On a linear Gromov–Wasserstein distance

Florian Beier, Robert Beinert, Gabriele Steidl

***Abstract*—Gromov–Wasserstein distances are generalization of Wasserstein distances, which are invariant under certain distance preserving transformations. Although a simplified version of optimal transport in Wasserstein spaces, called linear optimal transport (LOT), was successfully used in certain applications, there does not exist a notation of linear Gromov–Wasserstein distances so far. In this paper, we propose a definition of linear Gromov–Wasserstein distances. We motivate our approach by a generalized LOT model, which is based on barycentric projection maps of transport plans. Numerical examples illustrate that the linear Gromov–Wasserstein distances, similarly as LOT, can replace the expensive computation of pairwise Gromov–Wasserstein distances in certain applications.**

***Index Terms*—Optimal transport, linear Wasserstein distance, Wasserstein spaces, Gromov–Wasserstein distance, shape spaces.**

I. Introduction

RECENTLY, a simplified version of optimal transport in Wasserstein spaces, called linear optimal transport (LOT), was introduced by Wang et al. [1]. Indeed all theoretical justifications can be found in the book of Ambrosio, Gigli and Savaré [2]. From a geometric point of view, this approach just transfers measures from the geodesic Wasserstein space by the inverse exponential map to the tangent space at some fixed reference measure which is assumed to be absolutely continuous (with respect to the Lebesgue measure). Then the LOT distance can be characterized by the optimal transport maps between the reference measure and the considered measures. This approach allows to work in the linear tangent space instead of the Wasserstein space so that subsequent computations can utilize known methods from data science as for instance classification techniques. This is especially suited for the approximate computation of pairwise distances for large databases of images and signals. Meanwhile LOT has been successfully applied for several tasks in nuclear structure-based pathology [3], parametric signal estimation [4], signal and image classification [5], [6], modeling of turbulences [7], cancer detection [8]–[10], Alzheimer disease detection [11], vehicle-type recognition [12] as well as for de-multiplexing vortex modes in optical communications [13]. On the real line, LOT can further be written using the cumulative density function of the random variables associated to the involved measures. This was used in combination with the Radon transform under the name Radon-CDT [6], [14]. We like to mention that (inverse) exponential mappings were also used for the iterative computation of Fréchet means, also known as barycenters, in Wasserstein spaces in [15]. Furthermore, in [16], the determination of conditions which allow the transformation of signals created by algebraic generative models into convex sets by applying LOT has been addressed. In [17], the authors characterized settings in which LOT embeds families of distributions into a space in which they are linearly separable and provided conditions such that the LOT distance between two measures is nearly isometric to the Wasserstein distance. Finally, note that a linear version of the Hellinger–Kantorovich distance is also available [18].

However, when dealing with reference measures which are not absolutely continuous, e.g., discrete measures, then optimal transport maps are in general not available such that a generalized setting of LOT is needed. In this paper, we propose a generalized LOT which relies on barycentric averaging maps of optimal transport plans instead of optimal transport maps. For discrete measures, such an approach was also considered in [1]. In this paper, we actually need this generalized LOT concept to motivate our framework of (generalized) linear Gromov–Wasserstein distances.

Gromov–Wasserstein distances were first considered by Mémoli in [19] as a modification of Gromov–Hausdorff and Wasserstein distances. A survey of the geometry of Gromov–Wasserstein spaces was given by Sturm in [20]. Due to its invariance on isomorphism classes of so-called metric measure spaces, the Gromov–Wasserstein distance is more suited for certain practical computations like shape comparison and matching while retaining several desirable theoretical properties of its predecessors. A combination with inverse problems has been considered in [21]. Further, a sliced version of the Gromov–Wasserstein distance has been proposed in [22]. For the one-dimensional discrete setting, uniform distributions and the squared distance $|\cdot|^2$ instead of $|\cdot|$ within the definition of the Gromov-Wasserstein distance, the authors gave a reformulation of the distance such that its minimization requires essentially a sorting of sets. Recently, Gromov–Wasserstein distances were examined for Gaussian measures in [23].

In this paper, we introduce a linear variant of the Gromov–Wasserstein distance that has the same advantages as LOT, namely the efficient computation of pairwise distances in larger data sets, which can be subsequently coupled with standard methods from image and signal processing. Since the Brenier theorem that relates optimal transport maps with transport plans in Wasserstein spaces is not available for the Gromov–Wasserstein setting, we rely on optimal transport plans with respect to Gromov–Wasserstein distances which always exist. Numerical examples illustrate how our linear Gromov–Wasserstein distances perform in certain applications.

Outline of the paper: In Section II, we deal with linear optimal transport and its generalization via barycentric projection maps of transport plans. In Section III, we consider Gromov–

F. Beier is with the Institute of Mathematics, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany.

R. Beinert is with the Institute of Mathematics, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany.

G. Steidl is with the Institute of Mathematics, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany.

Wasserstein distances. We introduce the basic notation and properties which are quite technical and we tried to keep things as simple as possible. Then, following the definition of generalized LOT, we propose generalized linear Gromov–Wasserstein distances. Section IV demonstrates how linear Gromov–Wasserstein distances perform in several applications. Finally, conclusions are drawn in Section V.

II. Linear Optimal Transport

In this section, we introduce a general version of LOT. This will provide the idea how to get a linear Gromov–Wasserstein distance.

A. Optimal Transport

By $L^2_\mu(\mathbb{R}^d, \mathbb{R}^d)$ we denote the space of (equivalence classes of) functions $T\colon \mathbb{R}^d \to \mathbb{R}^d$ fulfilling

$$\|T\|_{L^2_\mu} \coloneqq \Big(\int_{\mathbb{R}^d} \|T(x)\|^2 \,\mathrm{d}\mu(x) \Big)^{\frac{1}{2}} < \infty.$$

Let $\mathcal{P}(\mathbb{R}^d)$ be the space of probability measures on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ and $\mathcal{P}_2(\mathbb{R}^d)$ be the space of measures with finite second moments. The *push forward measure* $T_\#\mu$ of $\mu \in \mathcal{P}(\mathbb{R}^d)$ by a measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ is defined by $T_\#\mu(B) \coloneqq \mu(T^{-1}(B))$ for all $B \in \mathcal{B}(\mathbb{R}^d)$. By $\|\cdot\|$ we denote the Euclidean norm on $\mathbb{R}^d$. Together with the *Wasserstein distance*

$$W(\mu,\nu) \coloneqq \min_{\pi \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^2 \,\mathrm{d}\pi(x,y) \right)^{\frac{1}{2}}, \quad (1)$$

where $\Pi(\mu,\nu)$ denotes the set of transport plans $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ and ν, the space $\mathcal{P}_2(\mathbb{R}^d)$ becomes a metric space, known as (2-)*Wasserstein space*. For the more general definition of p-Wasserstein spaces, $p \in [1,\infty)$, see, e.g., [24]. The Wasserstein space is a *geodesic space*, meaning, that for every $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a continuous curve $\gamma : [0,1] \to \mathcal{P}_2(\mathbb{R}^d)$ with $\gamma(0) = \mu$, $\gamma(1) = \nu$ and

$$W(\gamma(t),\gamma(s)) = |t-s| W(\gamma(0),\gamma(1)) \quad (2)$$

for all $t,s \in [0,1]$. A continuous curve with property (2) is called (constant speed) *geodesic*.

If the measure μ is absolutely continuous, then, by the following theorem of Brenier [25], optimal transport plans in (1) are unique and can be characterized by transport maps.

Theorem II.1 (Brenier's Theorem)**.** *Let $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, where μ is absolutely continuous. Then the minimization problem in* (1) *admits a unique solution π^ν_μ. Moreover, there exists a unique optimal transport map $T^\nu_\mu \in L^2_\mu(\mathbb{R}^d,\mathbb{R}^d)$ which solves*

$$\min_T \int_{\mathbb{R}^d} \|x - T(x)\|^2 \,\mathrm{d}\mu(x) \quad \text{subject to} \quad T_\#\mu = \nu.$$

This optimal map is related to the optimal transport plan by

$$\pi^\nu_\mu = (\mathrm{id}, T^\nu_\mu)_\#\mu.$$

The situation changes if μ is not absolutely continuous. Then there still exists an optimal transport plan, but it may not be unique. In contrast, the existence of an optimal transport map is not guaranteed. However, if there exists T such that $\nu = T_\#\mu$ and $\pi \coloneqq (\mathrm{id}, T)_\#\mu$ is an optimal plan, then T is an optimal map. Conversely, if T is an optimal map, then $\pi \coloneqq (\mathrm{id}, T)_\#\mu$ fulfills the marginal conditions, but must not be an optimal plan, as the example $\mu \coloneqq \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1$ and $\nu \coloneqq \frac{3}{4}\delta_0 + \frac{1}{4}\delta_1$ shows. In the following, we denote by $\Pi_\mathrm{o}(\sigma,\mu)$ the set of optimal transport plans in (1).

