vector fields, the computation of the distance can be formulated as an optimization problem over M (instead of $\mathbb{S}M$), given that we have an explicit formula for $d^2_{\mathbb{S}M}(s_X(x), s_Y(y))$.

4.2.8 A Benamou-Brenier Formula

In this part, we assume the special case where the Wasserstein geodesic between two smooth vector fields in the space of fuzzy vector fields is a smooth vector field for all times. In a similar vein to the computations above, we can then use the Benamou-Brenier formula and pull the terms back to the base manifold in order to get a simplified formula.
For this, let

- X_t be a time-dependent vector field on M with $\int_M |X_t| \mathrm{dvol} = 1$ for all times $t \in [0, 1]$,
- $s_t = s_{X_t} = \frac{X_t}{|X_t|}$, $\mu_t = |X_t|\mathrm{dvol}$ and $\tilde{\mu_t} = (s_t)_\# \mu_t$ as above,
- v_t time dependent vector field on $\mathbb{S}M$ related to $\tilde{\mu_t}$ via the continuity equation.

Recall the Benamou-Brenier formula

$$W(\tilde{\mu_0},\tilde{\mu_1})^2 = \inf_{v_t}\{\int_0^1\int_{\mathbb{S}M}|v_t|^2\mathrm{d}\tilde{\mu_t}\mathrm{d}t \quad | \quad \frac{\partial}{\partial t}\tilde{\mu_t}+\nabla(v_t\tilde{\mu})=0\}$$

Now,we want to 'solve' the push-forwards.
The energy-term then becomes

$$\int_{\mathbb{S}M}|v_t|^2\mathrm{d}\tilde{\mu_t} = \int_{\mathbb{S}M}|v_t|^2\mathrm{d}(s_t)_{\#}\mu_t = \int_M|v_t|^2\mathrm{d}\mu_t,$$

where the vector field and its norm are both on $\mathbb{S}M$.
In the continuity equation

$$\frac{\partial}{\partial t}\int_{\mathbb{S}M}\phi\mathrm{d}\tilde{\mu_t} = \int_{\mathbb{S}M}\langle\nabla\phi,v_t\rangle\mathrm{d}\tilde{\mu_t}$$

there are two integrals to investigate. Before doing that, let us first define the vertical vector field along $im(s_t)$: $u := \frac{\partial}{\partial t}s_t$ and write $|X_t|'$ for $\frac{\partial}{\partial t}|X_t|$. Then we have

$$\frac{\partial}{\partial t}\int_{\mathbb{S}M}\phi\mathrm{d}(s_t)_{\#}\mu_t = \frac{\partial}{\partial t}\int_M\phi\circ s_t\mathrm{d}\mu_t \ = \int_M\langle\nabla\phi,u\rangle_{s_t}|X_t|\mathrm{dvol}+\int_M\phi\circ s_t\cdot|X_t|'\mathrm{dvol}$$

and

$$\int_{\mathbb{S}M}\langle\nabla\phi,v_t\rangle\mathrm{d}(s_t)_{\#}\mu_t = \int_M\langle\nabla\phi,v_t\rangle_{s_t}|X_t|\mathrm{dvol}.$$

Putting them together yields the equation

$$\int_M\phi\circ s_t\cdot|X_t|'\mathrm{dvol}+\int_M\langle\nabla\phi,u-v_t\rangle_{s_t}|X_t|\mathrm{dvol}=0$$

for all smooth and compactly supported test functions $\phi\in C(\mathbb{S}M)$. However, the first integral only depends on ϕ on the image of s_t. In particular, the term $u-v_t$ has to be tangent to the image of s_t in $\mathbb{S}M$. Now consider the projection $w_t := \mathrm{d}\pi(u-v_t)$, which is a vector field in M. Clearly, $u-v_t$ can be recovered as $u-v_t = Ds_t(w_t)$. If we define $\psi=\phi\circ s_t$, we then have

$$\langle\nabla\phi,u-v_t\rangle_{s_t} = \langle\nabla\phi,Ds_t(w_t)\rangle_{s_t} = Ds_t(w_t)(\phi) = w_t(\phi\circ s_t) = \langle\nabla\psi,w_t\rangle.$$