B. Linear Optimal Transport

For discrete measures with a maximum of n support points, the optimal transport amounts to solving a linear program that has worst-case complexity of n^3. Computing the pairwise Wasserstein distances of N such measures results in $\binom{N}{2}$ optimal transport computations, which becomes numerically intractable for large N. To speed up the numerical comparison, Wang et al. [1] proposed LOT, which exploits the geometric structure of the Wasserstein space. Following [2, Eq (8.5.1)], the *reduced tangent space (cone)* $\mathrm{Tan}^\mathrm{r}_\sigma \mathcal{P}_2(\mathbb{R}^d) \subset L^2_\mu(\mathbb{R}^d,\mathbb{R}^d)$ with base $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ is given by

$$\begin{aligned}&\mathrm{Tan}^\mathrm{r}_\sigma \mathcal{P}_2(\mathbb{R}^d)\\ &\coloneqq \overline{\big\{ r(T-\mathrm{id}) : (\mathrm{id}\times T)_\#\sigma \in \Pi_\mathrm{o}(\sigma, T_\#\sigma), r>0 \big\}}^{L^2_\sigma}.\end{aligned}$$

Note that the mapping T in this definition is always an optimal transport map between σ and $T_\#\sigma$. If σ is absolutely continuous, then the mapping

$$F_\sigma : \mathcal{P}_2(\mathbb{R}^d) \to \mathrm{Tan}^\mathrm{r}_\sigma \mathcal{P}_2(\mathbb{R}^d), \quad \mu \mapsto T^\mu_\sigma - \mathrm{id}$$

is the inverse exponential map. The key idea of LOT is to approximate $W(\mu,\nu)$ by the distance of the liftings to the tangent space, i.e.

$$\mathrm{LOT}_\sigma(\mu,\nu) \coloneqq \|F_\sigma(\mu) - F_\sigma(\nu)\|_{L^2_\sigma} = \|T^\mu_\sigma - T^\nu_\sigma\|_{L^2_\sigma}. \quad (3)$$

Then LOT is length preserving, i.e. $W(\mu,\sigma) = \mathrm{LOT}_\sigma(\mu,\sigma)$ and gives an upper bound of the Wasserstein distance

$$W(\mu,\nu) \le \mathrm{LOT}_\sigma(\mu,\nu).$$

For a fixed $\sigma \in \mathcal{P}_2(\mathbb{R})$, the computation of all pairwise LOT_σ distances by (3) between N measures requires only N transport map computations.

One shortcoming of LOT_σ in (3) is that the base measure σ has to be absolutely continuous to ensure that the inverse exponential map to the reduced tangential space is well-defined for all measures in $\mathcal{P}_2(\mathbb{R}^d)$. As a remedy, we replace the reduced tangent space by the geometric tangent space. Given $\pi^\mu_\sigma \in \Pi_\mathrm{o}(\sigma,\mu)$, the mapping

$$t \mapsto \pi^{\sigma\to\mu}_t \coloneqq ((1-t)P^1 + tP^2)_\#\pi^\mu_\sigma, \quad t \in [0,1],$$

with the projections $P^1(s,x) \coloneqq s$ and $P^2(s,x) \coloneqq x$ defines a geodesic between σ and μ. Moreover, every geodesic corresponds one-to-one to an optimal plan [2, Thm 7.2.2]. Henceforth we identify each geodesic by its plan. Let G_σ denote the set of equivalence classes of all geodesics starting in σ, where two geodesics are equivalent, if there exists an $\epsilon > 0$ such that they coincide for $t \in [0,\epsilon]$. The *geometric*

tangent space $\operatorname{Tan}^{\mathrm{g}}_\sigma \mathcal{P}_2(\mathbb{R}^d)$ is the closure of G_σ with respect to the metric

$$W^2_\sigma(\pi^\mu_\sigma, \pi^\nu_\sigma) := \min_{\pi \in \Pi_\sigma(\pi^\mu_\sigma, \pi^\nu_\sigma)} \int_{\mathbb{R}^{3d}} |x-y|^2 \,\mathrm{d}\pi(s,x,y), \quad (4)$$

where $\Pi_\sigma(\pi^\mu_\sigma, \pi^\nu_\sigma)$ consists of all 3-plans $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ with $P^{12}_\# \pi = \pi^\mu_\sigma$ and $P^{13}_\# \pi = \pi^\nu_\sigma$, and where $P^{12}(s,x,y) := (s,x)$ and $P^{13}(s,x,y) := (s,y)$, cf. [2, § 12.4]. Note that the plans $\pi \in \Pi_\sigma(\pi^\mu_\sigma, \pi^\nu_\sigma)$ give also rise to so-called generalized geodesics between μ and ν, c.f. [2, § 9.2].

If σ is not absolutely continuous, there may exist more than one geodesic between σ and μ, ν, i.e. $\Pi_{\mathrm{o}}(\sigma,\mu)$ and $\Pi_{\mathrm{o}}(\sigma,\nu)$ are no singletons; so a proper extension of LOT to not absolutely continuous bases is

$$\operatorname{LOT}_\sigma(\mu,\nu) := \inf_{\substack{\pi^\mu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\mu) \\ \pi^\nu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\nu)}} W_\sigma(\pi^\mu_\sigma, \pi^\nu_\sigma). \quad (5)$$

A simple computation shows that LOT in (3) and (5) coincides for absolutely continuous σ. In general $\operatorname{LOT}_\sigma$ is only a semi-metric, i.e., the triangular inequality is not fulfilled. Taking the supremum instead of the infimum in (5) would fix this issue. Moreover, we have again $W(\mu,\nu) \le \operatorname{LOT}_\sigma(\mu,\nu)$.

Remark II.2. *Besides the geometric interpretation, we may interpret* $\operatorname{LOT}_\sigma$ *also as constrained optimization of the Wasserstein distance* (1)*, where only those plans* $\pi \in \Pi(\mu,\nu)$ *are considered that transport the mass optimally over the reference measure* σ*. More precisely, if we are given two plans* $\pi^\mu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\mu)$ *and* $\pi^\nu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\nu)$*, then the gluing lemma of Dudley [26, Lem. 8.4] ensures the existence of* $\pi_{\mathrm{g}} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ *such that* $P^{12}_\# \pi_{\mathrm{g}} = \pi^\mu_\sigma$ *and* $P^{13}_\# \pi_{\mathrm{g}} = \pi^\nu_\sigma$*. The two plans* π^μ_σ *and* π^ν_σ *are glued together along the first axis. If the two marginal plans are related to maps, i.e.* $\pi^\mu_\sigma = (\mathrm{id}, T^\mu_\sigma)_\# \sigma$ *and* $\pi^\nu_\sigma = (\mathrm{id}, T^\nu_\sigma)_\# \sigma$*, then the gluing is unique and given by* $\pi_{\mathrm{g}} := (\mathrm{id}, T^\mu_\sigma, T^\nu_\sigma)_\# \sigma$.

C. Generalized Linear Optimal Transport

Although $\operatorname{LOT}_\sigma$ in (5) is also well defined for point reference measures, the numerical implementation requires the computation of an optimal 3-plan, which completely counteracts the intention behind LOT. Instead we remain in the comfortable setting of transport maps by using barycentric projection maps, which are based on the disintegration of transport plans [2, Thm 5.3.1]. More precisely, given $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with $P^1_\# \pi = \sigma$, there exists a σ-almost everywhere uniquely defined family of measures $(\pi_s)_{s \in \mathbb{R}^d} \subset \mathcal{P}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^{2d}} f(s,x) \,\mathrm{d}\pi(s,x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(s,x) \,\mathrm{d}\pi_s(x) \,\mathrm{d}\sigma(s).$$

The *barycentric projection map* $\mathcal{T}_\pi : \mathbb{R}^d \to \mathbb{R}^d$ of $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with first marginal σ is defined for σ-almost every $s \in \mathbb{R}^d$ by

$$\mathcal{T}_\pi(s) := \int_{\mathbb{R}^d} x \,\mathrm{d}\pi_s(x) = \operatorname*{argmin}_{x' \in \mathbb{R}^d} \int_{\mathbb{R}^d} \|x - x'\|^2 \,\mathrm{d}\pi_s(x) \quad (6)$$

provided that π_s has finite second moments σ-a.e., see [2, p. 126].

Example II.3. *For a discrete probability measure*

$$\sigma = \sum_{i=1}^n \sigma_i \delta_{s_i} \in \mathcal{P}(\mathbb{R}^d)$$

and

$$\pi = \sum_{i=1}^n \sum_{j=1}^m \pi_{i,j} \delta_{s_i, x_j} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$$

with first marginal σ*, the barycentric projection reads as*

$$\mathcal{T}_\pi(s_i) = \frac{1}{\sigma_i} \sum_{j=1}^m \pi_{i,j} x_j, \quad i = 1, \dots, n.$$

Such maps are also used in [1].

By the following proposition, the barycentric projection map (6) of an optimal transport plan π^μ_σ is always an optimal transport map $T^{\tilde\mu}_\sigma$ between σ and $\tilde\mu = (\mathcal{T}_{\pi^\mu_\sigma})_\# \sigma$.

Proposition II.4. *For each* $\pi^\mu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\mu)$*, the barycentric projection map* $\mathcal{T}_{\pi^\mu_\sigma} : \mathbb{R}^d \to \mathbb{R}^d$ *in* (6) *defines an optimal transport map from* σ *to the measure* $\tilde\mu := (\mathcal{T}_{\pi^\mu_\sigma})_\# \sigma$*, i.e.,*

$$\mathcal{T}_{\pi^\mu_\sigma} = T^{\tilde\mu}_\sigma.$$

Although the statement may be implicitly derived from [2, § 12.4], we give a direct proof in the appendix. On the basis of the barycentric projection, we propose to extend the LOT formulation in (3) by considering *generalized LOT (gLOT)*

$$\operatorname{gLOT}_\sigma(\mu,\nu) := \inf_{\substack{\pi^\mu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\mu) \\ \pi^\nu_\sigma \in \Pi_{\mathrm{o}}(\sigma,\mu)}} \|\mathcal{T}_{\pi^\mu_\sigma} - \mathcal{T}_{\pi^\nu_\sigma}\|_{L^2_\sigma}.$$

If $\pi^\mu_\sigma = (\mathrm{id}, T^\mu_\sigma)_\# \sigma$, then $\mathcal{T}_{\pi^\mu_\sigma} = T^\mu_\sigma$, so that gLOT coincides with LOT in particular for absolutely continuous bases. In the numeric implementation of gLOT, we stick with fixed optimal plans π^μ_σ and π^ν_σ instead of minimizing over all optimal plans.

Remark II.5. *gLOT has actually a geometric interpretation. The barycentric projection* $\pi \to \mathcal{T}_\pi$ *defines a map from the geometric tangent space to the reduced tangent space by*

$$\pi^\mu_\sigma \in \operatorname{Tan}^{\mathrm{g}}_\sigma \mathcal{P}_2(\mathbb{R}^d) \quad \mapsto \quad (\mathcal{T}_{\pi^\mu_\sigma} - \mathrm{id}) \in \operatorname{Tan}^{\mathrm{r}}_\sigma \mathcal{P}_2(\mathbb{R}^d),$$

see Proposition II.4. From this point of view, gLOT takes two geodesics corresponding to the optimal plans π^μ_σ *and* π^ν_σ *from the geometric tangent space, maps them to the reduced tangent space, and computes the distance there. In this way, we overcome the issue that the inverse exponential map may not be defined for the whole* $\mathcal{P}_2(\mathbb{R}^d)$.

III. Linear Gromov–Wasserstein Distance

For certain applications like shape matching, a weakness of the optimal transport distance is its lack of important invariances such as translations and rotations. Mémoli [19] introduced an optimal transport like distance, where the aim was to match measures according to pairwise distance perturbations. To this end, we need the definition of a *metric measure space (mm-space)* which is a triple $\mathbb{X} = (X, d_X, \mu)$, where

i) (X, d_X) is a compact metric space,
ii) μ is a Borel probability measure on X with full support.

A. Gromov–Wasserstein Distance

For two mm-spaces $\mathbb{X} = (X, d_X, \mu)$ and $\mathbb{Y} = (Y, d_Y, \nu)$, the *Gromov–Wasserstein (GW) distance* is defined by

$$\mathrm{GW}(\mathbb{X}, \mathbb{Y}) \coloneqq \inf_{\pi \in \Pi(\mu,\nu)} \Bigg(\int_{(X\times Y)^2} |d_X(x,x') - d_Y(y,y')|^2 \times \mathrm{d}\pi(x,y)\,\mathrm{d}\pi(x',y') \Bigg)^{\frac{1}{2}}. \quad (7)$$

Here $\pi \in \Pi(\mu,\nu)$ means that $\pi \in \mathcal{P}(X \times Y)$ has marginals μ and ν. Further, we denote by $\Pi_{\mathrm{o}}(\mathbb{X}, \mathbb{Y})$ the set of optimal GW plans in (7). In the literature, the above quantity is also called 2-Gromov–Wasserstein distance, and analogous definitions for $p \in [1, \infty)$ as well as further generalizations are possible. For an overview, we refer also to [27]. Due to the Weierstraß theorem, a minimizer in (7) always exists [19, Cor 10.1]. Two mm-spaces $\mathbb{X} = (X, d_X, \mu)$ and $\mathbb{Y} = (Y, d_Y, \nu)$ are called *isomorphic* if and only if there exists a (bijective) isometry $\psi : X \to Y$ such that $\psi_{\#}\mu = \nu$. We denote the corresponding equivalence classes by $[\![\cdot]\!]$. The GW distance defines a metric on these equivalence classes [19, Thm 5.1]. The resulting (incomplete) metric space is called the *Gromov–Wasserstein space*. In particular, the GW distance is invariant under translation and rotation of the mm-space.

Up to now, there does not exist a general analogue to Brenier's Theorem for the GW setting, see also [21, Rem 3.3]. In [20], Sturm has shown that in the Euclidean setting optimal GW plans between rotationally invariant probability spaces are realized by optimal transport maps.

Due to its invariance properties and independence of the ambient spaces, the GW metric provides a valuable tool for data science, shape analysis and object classification. However, its exact computation is NP-hard. Even its approximation is computationally challenging and requires, if a gradient descent algorithm is used, $O(n^3 \log(n))$ arithmetic operations, where n is the cardinality of the underlying mm-spaces [28]. For further improvements in the setting of sparse graphs, see [29]. Hence its use for comparing a larger number of mm-spaces is limited which motivates the following considerations.

B. Linear Gromov–Wasserstein Distance

We consider the (equivalence classes of) mm-spaces $\mathbb{S} = [\![S, d_S, \sigma]\!]$, $\mathbb{X} = [\![X, d_X, \mu]\!]$, and $\mathbb{Y} = [\![Y, d_Y, \nu]\!]$. In contrast to the above definition from Mémoli [19], we allow that the measures σ, μ, and ν may not have full support. Similarly to the Wasserstein setting, the Gromov–Wasserstein space is geodesic. Unfortunately, the construction of the tangent space is a little technical. We follows the lines of Sturm in [20]. Each geodesic from $\mathbb{S}$ to $\mathbb{X}$ has the form

$$t \mapsto \pi_t^{\mathbb{S}\to\mathbb{X}} \coloneqq [\![S \times X, (1-t)d_S + t d_X, \pi]\!], \quad t \in [0,1], \quad (8)$$

where $\pi \in \Pi_{\mathrm{o}}(\mathbb{S}, \mathbb{X})$, and where d_S acts on the S components and d_X on the X components of $(S \times X)^2$, respectively. Conversely, every optimal plan defines a geodesic. Note that $\pi_0^{\mathbb{S}\to\mathbb{X}}$ and $\pi_1^{\mathbb{S}\to\mathbb{X}}$ are isomorphic to $\mathbb{S}$ and $\mathbb{X}$ by $P^1(s,x) \coloneqq s$ and $P^2(s,x) \coloneqq x$, respectively.