In particular, we see that the measure curve $|X_t|\mathrm{dvol}$ satisfies the continuity equation

$$\frac{\partial}{\partial t}\int_M\psi\cdot|X_t|\mathrm{dvol}+\int_M\langle\nabla\psi,w_t\rangle|X_t|\mathrm{dvol}=0$$

for a different vector field w_t on M. Now, we also express v_t in terms of w_t and decompose it into horizontal and vertical parts to compute its squared norm

$$|v_t|^2 = |u-Ds_t(w_t)|^2 = |w_t|^2+|u-\nabla_{w_t}s_t|^2,$$

where $\nabla_{w_t}s_t$ denotes the covariant derivative of the unit vector field s_t interpreted as an infinitesimal rotation and thus an element in the vertical space.
Thus, the measure curve $|X_t|\mathrm{dvol}$ minimizes the energy

$$\int_M|w_t|^2+|\frac{\partial}{\partial t}s_t-\nabla_{w_t}s_t|^2|X_t|\mathrm{dvol},$$

subject to all w_ts that satisfy the above continuity equation.

4.2.9 An Extension to Cone Spaces

A possible generalization of the previously introduced idea will be presented now. This generalization is based on the underlying intuition of a vector consisting of a length and a direction, which was also used above. Let us start by defining a cone space.

Definition 10. *Let M be a manifold. Consider a fibre bundle $\pi : B \to M$ over M with fibre F. Then a* cone space $C(B)$ *is defined as a fibre bundle over M whose fibres consist of cones of the fibres of B. More precisely, $C(B)$ is defined as*

$$C(B) = (B \times \mathbb{R}_{\geq 0}) \,/\, \sim,$$

where $(f,t) \sim (g,s)$ iff $t = s = 0$ and $\pi(f) = \pi(g)$ and the projection to M is defined via $[(f,t)] \mapsto \pi(f)$ (where $[(f,t)]$ is the equivalence class of (f,t)).

Let us assume that the fibre $B_m \cong F$ and the fibre bundle B are manifolds themselves. Further, we want to impose that on the fibres B_m we can consistently define a Riemannian metric. That is, we equip the fibre F with a Riemannian metric and require the transition functions for two trivializations ϕ_i, ϕ_j

$$\phi_i \circ \phi_j^{-1} = (\mathrm{Id}, t_{ij}) : U_i \cap U_j \times F \to U_i \cap U_j \times F$$

to be such that all the t_{ij}s are isometries. In other words, we want the group of isometries of F as the structure group for the bundle.
Additionally, let us assume that B comes with a connection, i.e. there is a splitting on each tangent space $T_f B = H_f \oplus V_f$ such that $\mathrm{d}\pi_{|H_f} : H_f \to T_{\pi(f)} M$ is an isomorphism and $V_f = \ker \mathrm{d}\pi$ for each $f \in B$. In this case, we can define a metric on B in the same way as we did in the previous section. In accordance with the nomenclature from above, let us call this metric *Kaluza-Klein metric* as well.

Definition 11. *For a Riemannian manifold M and a fibre bundle $\pi : B \to M$ satisfying the above assumptions, we can define the* Kaluza-Klein-metric *on B by using the metric of M on the horizontal space H_f (using the isomorphism $\mathrm{d}\pi_{|H_f}$) and the metric of the fibre F on the vertical space V_f.*
The latter can be done using a local trivialization

$$\phi_j^{-1}(m, \cdot) : \{m\} \times F \cong F \to B_m.$$

This is an isometry and clearly $im(\mathrm{D}_{\tilde{f}} \phi_j^{-1}(m,\cdot)) = \ker \mathrm{d}\pi_f = V_f$ for $\phi_j^{-1}(m,\cdot)(\tilde{f}) = f$. So $\mathrm{D}\phi_j^{-1}(m,\cdot)$ induces a metric on V and since our structure group is given by isometries of F, this metric does not depend on the chosen trivialization.
By requiring that H_f and V_f are orthogonal for each f, this then yields a metric for B.