In order to introduce tangent spaces and to derive their explicit representations, the GW space is embedded into the more regular space of gauged measure spaces. A *gauged measure space (gm-space)* is as before a triple $\mathbb{S}_{\mathrm{s}} \coloneqq (S, k_S, \sigma)$, where the distance is replaced by a so-called *gauge function* k_S in $L^2_{\mathrm{s}}(S\times S, \sigma\otimes\sigma)$, where the subscript „s" signifies „symmetric" and the latter space consists of all *symmetric*, square-integrable functions with respect to $\sigma \otimes \sigma$. Here, S can be a Polish space. The extension of the GW distance to the gm-spaces $\mathbb{X}_{\mathrm{s}} = (X, k_X, \mu)$ and $\mathbb{Y}_{\mathrm{s}} = (Y, k_Y, \nu)$ is given by

$$\mathrm{GW}(\mathbb{X}_{\mathrm{s}}, \mathbb{Y}_{\mathrm{s}}) = \inf_{\pi \in \Pi(\mu,\nu)} \Bigg(\int_{(X\times Y)^2} |k_X(x,x') - k_Y(y,y')|^2 \times \mathrm{d}\pi(x,y)\,\mathrm{d}\pi(x',y') \Bigg)^{\frac{1}{2}}. \quad (9)$$

A minimizing coupling always exists [20, Thm 5.8]. The set of all plans minimizing the integral in (9) with respect to $\mathbb{X}_{\mathrm{s}}$ and $\mathbb{Y}_{\mathrm{s}}$ is henceforth denoted by $\Pi_{\mathrm{o}}(\mathbb{X}_{\mathrm{s}}, \mathbb{Y}_{\mathrm{s}})$. Clearly, every mm-space is a gm-space. Two gauged measure spaces $\mathbb{X}_{\mathrm{s}}$ and $\mathbb{Y}_{\mathrm{s}}$ are called *homomorphic* if $\mathrm{GW}(\mathbb{X}_{\mathrm{s}}, \mathbb{Y}_{\mathrm{s}}) = 0$. The space $\mathfrak{G}$ of homomorphic equivalent classes—again denoted by $[\![\cdot]\!]$—equipped with the GW distance (9) is complete and geodesic. To simplify notation, we denote such equivalence classes again by $\mathbb{X}_{\mathrm{s}}$. The geodesics from $\mathbb{S}_{\mathrm{s}}$ to $\mathbb{X}_{\mathrm{s}}$ have the form

$$t \mapsto \pi_t^{\mathbb{S}_{\mathrm{s}}\to\mathbb{X}_{\mathrm{s}}} \coloneqq [\![S\times X, (1-t)k_S + t k_X, \pi]\!], \quad t \in [0,1], \quad (10)$$

where $\pi \in \Pi_{\mathrm{o}}(\mathbb{S}_{\mathrm{s}}, \mathbb{X}_{\mathrm{s}})$. Conversely, each $\pi \in \Pi_{\mathrm{o}}(\mathbb{S}_{\mathrm{s}}, \mathbb{X}_{\mathrm{s}})$ defines a geodesic.

Formally, the *tangent space* $\mathrm{Tan}_{\mathbb{S}_{\mathrm{s}}}\mathfrak{G}$ with base $\mathbb{S}_{\mathrm{s}} \in \mathfrak{G}$ is defined as

$$\mathrm{Tan}_{\mathbb{S}_{\mathrm{s}}}\mathfrak{G} \coloneqq \Big(\bigcup_{[\![S,k_S,\sigma]\!]=\mathbb{S}_{\mathrm{s}}} L^2_{\mathrm{s}}(S\times S, \sigma\otimes\sigma) \Big) \Big/ \sim,$$

where the union is taken over all gm-spaces (S, k_S, σ) in the equivalence class $\mathbb{S}_{\mathrm{s}}$ and two functions $g \in L^2_{\mathrm{s}}(S\times S, \sigma\otimes\sigma)$ and $g' \in L^2_{\mathrm{s}}(S'\times S', \sigma'\otimes\sigma')$ defined on the representatives (S, k_S, σ) and (S', k'_S, σ') of $\mathbb{S}_{\mathrm{s}}$ are equivalent, if there exists $\pi \in \Pi_{\mathrm{o}}\left((S, k_S, \sigma), (S', k_{S'}, \sigma')\right)$ such that

$$g(s_1, s_2) = g'(s'_1, s'_2)$$

almost everywhere with respect to $\pi(s_1, s'_1) \otimes \pi(s_2, s'_2)$. Note that each tangent $g \in \mathrm{Tan}_{\mathbb{S}_{\mathrm{s}}}\mathfrak{G}$ is implicitly associated with its representative (S, k_S, σ). A (cone) *metric on* $\mathrm{Tan}_{\mathbb{S}_{\mathrm{s}}}\mathfrak{G}$ is given by

$$\begin{aligned}&\mathrm{GW}_{\mathbb{S}_{\mathrm{s}}}(g,h)\\ &\coloneqq \inf\big\{\|g-h\|_{L^2((S\times S')^2, \pi\otimes\pi)} : \pi \in \Pi_{\mathrm{o}}(\mathbb{T}_g, \mathbb{T}_h)\big\},\end{aligned} \quad (11)$$

where $\mathbb{T}_g$ and $\mathbb{T}_h$ denote the representatives associated with g and h. Given $g \in \mathrm{Tan}_{\mathbb{S}_{\mathrm{s}}}\mathfrak{G}$ defined on the representative (S, k_S, σ) of the equivalence class $\mathbb{S}_{\mathrm{s}}$, the *exponential map* $E_{\mathbb{S}_{\mathrm{s}}} : \mathrm{Tan}_{\mathbb{S}_{\mathrm{s}}}\mathfrak{G} \to \mathfrak{G}$ is defined by

$$E_{\mathbb{S}_{\mathrm{s}}}(g) = [\![S, k_S + g, \sigma]\!].$$

As a consequence, every geodesic in (10) may be written as

$$\pi_t^{\mathbb{S}_{\mathrm{s}}\to\mathbb{X}_{\mathrm{s}}} = E_{\mathbb{S}_{\mathrm{s}}}(th) \quad \text{with} \quad h \coloneqq k_X - k_S,$$

where h is defined on the representative $(S\times X, k_S, \pi)$ with $\pi \in \Pi_{\mathrm{o}}(\mathbb{S}_{\mathrm{s}}, \mathbb{X}_{\mathrm{s}})$. Note that two geodesics which coincide for all $t \in [0, \epsilon]$ for some $\epsilon > 0$ correspond to the same tangent; so the

tangent space embrace all geodesics starting in $\mathbb{S}_s$. Associating any geodesic $\pi_t^{\mathbb{S}_s \to \mathbb{X}_s}$ with its optimal plan $\pi_{\mathbb{S}_s}^{\mathbb{X}_s} \in \Pi_o(\mathbb{S}_s, \mathbb{X}_s)$, we define $F_{\mathbb{S}_s} : \mathfrak{G} \to \mathrm{Tan}_{\mathbb{S}_s} \mathfrak{G}$ by

$$F_{\mathbb{S}_s}(\pi_{\mathbb{S}_s}^{\mathbb{X}_s}) = k_X - k_S \quad (\text{acting on } (S \times X, k_S, \pi_{\mathbb{S}_s}^{\mathbb{X}_s})).$$

For the geodesics (8) between mm-spaces, we especially have

$$F_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}) = d_X - d_S \quad (\text{acting on } (S \times X, d_S, \pi_{\mathbb{S}}^{\mathbb{X}})).$$

Against this background, we define the distance between two geodesics $\pi_{\mathbb{S}}^{\mathbb{X}}$ and $\pi_{\mathbb{S}}^{\mathbb{Y}}$ as

$$\mathrm{GW}_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}, \pi_{\mathbb{S}}^{\mathbb{Y}}) \coloneqq \mathrm{GW}_{\mathbb{S}}(F_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}), F_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{Y}})). \tag{12}$$

Then we have the following relation which proof is given in the appendix.