With this definition we can now define a Wasserstein-like distance for sections of the cone bundle $C(B)$ just the same way we did for vector fields:

Definition 12. *Let $\pi : B \to M$ be a fibre bundle over a Riemannian manifold M satisfying all the assumptions above. For two sections $s_1 = [(f_1, t_1)], s_2 = [(f_2, t_2)]$ of $C(B)$ satisfying $\|t_1\|_{L^1(M)} = \|t_2\|_{L^1(M)} = 1$, we define a distance via*

$$W_2(s_1, s_2) := W_2^B(\mu_{s_1}, \mu_{s_2}),$$

where μ_{s_1} and μ_{s_2} are defined by $\mu_{s_1} = (f_1)_\#(t_1 \cdot \mathrm{dvol}_M), \mu_{s_2} = (f_2)_\#(t_2 \cdot \mathrm{dvol}_M)$ for the Riemannian volume form dvol_M. Note that, just as in section 4.2.4, f_1 might not be well-defined whenever t_1 is zero (and similarly for f_2 and t_2). But since there the measure $t_1 \cdot \mathrm{dvol}_M$ has no mass, the push-forwards are still well-defined.

Note, that if we understand the tangent bundle TM over a Riemannian manifold M as the cone bundle of the unit circle bundle $\mathbb{S}M$, this definition recovers exactly the definition from section 4.2.4.
However, since a Riemannian metric g on M also induces a metric on the k-exterior algebra of T^*M as well as a splitting into vertical and horizontal spaces of the bundle $\Lambda^k T^*M$ (via the Levi-Civita connection), we can here too view the space $\Lambda^k T^*M$ as a cone bundle of the space of k-forms of norm 1. In particular, this also leads to a definition of a Wasserstein-like distance on k-forms. Understanding this distance could be an interesting subject for further research.

5 A Possible Application

For the distance on vector fields we came up with one particular application, which will be described in the following: The Wasserstein distance is often used to interpolate images. For this use case, two grayscale images are interpreted as functions (or rather as measures after potential rescaling) and then used for dynamical optimal transport. The Wasserstein distance seems to be indeed the *correct* metric to do that. However, for natural images (like a waving arm), this can lead to distorsions since the Wasserstein metric will always move particles along geodesics. In particular a circular movement (like the waving arm) could not be replicated by Wasserstein interpolation and instead we get movement along a straight line (so the arm would change its length and thus not follow a natural movement). Another approach could therefore be to take the gradient of two grayscale images and, instead of transporting the images, transport their gradients using the distance we introduced in the previous section. This interpolation yields vector fields and the interpolated states of the grayscale images can then be recovered by solving the minimisation problem $\min_f \int |\nabla f - X|^2$ with some scaling constraints on f and where X denotes the interpolated vector field, which leads to a Poisson problem.
However, recall that the optimal transport of vector fields tries to move every vector on a path that minimizes the energy $\mathcal{E}$. For the Euclidean case though, a minimizer for the energy

$$\mathcal{E}(\gamma, U) = \int_M |\dot{\gamma}|^2 + |\nabla_{\dot{\gamma}} U|^2 \mathrm{d}t$$

is given by γ just being a straight line and U turning with constant speed from a given angle to its target angle along the shortest path in $\mathbb{S}^1$ (this follows immediately from the variational analysis of the energy). In particular, this would not solve the above mentioned defect since particles of an image would still travel along straight lines, which we wanted to avoid. Only if the manifold M has curvature, there exists holonomy and thus it might indeed be better to travel along a non-geodesic line in order to turn the vector U less (since then, compared to the geodesic path, some of the turning can be compensated by travelling more parallel along a less optimal path). Therefore, the application described above can only yield interesting results for curved surfaces M.
Thus, we are interested in a numerical treatment of the Wasserstein problem on fuzzy vector fields for a curved base manifold. For this we see two fundamental approaches:

Discretisation of the whole Bundle Space One idea is to discretise the whole unit circle bundle. More precisely, we can consider a discrete surface given by a (compact, oriented) triangulated mesh embedded in $\mathbb{R}^3$ and give a concept for a discrete unit circle bundle. We want to interpret vector fields again as constant per triangle, i.e. a vector field is indexed by the triangles and at a triangle f it is also tangent to f as in section 3. The circle bundle with smooth fibre then consists of a unit circle on each triangle. A discrete connection is given by an $\mathbb{S}^1$-equivariant map for each edge between the respective unit circles of the adjacent triangles (this map represents the parallel transport of a circle

to a circle on a neighbouring triangle). A discrete Levi-Civita-connection is defined by the following $\mathbb{S}^1$-equivariant map: for two adjacent triangles we first isolate them from the structure and turn one triangle, using their common edge as a hinge, such that both triangles lie in a plane. Then just move one circle to the other parallelly in that plane and turn the triangles back.

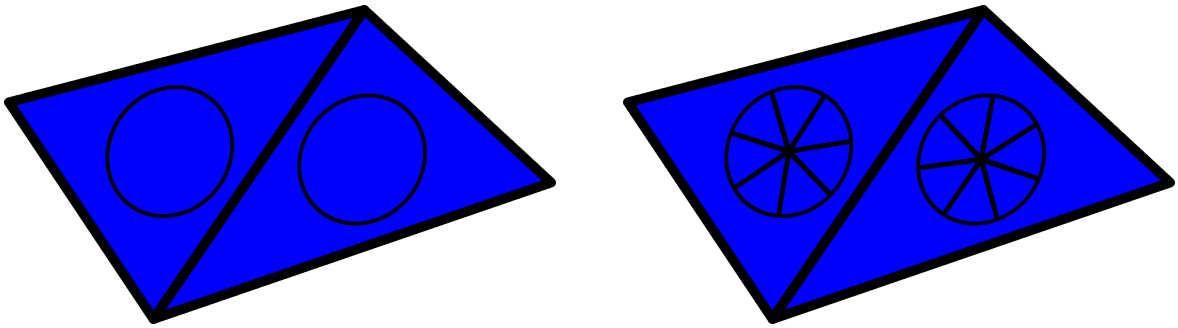

Figure 2: Discrete circle bundle with smooth and discrete fibre.

With this definition, one can recover holonomy by transporting a point of a circle (a unit vector) parallelly along a closed path of triangles (or more precisely, along a path in the dual graph) and measuring the angular difference of the resulting point to the starting point. If performed around a vertex we recover the usual angle deficit as a discrete curvature notion.
In order to discretise the whole bundle, one can now additionally discretise the fibres over each triangle equidistantly (say into l discrete steps such that multiplication with $2\pi/l$ would rotate every discrete step of a circle into the next one with respect to the orientation). Another way to visualize this as a 3-dimensional bundle over the triangulated surface, would be to think of prisms stacked onto each other in a way that the lth prism lies beneath the first one again (see figure 3).
However, generally there is no way to do these discretisations of each fibre in such a way that the discrete steps of every circle get mapped to the discrete steps of every adjacent circle via the connection, since that would imply that the recoverable holonomies can only be multiples of the discretisation step $2\pi/l$. But the holonomies (e.g. the angle deficit) can clearly take any value in $\mathbb{S}^1$. Thus, we consider a discretisation of the unit circle bundle in terms of collections of prisms over each triangle in such a way that the prisms over adjacent triangles will be shifted to each other (when compared using the connection). By constructing incidence relations and discrete differential operators for this space one can then treat this space numerically.
This is a triangle based discretisation in the sense that the fibres live on the triangles. Alternatively, one could also construct a vertex based discretisation of the unit circle bundle and a similar thing is actually done in [17] (without the discretization of the fibre).
One could then try to run an adapted heat method (as in [15]) on this space in order to compute distances on this space and then use this to solve the static optimal transport problem.
Moreover, one could try to adapt the common algorithms for the dynamical optimal

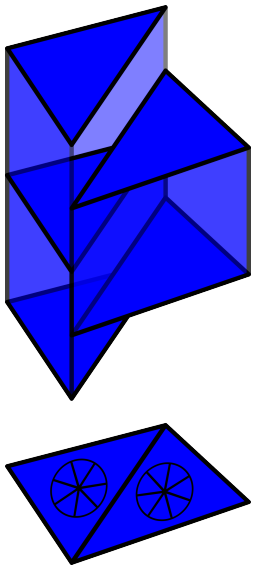

Figure 3: Total space of the discrete circle bundle with discrete fibre.

transport problem based on discretisations of the Benamou-Brenier formulation as in [32]. However, all of these approaches rely on certain discrete differential operators (in particularly a discrete Laplacian) and therefore they seem to be rather slow since the dimension of the unit circle bundle is three-dimensional, which means that the dimension of a discrete Laplacian and related discrete differential operators will be rather large.