Proposition III.1. *Consider the mm-spaces $\mathbb{S} = [\![S, d_S, \sigma]\!]$, $\mathbb{X} = [\![X, d_X, \mu]\!]$, $\mathbb{Y} = [\![Y, d_Y, \nu]\!]$. The distance* (12) *between the geodesics related to $\pi_{\mathbb{S}}^{\mathbb{X}} \in \Pi_o(\mathbb{S}, \mathbb{X})$ and $\pi_{\mathbb{S}}^{\mathbb{Y}} \in \Pi_o(\mathbb{S}, \mathbb{Y})$ is given by*

$$\mathrm{GW}_{\mathbb{S}}^2(\pi_{\mathbb{S}}^{\mathbb{X}}, \pi_{\mathbb{S}}^{\mathbb{Y}}) = \inf_{\pi \in \Pi_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}, \pi_{\mathbb{S}}^{\mathbb{Y}})} \int_{(S \times X \times Y)^2} |d_X(x, x') - d_Y(y, y')|^2 \times \mathrm{d}\pi(s, x, y)\, \mathrm{d}\pi(s', x', y'),$$

where $\Pi_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}, \pi_{\mathbb{S}}^{\mathbb{Y}})$ consists of all 3-plans $\pi \in \mathcal{P}(S \times X \times Y)$ with $P_{\#}^{12}\pi = \pi_{\mathbb{S}}^{\mathbb{X}}$ and $P_{\#}^{13}\pi = \pi_{\mathbb{S}}^{\mathbb{Y}}$.

The minimization over the 3-plans in the proposition goes hand in hand with 3-plans in the definition of W_σ in (4).

In the spirit of LOT, we now propose to approximate the GW distance $\mathrm{GW}(\mathbb{X}, \mathbb{Y})$ by lifting $\mathbb{X}$ and $\mathbb{Y}$ to $\mathrm{Tan}_{\mathbb{S}} \mathfrak{G}$ via geodesics and using the metric on the tangent space. More precisely, we define the *linear Gromov–Wasserstein distance* by

$$\mathrm{LGW}_{\mathbb{S}}(\mathbb{X}, \mathbb{Y}) \coloneqq \inf_{\substack{\pi_{\mathbb{S}}^{\mathbb{X}} \in \Pi_o(\mathbb{S}, \mathbb{X}) \\ \pi_{\mathbb{S}}^{\mathbb{Y}} \in \Pi_o(\mathbb{S}, \mathbb{Y})}} \mathrm{GW}_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}, \pi_{\mathbb{S}}^{\mathbb{Y}}).$$

A simple calculation shows that $\mathrm{GW}(\mathbb{X}, \mathbb{Y}) \leq \mathrm{LGW}_{\mathbb{S}}(\mathbb{X}, \mathbb{Y})$. In comparison with (5), we can consider LGW as an analogue to LOT in GW spaces.

C. Generalized Linear Gromov–Wasserstein Distance

Assume for the moment that we are in the comfortable position, where $\pi_{\mathbb{S}}^{\mathbb{X}} \in \Pi_o(\mathbb{S}, \mathbb{X})$ and $\pi_{\mathbb{S}}^{\mathbb{Y}} \in \Pi_o(\mathbb{S}, \mathbb{Y})$ are unique and induced by optimal maps $T_{\mathbb{S}}^{\mathbb{X}}$ and $T_{\mathbb{S}}^{\mathbb{Y}}$. In this situation, $\Pi_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}, \pi_{\mathbb{S}}^{\mathbb{Y}})$ becomes the singleton $(\mathrm{id}, T_{\mathbb{S}}^{\mathbb{X}}, T_{\mathbb{S}}^{\mathbb{Y}})_{\#}\sigma$, and we obtain

$$\mathrm{GW}_{\mathbb{S}}(\mathbb{X}, \mathbb{Y}) = \big\| d_X(T_{\mathbb{S}}^{\mathbb{X}}(\cdot_1), T_{\mathbb{S}}^{\mathbb{X}}(\cdot_2)) - d_Y(T_{\mathbb{S}}^{\mathbb{Y}}(\cdot_1), T_{\mathbb{S}}^{\mathbb{Y}}(\cdot_2)) \big\|_{L^2_{\sigma \otimes \sigma}},$$

where $\cdot_1$ and $\cdot_2$ are the first and second argument with respect to $S \times S$.

Similarly to LOT, LGW does not alleviate the computational costs of calculating pairwise GW distances. For this reason, we recommend to use the barycentric projection mapping to transform $\pi_{\mathbb{S}}^{\mathbb{X}}$ and $\pi_{\mathbb{S}}^{\mathbb{Y}}$ into maps $\mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{X}}}$ and $\mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{Y}}}$. Since the metric spaces may be more general than the measure spaces considered in Section II, we introduce the *generalized barycentric projection*

$$\mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{X}}}(s) \coloneqq \operatorname*{argmin}_{x' \in X} \int_X d_X^2(x', x)\, \mathrm{d}\pi_{\mathbb{S},s}^{\mathbb{X}}(x), \tag{13}$$

where $\pi_{\mathbb{S},s}^{\mathbb{X}}$ is the disintegration of the chosen $\pi_{\mathbb{S}}^{\mathbb{X}}$. Based on Weierstraß' theorem, we see that the minimum is attained. In the special case that $X \subset \mathbb{R}^d$ is convex and $d_X(x_1, x_2) = \|x_1 - x_2\|$, the generalized barycentric projection coincides with (6).

Analogously to gLOT, we define *generalized LGW (gLGW)* by

$$\mathrm{gLGW}_{\mathbb{S}}(\mathbb{X}, \mathbb{Y}) \coloneqq \inf_{\substack{\pi_{\mathbb{S}}^{\mathbb{X}} \in \Pi_o(\mathbb{S}, \mathbb{X}) \\ \pi_{\mathbb{S}}^{\mathbb{Y}} \in \Pi_o(\mathbb{S}, \mathbb{Y})}} \big\| d_X(\mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{X}}}(\cdot_1), \mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{X}}}(\cdot_2)) - d_Y(\mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{Y}}}(\cdot_1), \mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{Y}}}(\cdot_2)) \big\|_{L^2_{\sigma \otimes \sigma}}.$$

For numerical computations, we again propose to use fixed optimal plans instead of minimizing over $\Pi_o(\mathbb{S}, \mathbb{X})$ and $\Pi_o(\mathbb{S}, \mathbb{Y})$.

Remark III.2. *Recall that under the conditions of the Brenier theorem, the OT and LOT distances coincide in the one-dimensional setting. The linear GW distance differs in general from the GW distance also in one dimension. We verified this by computing the corresponding GW and LGW distances for*

$$S = \{0, 1, 2, 3, 6\}, \quad X = \{0, 1, 2, 5, 7\}, \quad Y = \{0, 2, 3, 6, 7\}$$

the absolute value distances and the corresponding discrete measures with weights $\frac{1}{5}$. For this specific instance, we obtain $\mathrm{GW}(\mathbb{X}, \mathbb{Y}) \approx 0.69$ and $\mathrm{gLGW}(\mathbb{X}, \mathbb{Y}) \approx 1.13$. Note that the same example delivers also different values in the setting of squared absolute values considered in [22]. Here it can be easily shown that S must have cardinality 5 at least to construct a counterexample.

IV. Numerical Examples

All numerical experiments in this section have been implemented in *Python 3*, where we mainly rely on the packages *Python Optimal Transport (POT)* [30], *scikit-learn* [31], and *NetworkX* [32]. POT contains a Gromov–Wasserstein module allowing the numerical computation of the GW distance (7), a corresponding optimal plan, and GW barycenters for discrete mm-spaces, where the measure space consists of finitely many points, and where the measure thus becomes a point measure. The used methods are based on the work of Peyré, Cuturi & Solomon [28] and Mémoli [19]. The GW barycenter $\mathbb{S}$ between the discrete mm-spaces $\mathbb{X}_k$ for $k = 1, \ldots, K$ is defined via

$$\mathbb{S} = \operatorname*{argmin}_{\tilde{\mathbb{S}}} \sum_{k=1}^{K} \mathrm{GW}(\tilde{\mathbb{S}}, \mathbb{X}_k).$$

The minimization here goes over the set of all discrete mm-spaces with a certain number of points, where the weights of the corresponding point measure are fixed beforehand; so only the metric $d_{\mathbb{S}}$ has to be computed and is returned. To visualize the computed GW barycenters and the computed pairwise gLGW distances, we use the scikit-learn implementation of multi-dimensional scaling (MDS) from [31], which allows to

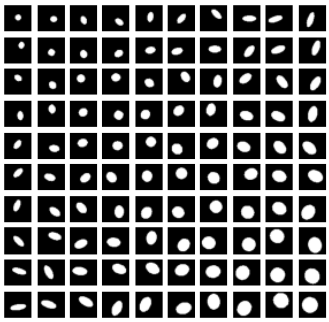

Fig. 1. Elliptical disks used for the first numerical experiment. The opposing images with respect to the diagonal form isometrical pairs resulting from rotations and shifts.