Discretisation of the Measures An alternative and simpler approach is to discretise the measures but keep the circle bundle smooth. Assume that the measures are given by collections of particles of different masses which in our interpretation corresponds to a collection of vectors whose sum of norms is 1. If we can explicitly compute the distances between the particles of the two measures, we can solve the static optimal transport problem using the Sinkhorn algorithm, a standard algorithm for the static optimal transport problem (for a detailed description of the algorithm refer to [34]).
This algorithm yields an optimal transport plan. Therefore, given that we can also efficiently compute the shortest geodesic for any given two points in our bundle, we can also treat the dynamical optimal transport problem this way.
In the following we will consider this approach for the special case of the hyperbolic disk.

5.1 The Hyperbolic Case

Of the three model geometries in two dimensions, we noted that the Euclidean case is not interesting for applications, while for the spherical case, the minimization of the energy involved, is equivalent to the problem of finding elastic rods, which has been researched with by many people. For this reason we focus on the hyperbolic case, i.e. the hyperbolic

disk $\mathbb{D}^2$ with constant curvature -1. As mentioned above, for the approach for discretised measures, we need to be able to do two things: compute the distance in $\mathbb{SD}^2$ and finding shortest geodesics on $\mathbb{SD}^2$. These formulas will be derived here. The computations use well-known facts from 2-dimensional hyperbolic geometry, for which we refer the reader to [35].

5.1.1 The Hyperbolic Disk

We will make all our computations using the hyperbolic disk model $\mathbb{D}^2$. Then the unit circle bundle can be parametrised by a point in $\mathbb{D}^2$ and an angle ϕ, that corresponds to the unit vector, which has the angle ϕ with the x-axis. Note, however, that a constant angle ϕ along a curve γ is by no means parallel in the hyperbolic space. For example, for a circle that hits the boundary of the disk in right angles (i.e. a geodesic in $\mathbb{D}^2$), an angle with constant angle to the derivative of γ would be parallel instead.
Since any pair of points with a given distance can be mapped isometrically to any other pair of points with the same distance in $\mathbb{D}^2$, we can without loss of generality assume that our starting points z_0, z_1 lie on the x-axis and have the same distance to the origin. For two angles, ϕ_0 sitting at z_0, and ϕ_1 sitting at z_1, we want to find the geodesic in $\mathbb{SD}^2$. Computing their distance then becomes a trivial task.
Recall the variational analysis we did in section 4. There we showed that for constant curvature, a critical point of $\mathcal{E}$ will satisfy that γ has constant geodesic curvature. On the hyperbolic disk, these curves are circles and there are three kinds of circles:

- Circles contained in $\mathbb{D}^2$ which are just called circles or cycles. Those are indeed circles in the sense that there exists a point z such that the circle consists exactly of the points in $\mathbb{D}^2$ that have a given (hyperbolic) distance to z.
- Circles that touch the boundary of $\mathbb{D}^2$. These are called horocycles.
- Circles that intersect the boundary of $\mathbb{D}^2$ at two points. These are called hypercycles and geodesics would be one example.

So in our case, we have to consider all the circles that pass through z_0 and z_1. For simplicity, let us consider only those circles for which the part below the x-axis is finite (i.e. does not leave the disk) and parametrise the circle such that the curve γ is unit speed, starts at z_0, ends at z_1, and has the part of the circle below the x-axis as its image. These curves will always have positive constant geodesic curvature and can be parametrised by their angle θ at z_1 with the x-axis (i.e. $\gamma = \gamma_\theta$).
Denote the geodesic curvature of γ_θ by $c(\theta)$ and its length (between z_0 and z_1) by $L(\theta)$. Note that in the proof of the theorem for the variation of the energy, we showed that

$$c(\theta) = \kappa_\gamma = \frac{1}{|\dot{\gamma}|} K \cdot const_2,$$

where K is the curvature and $const_2 = \langle \nabla_{\dot\gamma} U, JU \rangle$ (which has the turning velocity as absolute value). Since in this case, $K = -1$ and $|\dot\gamma_\theta| = 1$ and $c(\theta) > 0$, we have

$$c(\theta) = -\langle \nabla_{\dot\gamma} U, JU \rangle.$$

In particular, U is turning away from γ_θ with velocity $c(\theta)$.
What does that mean for ϕ_t? Well, since $\dot\gamma_\theta$ is turning positively with speed $c(\theta)$, a parallel vector field along γ would be given by vector field $\tilde\phi_t$ whose angle to γ is given by $-c(\theta) \cdot t$ for all times $t \in [0, L(\theta)]$. But since we just established that for a critical point, the vector field has to point away from γ with speed $c(\theta)$, we can conclude that the angle of ϕ_t and γ_θ will be $-2c(\theta) \cdot t$ for all times $t \in [0, L(\theta)]$.
Since this angle difference is given by $\phi_0 + \theta$ at time $t = 0$ and $\phi_1 - \theta$ at time $t = L(\theta)$, we get the following equation: $\phi_0 + \theta - 2c(\theta)L(\theta) = \phi_1 - \theta$ or, equivalently

$$\phi_1 - \phi_0 = 2\theta - 2c(\theta)L(\theta).$$

If we can compute $c(\theta)$ and $L(\theta)$, this becomes an equation, which we can solve (numerically) for θ and from that we can recover the curve γ_θ as well as its length $L(\theta)$ and thus find the geodesic in $\mathbb{SD}^2$. Computing these is a rather standard computation in hyperbolic geometry. The computations differ slightly depending on whether γ_θ is a cycle, a horocycle or a hypercycle.

5.1.2 Length and Curvature

We have to distinguish three cases. In all of them we need the two geodesics through z_0 and z_1 that cut the circle at right angles. Let us call these η^0 and η^1. Moreover, denote the geodesic connecting z_0 and z_1 by η (i.e. the part of the x-axis between them) and the distance of z_0 and z_1 by a.

Cycle In this case, the curves η^0 and η^1 meet in a point, which is the (hyperbolic) midpoint of the circle γ.
Let us denote the angle between η^1 and η by ψ, then $\theta + \pi/2 + \psi = \pi$, i.e. $\psi = \pi/2 - \theta$.
Next, we connect the midpoint of η with the center of γ, a curve of length h. Then we consider the triangle (center of γ, midpoint of η, z_1). By symmetry, we have a right angle at the midpoint of η (i.e. the origin) and we denote the angle at the center of γ by $\chi/2$.
Then by the formulas of hyperbolic trigonometry, we have $\tan(\psi) = \frac{\tanh(h)}{\sinh(\frac{a}{2})}$. Thus, we can determine h via

$$h = \operatorname{arctanh}(\tan(\psi)\sinh(\frac{a}{2})).$$

The same formula can be applied to $\chi/2$ so that $\tan(\chi/2) = \frac{\tanh(\frac{a}{2})}{\sinh(h)}$. Thus, we can determine χ via

$$\chi = 2\arctan(\frac{\tanh(\frac{a}{2})}{\sinh(h)}).$$

Moreover, by the cosine theorem from hyperbolic geometry for this triangle, we get $\cosh(r) = \cosh(h)\cosh(\frac{a}{2})$, where r is the radius of the circle, i.e. the length of the curve η^1 from the center of γ to z_1. Thus,

$$r = \operatorname{arccosh}(\cosh(h)\cosh(\frac{a}{2})).$$

From that we can finally compute the length

$$L = \chi \sinh(r)$$

and the curvature

$$c = \frac{1}{\tanh(r)},$$

from standard equations from hyperbolic geometry.

Horocycle For horocycles, the geodesic curvature is always 1 and the length can be computed as $2\sinh(a/2)$, where a is the distance between the points.