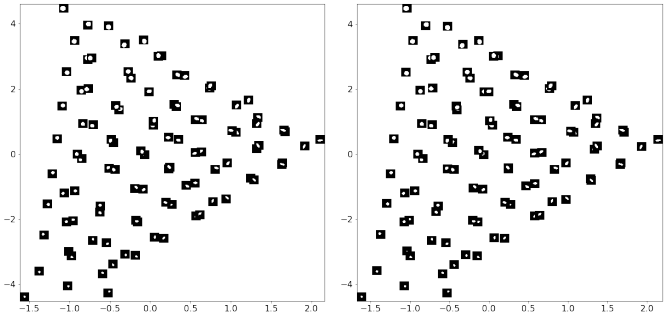

Fig. 2. MDS embeddings according to the computed pairwise GW distances (left) and gLGW distances (right) of the elliptical disks in Figure 1, where the reference corresponds to the uniform distribution on two circular disks.

embed a series of points with given distances into $\mathbb{R}^2$ or $\mathbb{R}^3$ such that the distances are approximately preserved. For all computations on graphs, we use the NetworkX package, which contains an implementation of Dijkstra's algorithm to compute shortest paths.

A. Gromov–Wasserstein of elliptical disks

For our first example, we apply gLGW to a toy problem, where we want to compute the GW distance between a series of elliptical disks, see Figure 1. Each image here consists of 50×50 equispaced pixels in $[-1, 1]^2$. For the numerical simulations, we interpret these images as discrete mm-spaces $\mathbb{X}_1, \dots, \mathbb{X}_{100}$. For this we set $\mathbb{X}_k := ([-1, 1]^2, d_{\mathrm{E}}, \mu_k)$, where d_{E} is the Euclidean distance and μ_k corresponds the uniform distribution on the position of the white pixels. Notice that, except for the diagonal, all elliptical disks in Figure 1 occur in isometrical pairs (up to discretization errors). Since the GW distance is invariant under isometries, this should be reflected in the computed GW and gLGW distances.

For comparison, we first compute all pairwise GW distances, where we use the optimal Wasserstein coupling as starting value for the corresponding POT algorithm in the GW distance computation. We visualize them by embedding the images as points in the plane using MDS. The result is shown on the left-hand side of Figure 2. Here the GW distance behaves as expected meaning that the isometrical pairs are found and located close to each other—the small visible distances between them result from the chosen discretizations.

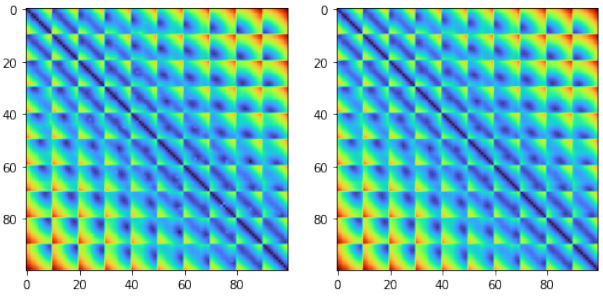

Fig. 3. Computed pairwise GW distances (left) and gLGW distances (right) of the elliptical disks in Figure 1, where the reference corresponds to the uniform distribution on two circular disks described in the text.

Fig. 4. Examples of the optimal GW plan between the reference mm-space (left) and some elliptical disks form Figure 1 (right). The color indicates the mass transport of the computed $\pi_{\mathbb{S}}^{\mathbb{X}_k}$ that is used for the barycentric projection.

Up to this expected doubling, we essentially obtain a triangle, whose corners correspond to the smallest as well as the largest (isotopic) elliptical disk and the most anisotropic elliptical disk. The 4950 pairwise GW distances of this toy example have been computed in 138.20 minutes.

To compute the generalized linear GW distance, we require a suitable reference mm-space $\mathbb{S}$. Here we have chosen an image with two circular disks and construct $\mathbb{S}$ in the same way as the mm-spaces $\mathbb{X}_k$, cf. the left-hand side image in Figure 4 without color. Moreover Figure 4 shows the optimal GW coupling between the reference $\mathbb{S}$ and some elliptical disks $\mathbb{X}_k$, where the color corresponds to the mass transport. Figuratively, the circular disks of the reference $\mathbb{S}$ are transported to the left-hand and right-hand side of the elliptical disks $\mathbb{X}_k$ and glued together along the semi-minor axis. On the basis of this couplings, we compute the barycentric projections (6) and the gLGW distances, which are visualized on the right-hand side of Figure 2. A direct comparison of the GW and gLGW distances is shown in Figure 3. Here gLGW gives a very good approximation of GW. The computation of the 4950 pairwise distances gLGW only requires 100 GW transport plans, therefore we merely needed 7.08 minutes.

B. Gromov–Wasserstein in 2D shape analysis

Next, we apply the GW distance and its linear form to distinguish different 2D shapes from each other. In this numerical experiment, we use the publicly available database [33] embracing over 1 200 shapes in 70 shape classes. For our example, we select 20 shapes of the classes bone, goblet, star, and horseshoe, respectively. The shapes are stored as black and white images of different sizes, where the white pixels corresponds to the objects. To speed-up the computations, each images is approximated by a point measure μ_k consisting of 50 points and uniform weights. For this preprocessing step, we use the dithering technique in [34], see also [35]. Afterwards, the obtained measures are randomly rotated yielding 80 mm-

Fig. 5. One example of each class (bone, goblet, star, horseshoe) of the employed 2D shape dataset.

Fig. 6. Embedding of the computed barycenter $\mathbb{S}$ into $\mathbb{R}^2$ using MDS.

spaces $\mathbb{X}_k = ([-1,1]^2, \|\cdot\|, \mu_k)$. A preprocessed example of each class is shown in Figure 5.

The performance of the generalized linear GW distance depends on the selection of an appropriate reference space $\mathbb{S}$. To ensure that the main features of each mm-space $\mathbb{X}_k$ are covered, we choose a representative for each class of shapes and compute their GW barycenter using the POT package, which takes 1.60 seconds. Numerically, it seems to be advantageous to compute the barycenter with less points than contained in the dithered images $\mathbb{X}_k$. In this example, the reference $\mathbb{S}$ is a barycenter with 35 points and uniform weights, which is shown in Figure 6. Since the reference $\mathbb{S}$ has less points than the spaces $\mathbb{X}_k$, the barycentric projection (6) is indeed a mapping to weighted means.

On the basis of the chosen reference $\mathbb{S}$, we now compute the pairwise GW distances (11.34 seconds) and based on the chosen reference $\mathbb{S}$ the pairwise generalized linear GW distances (0.72 seconds). The results are shown in Figure 7. Notice that the shapes in the horseshoe class significantly differ between each other explaining the greater distances than in other classes. In this example, gLGW gives a fair approximation of GW, and both distances are comparable.

Considering the results, it seems reasonable to use a nearest neighbor classification to distinguish the different classes of shapes with respect to some representatives. Numerically, this concept may be verified by computing a *confusion matrix* consisting of the probabilities to classify an instance of a class as another class. For this, we rely on [19, § 8.2], where the

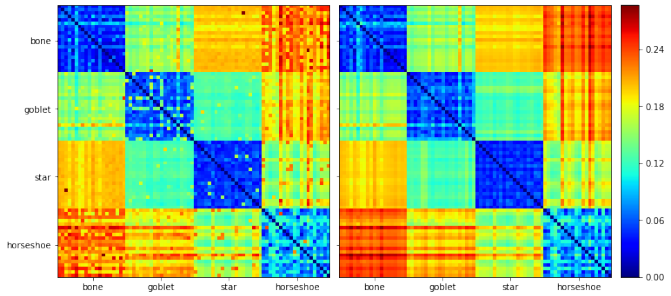

Fig. 7. Pairwise GW distances (left) and generalized linear GW distances for a barycenter reference $\mathbb{S}$. All images share the same color coding.