Hypercycle In this setting, the two curves η^0 and η^1 do not intersect. However, the two points on the boundary of the disk where γ hits the boundary can be joined by a geodesic called the axis of γ and η^0 and η^1 will intersect it in right angles at some points z_0' and z_1'. The distance from γ to the axis is denoted by r and can be seen as the length of η^1 connecting z_1 and z_1'. If we connect the midpoint m of η with the midpoint m' of z_0' and z_1', then the quadrilateral m, m', z_1, z_1' is a so-called Lambert quadrilateral. We call the length of the halved sides of this quadrilateral $a' = a/2$ and $d' = d/2$ where d is the distance between z_0' and z_1'. In these Lambert quadrilaterals, we have the relation $\sin(\psi) = \frac{\cosh(d')}{a'}$. Therefore, we have

$$d' = \operatorname{arccosh}(\sin(\psi)\cosh(a')).$$

Moreover, we can define r as the distance of the hypercycle to the axis. Then, there holds $\cos(\psi) = \tanh(a')\tanh(r)$ and thus

$$r = \operatorname{arctanh}(\frac{\cos(\psi)}{\tanh(a')}).$$

Finally, the length of the part of γ connecting z_0 and z_1 can be computed from the formula

$$L = d\cosh(r)$$

and the geodesic curvature is given by $\tanh(r)$.

6 References

[1] Amari, S. I. (2016). *Information geometry and its applications* (Vol. 194). Springer.

[2] Ambrosio, L., Gigli, N. (2013). *A user's guide to optimal transport*. In Modelling and optimisation of flows on networks (pp. 1-155). Springer, Berlin, Heidelberg.

[3] Bacon, X. (2019). *Optimal transportation of vector-valued measures*. arXiv preprint arXiv:1901.04765.

[4] Bartels, S., Reiter, P. (2020). *Numerical solution of a bending-torsion model for elastic rods*. Numerische Mathematik, 146(4), 661-697.

[5] Beckmann, M. (1952). *A continuous model of transportation*. Econometrica: Journal of the Econometric Society, 643-660.

[6] Bergou, M., Wardetzky, M., Robinson, S., Audoly, B., Grinspun, E. (2008). *Discrete elastic rods*. In ACM SIGGRAPH 2008 papers (pp. 1-12).

[7] Braess, D. (2013). *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer-Verlag.

[8] Brenier, Y., Duan, X. (2018). *An integrable example of gradient flow based on optimal transport of differential forms*. Calculus of Variations and Partial Differential Equations, 57(5), 1-16.

[9] Brenier, Y., Vorotnikov, D. (2020). *On optimal transport of matrix-valued measures*. SIAM Journal on Mathematical Analysis, 52(3), 2849-2873.

[10] Chavel, I. (2006). *Riemannian geometry: a modern introduction* (Vol. 98). Cambridge university press.

[11] Chen, Y., Georgiou, T. T., Tannenbaum, A. (2017). *Matrix optimal mass transport: a quantum mechanical approach*. IEEE Transactions on Automatic Control, 63(8), 2612-2619.

[12] Chen, Y., Georgiou, T. T., Tannenbaum, A. (2018). *Vector-valued optimal mass transport*. SIAM Journal on Applied Mathematics, 78(3), 1682-1696.

[13] Chen, Y., Haber, E., Yamamoto, K., Georgiou, T. T., Tannenbaum, A. (2018). *An efficient algorithm for matrix-valued and vector-valued optimal mass transport*. Journal of Scientific Computing, 77(1), 79-100.

[14] Ciosmak, K. J. (2021). *Optimal transport of vector measures*. Calculus of Variations and Partial Differential Equations, 60(6), 1-22.

[15] Crane, K., Weischedel, C., Wardetzky, M. (2017). *The heat method for distance computation*. Communications of the ACM, 60(11), 90-99.

[16] Dacorogna, B., Gangbo, W., Kneuss, O. (2015). *Optimal transport of closed differential forms for convex costs*. Comptes Rendus Mathematique, 353(12), 1099-1104.

[17] Knöppel, F., Pinkall, U. (2016). *Complex line bundles over simplicial complexes and their applications*. In Advances in Discrete Differential Geometry (pp. 197-239). Springer, Berlin, Heidelberg.

[18] Lavenant, H., Claici, S., Chien, E., Solomon, J. (2018). *Dynamical optimal transport on discrete surfaces*. ACM Transactions on Graphics (TOG), 37(6), 1-16.

[19] Lee, J. M. (2003). *Introduction to Smooth Manifolds*. Springer, New York, NY.