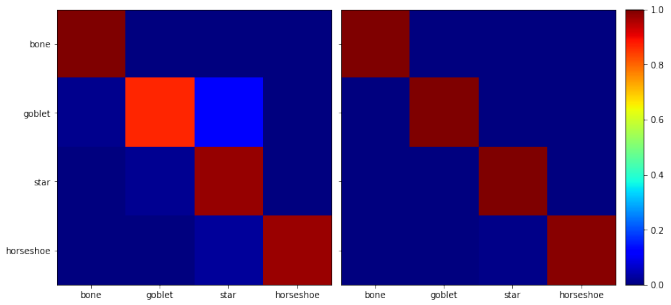

Fig. 8. Confusion matrix for the GW distances (left) and the generalized linear GW distances with barycenter reference (right).

confusion matrix is estimated by randomly choosing a representative for each class and then classifying all other shapes $\mathbb{X}_k$, $k = 1, \ldots, 80$, with respect to the nearest representative. This classification task is then repeated 10 000 times. The confusion matrix of GW and gLGW are shown in Figure 8. Note that the classification based on the generalized linear GW distance is even slightly better then based on the GW distance.

C. Gromov–Wasserstein in 3D shape analysis

The GW distance traces back to the comparison and matching of 3D shapes, which we take up in our final numerical example. Analytically, a 3D shape is a two-dimensional submanifold of $\mathbb{R}^3$ that may have a boundary. 3D shapes can be interpreted as mm-spaces $\mathbb{X} = (X, d_X, \mu)$, where X is a surface of the shape, where d_X corresponds to the length of the geodesics between two points, and where μ is some measure.

In practice, 3D shapes are usually triangulated and thus realized by a net of triangles. To handle them numerically, we approximate them by a discrete mm-space $\mathbb{X} = (X, d_X, \mu)$. The vertices of the net become the discrete points in X. To approximate the geodesic distance on X, a *weighted graph* $G = (X, E)$ consisting of all vertices X and all edges E of the triangulation may be used, where the edges are weighted by the Euclidean distance between the corresponding vertices. The geodesic distance between two vertices may now be approximated by the length of the shortest path between these vertices. This distance can be computed by the Dijkstra algorithm from the NetworkX package [32]. At the moment, μ may be some arbitrary point measure on X.

In this example, we consider 3D shapes of the publicly available database [36], which has already been used by Mémoli [19, § 8.2] in the context of GW distances. We use a similar setting to make the results comparable. As dataset for the experiment, we choose 3D shapes corresponding to the animals camel, cat, elephant, horse, and lion as well as to a human face and head. Each object is shown in 10 to 11 different poses totaling to 73 shapes. Figure 9 shows one example pose of each object. Every object is provided by a triangulation consisting of up to 43.000 vertices and up to 130.000 triangles.

Since the discrete mm-spaces of the full triangulations consist of too many points for our purpose, the 3D shapes are preprocessed by a two-step approximation similar to [19].

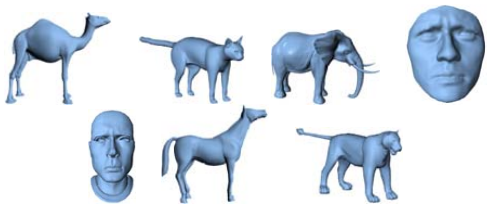

Fig. 9. One example of each class (camel, cat, elephant, face, head, horse, lion) of the employed 3D shape dataset.

1) Starting from a given triangulation with vertices X, we first reduce X to a set $\widetilde{X}$ consisting of 4 000 vertices. The first vertex is hereby chosen randomly and is sequentially followed by the points with the largest Euclidean distance to the already chosen points. This selection rule is also known as the furthest point procedure.
2) The set $\widetilde{X}$ is reduced further and an appropriate measure μ is constructed. For this, we again apply the furthest point procedure to reduce $\widetilde{X}$ to a subset $\widehat{X}$ consisting of 50 points, but this time with respect to the discrete geodesic distance d_X calculated using a weighted graph as explained above. Then we endow $\widehat{X}$ with a discrete probability measure, where the mass at $x \in \widehat{X}$ is proportional to the amount of closest neighbors within $\widetilde{X}$ with respect to the original geodesic distance d_X. In other words, we compute the Voronoi diagram of $\widetilde{X}$ to the points $\widehat{X}$ with respect to d_X and count the members of each Voronoi cell. Repeating this procedure for every given 3D shape, we end up with 73 discrete mm-spaces $\mathbb{X}_k = (\widehat{X}_k, d_{X_k}, \mu_k)$, $k = 1, \dots, 73$.

Since the distance d_X of the constructed mm-spaces $\mathbb{X}_k$ are discrete geodesic distances, which are restricted to the points in $\widehat{X}_k$, the barycentric projection (13) has the form

$$\mathcal{T}_{\pi_{\mathbb{S}}^{\mathbb{X}_k}}(s) \coloneqq \operatorname*{argmin}_{x_0 \in \widehat{X}_k} \sum_{x \in \widehat{X}_k} \pi_{\mathbb{S}}^{\mathbb{X}_k}(\{s\} \times \{x\})\, d_{X_k}^2(x_0, x),$$

where $\pi_{\mathbb{S}}^{\mathbb{X}_k}$ is the chosen optimal GW transport plan.

The pairwise GW distances (14.29 seconds) and the generalized linear GW distances (0.82 seconds) for the 3D shape dataset with respect to two different reference measures are shown in Figure 10. As reference $\mathbb{S}$ for gLGW, we first employ the POT package to compute the GW barycenter (4.67 seconds) between one model of each class, which corresponds to the middle image. The computed barycenter has 50 points corresponding to a uniform distribution. Secondly, we use a given reduced 3D shape, i.e. $\mathbb{S} \coloneqq \mathbb{X}_{10}$ (camel), which corresponds to the right-hand side image.

To evaluate the classification quality of the GW and gLGW distances, we again compute the confusion matrix C for both distances as in Section IV-B. The confusion matrix C consists of the probabilities to classify a 3D shape within one class (camel, cat, elephant, face, head, horse, lion) to another class. Following again [19, § 8.2], for this purpose, we first randomly chose a representative for each class and then classify all 3D shapes $\mathbb{X}_k$, $k = 1, \dots, 73$, with respect to the nearest representative. This classification task is repeated 10 000 times. The result is shown in Figure 11. Note that on the right image the classification probability for the cat, face, head, and lion class is nearly uniform along the corresponding rows, meaning that the classification has been nearly random. The reference $\mathbb{S} = \mathbb{X}_{10}$ is hence not suitable to distinguish the 3D shapes of the seven classes.

In conclusion, the quality of the computed gLGW distances essentially depends on the chosen reference space $\mathbb{S}$. Especially for classification tasks, the reference $\mathbb{S}$ has to contain features of all occurring classes. If this is the case, we obtain classification qualities comparable to the GW distance, however, with significantly less computational effort.

V. Conclusions

We proposed a linear version of the GW distance that was inspired by a generalized version of the linear Wasserstein distance. As the latter one, the approach appears to be efficient in applications, where pairwise distances of a larger amount of measures are of interest. We gave three examples indicating that our linear version of the GW distance gives reasonable approximations and circumvents the heavy computation of all pairwise distances. In contrast to Wasserstein distances, the mathematics behind GW distances is not well-examined so far and there are plenty of open problems which could be tackled in the future. For example, in generalized version of LOT, it would also be possible to use the concept of weak optimal transport [37]. This approach was neither considered for gLOT nor for gLGW so far. Further, multimarginals may be addressed, see [38]. Finally, we are interested in further meaningful applications of our approach.

Acknowledgment

The funding by the German Research Foundation (DFG) within the RTG 2433 DAEDALUS and by the BMBF project "VI-Screen" (13N15754) is gratefully acknowledged. Further, the authors would like to thank Johannes von Lindheim for valuable discussions and for assisting with numerical implementations.