[20] Lee, J. M. (2006). *Riemannian manifolds: an introduction to curvature* (Vol. 176). Springer Science and Business Media.

[21] Li, W. (2022). *Transport information geometry: Riemannian calculus on probability simplex*. Information Geometry, 5(1), 161-207.

[22] Mainini, E. (2012). *A description of transport cost for signed measures*. Journal of Mathematical Sciences, 181(6), 837-855.

[23] Martin, G. E. (2012). *The foundations of geometry and the non-Euclidean plane*. Springer Science and Business Media.

[24] Ning, L., Georgiou, T. T. (2014, December). *Metrics for matrix-valued measures via test functions*. In 53rd IEEE Conference on Decision and Control (pp. 2642-2647). IEEE.

[25] Otto, F. (2001). *The geometry of dissipative evolution equations: the porous medium equation*.

[26] Padilla, M., Chern, A., Knöppel, F., Pinkall, U., Schröder, P. (2019). *On bubble rings and ink chandeliers*. ACM Transactions on Graphics (TOG), 38(4), 1-14.

[27] Sasaki, S. (1962). *On the differential geometry of tangent bundles of Riemannian manifolds II*. Tohoku Mathematical Journal, Second Series, 14(2), 146-155.

[28] Albuquerque, R. (2019). *Notes on the Sasaki metric*. Expositiones Mathematicae, 37(2), 207-224.

[29] Abbassi, M. T. K., Calvaruso, G. (2012). *g-natural metrics of constant curvature on unit tangent sphere bundles*. Archivum Mathematicum, 48(2), 81-95.

[30] Abbassi, M. T. K., Sarih, M. (2005). *On natural metrics on tangent bundles of Riemannian manifolds*. Archivum Mathematicum, 41(1), 71-92.

[31] Boeckx, E., Vanhecke, L. (2001). *Unit tangent sphere bundles with constant scalar curvature*. Czechoslovak Mathematical Journal, 51, 523-544.

[32] Papadakis, N., Peyré, G., Oudet, E. (2014). *Optimal transport with proximal splitting.* SIAM Journal on Imaging Sciences, 7(1), 212-238.

[33] Peyré, G., Chizat, L., Vialard, F. X., Solomon, J. (2019). *Quantum entropic regularization of matrix-valued optimal transport.* European Journal of Applied Mathematics, 30(6), 1079-1102.

[34] Peyré, G., Cuturi, M. (2017). *Computational optimal transport.* Center for Research in Economics and Statistics Working Papers, (2017-86).

[35] Ratcliffe, J. G., Axler, S., Ribet, K. A. (1994). *Foundations of hyperbolic manifolds* (Vol. 149, pp. xii+-747). New York: Springer.

[36] Ryu, E. K., Chen, Y., Li, W., Osher, S. (2018). *Vector and matrix optimal mass transport: theory, algorithm, and applications.* SIAM Journal on Scientific Computing, 40(5), A3675-A3698.

[37] Sawhney, R., Crane, K. (2020). *Monte Carlo geometry processing: A grid-free approach to PDE-based methods on volumetric domains.* ACM Transactions on Graphics, 39(4).

[38] Schmitzer, B. (2019). *Stabilized sparse scaling algorithms for entropy regularized transport problems.* SIAM Journal on Scientific Computing, 41(3), A1443-A1481.

[39] Simon, D., Aberdam, A. (2020). *Barycenters of natural images constrained wasserstein barycenters for image morphing.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7910-7919).

[40] Solomon, J., Rustamov, R., Guibas, L., Butscher, A. (2014). *Earth mover's distances on discrete surfaces.* ACM Transactions on Graphics (ToG), 33(4), 1-12.

[41] Villani, C. (2009). *Optimal transport: old and new* (Vol. 338, p. 23). Berlin: springer.

[42] Wardetzky, M. (2007). *Discrete Differential Operators on Polyhedral Surfaces - Convergence and Approximation* (Doctoral dissertation, Freie Universität Berlin).

[43] Zhu, L., Yang, Y., Haker, S., Tannenbaum, A. (2007). *An image morphing technique based on optimal mass preserving mapping.* IEEE Transactions on Image Processing, 16(6), 1481-1495.