Appendix A
Proofs

A. Proof of Proposition II.4

We have that $\tilde{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ since by Jensen's inequality

$$\begin{aligned}
&\int_{\mathbb{R}^d} \|x\|^2 \,\mathrm{d}\tilde{\mu}(x) \\
&= \int_{\mathbb{R}^d} \|\mathcal{T}_{\pi_\sigma^\mu}(s)\|^2 \,\mathrm{d}\sigma(s) = \int_{\mathbb{R}^d} \left\| \int_{\mathbb{R}^d} x \,\mathrm{d}\pi_s(x) \right\|^2 \mathrm{d}\sigma(s) \\
&\le \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \|x\|^2 \,\mathrm{d}\pi_s(x) \,\mathrm{d}\sigma(s) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x\|^2 \,\mathrm{d}\pi_\sigma^\mu(s, x) \\
&= \int_{\mathbb{R}^d} \|x\|^2 \,\mathrm{d}\mu(x) < \infty.
\end{aligned}$$

Let $\pi_\sigma^{\tilde{\mu}}$ be an optimal transport plan with respect to $W(\sigma, \tilde{\mu})$. By the dual formulation of the optimal transport problem, see [26, Thm 4.2] & [24, Thm 5.10], we know that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2} \|s - x\|^2 \,\mathrm{d}\pi_\sigma^{\tilde{\mu}}(s, x)$$

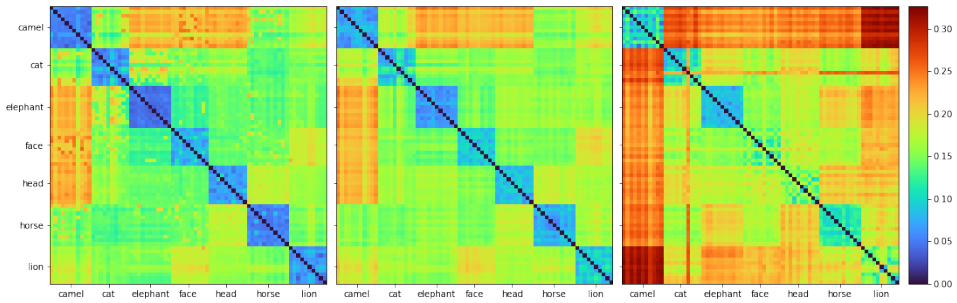

Fig. 10. Pairwise GW distances (left), generalized linear GW distances for a barycenter reference $\mathbb{S}$ (middle), and generalized linear GW distances with reference $\mathbb{S} = \mathbb{X}_{10}$ (camel) (right). All images share the same color coding.

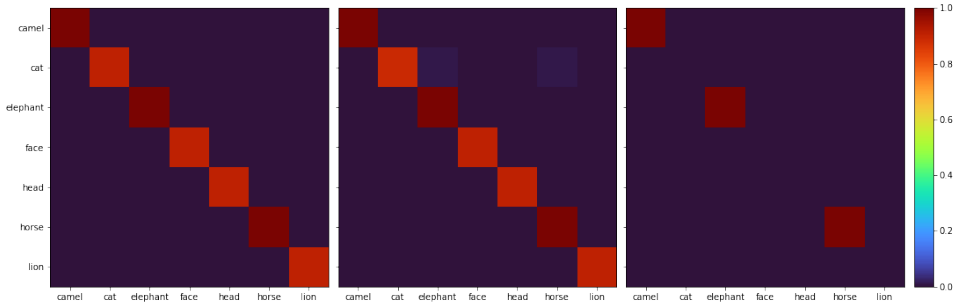

Fig. 11. Confusion matrices for the GW distances (left), the generalized linear GW distance with barycenter reference (middle), and the generalized linear GW distance with reference $\mathbb{S} = \mathbb{X}_{10}$ (right).

$$= \sup_{\phi \in L^1_\sigma(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \phi(s)\,\mathrm{d}\sigma(s) + \int_{\mathbb{R}^d} \phi^c(x)\,\mathrm{d}\tilde{\mu}(x) \right\},$$

where ϕ^c denotes the c-concave function given by

$$\phi^c(x) = \inf_{y \in X} \{ \tfrac{1}{2}\|x-y\|^2 - \phi(y) \}.$$

To yield a contradiction, assume that $\mathcal{T}_{\pi^\mu_\sigma}$ is not an optimal transport map. Then

$$\tilde{\pi} \coloneqq (\mathrm{id}, \mathcal{T}_{\pi^\mu_\sigma})_{\#}\sigma$$

is not an optimal transport plan with respect to $W(\sigma, \tilde{\mu})$ and

$$\begin{aligned}
&\int_{\mathbb{R}^d} \frac{1}{2}\|s - \mathcal{T}_{\pi^\mu_\sigma}(s)\|^2\,\mathrm{d}\sigma(s) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|s-x\|^2\,\mathrm{d}\tilde{\pi}(s,x) \\
&> \sup_{\phi \in L^1_\sigma(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \phi(s)\,\mathrm{d}\sigma(s) + \int_{\mathbb{R}^d} \phi^c(x)\,\mathrm{d}\tilde{\mu}(x) \right\}. \qquad (14)
\end{aligned}$$

Now we obtain for the optimal transport plan π^μ_σ of $W(\sigma, \mu)$ that

$$\begin{aligned}
&\int_{\mathbb{R}^d \times \mathbb{R}^d} \|s-x\|^2\,\mathrm{d}\pi^\mu_\sigma(s,x) \\
&= \int_{\mathbb{R}^d} \|s\|^2\,\mathrm{d}\sigma(s) - 2\int_{\mathbb{R}}^{d}\int_{\mathbb{R}}^{d} \langle s, x\rangle\,\mathrm{d}\pi_s(x)\,\mathrm{d}\sigma(s) \\
&\quad + \int_{\mathbb{R}^d} \|x\|^2\,\mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} \|s\|^2\,\mathrm{d}\sigma(s) - 2\int_{\mathbb{R}}^{d} \langle s, \mathcal{T}_{\pi^\mu_\sigma}(s)\rangle\,\mathrm{d}\sigma(s) \\
&\quad + \int_{\mathbb{R}^d} \|x\|^2\,\mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} \|s - \mathcal{T}_{\pi^\mu_\sigma}(s)\|^2\,\mathrm{d}\sigma(s) + \int_{\mathbb{R}^d} \|x\|^2\,\mathrm{d}\mu(x) \\
&\quad - \int_{\mathbb{R}^d} \|x\|^2\,\mathrm{d}\tilde{\mu}(x)
\end{aligned}$$

and by (14) further

$$\begin{aligned}
&\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|s-x\|^2\,\mathrm{d}\pi^\mu_\sigma(s,x) \\
&> \int_{\mathbb{R}^d} \frac{1}{2}\|x\|^2\,\mathrm{d}\mu(x) - \int_{\mathbb{R}^d} \frac{1}{2}\|x\|^2\,\mathrm{d}\tilde{\mu}(x) \\
&\quad + \sup_{\phi \in L^1_\sigma(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \phi(s)\,\mathrm{d}\sigma(s) + \int_{\mathbb{R}^d} \phi^c(x)\,\mathrm{d}\tilde{\mu}(x) \right\} \\
&= \sup_{\phi \in L^1_\sigma(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \frac{1}{2}\|x\|^2\,\mathrm{d}\mu(x) - \int_{\mathbb{R}^d} \frac{1}{2}\|x\|^2\,\mathrm{d}\tilde{\mu}(x) \right. \\
&\quad \left. + \int_{\mathbb{R}^d} \phi(s)\,\mathrm{d}\sigma(s) + \int_{\mathbb{R}^d} \phi^c(x)\,\mathrm{d}\tilde{\mu}(x) \right\} \\
&= \sup_{\phi \in L^1_\sigma(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \frac{1}{2}\|x\|^2\,\mathrm{d}\mu(x) + \int_{\mathbb{R}^d} \phi(s)\,\mathrm{d}\sigma(s) \right. \\
&\quad \left. + \int_{\mathbb{R}^d} \phi^c(x) - \frac{1}{2}\|x\|^2\,\mathrm{d}\tilde{\mu}(x) \right\}, \qquad (15)
\end{aligned}$$

where $L^1_\sigma(\mathbb{R}^d)$ is the space of functions which absolute values are integrable with respect to σ. Since ϕ^c is c-concave, we know that $h \coloneqq \phi^c - \frac{1}{2}\|\cdot\|^2$ is concave, see [26, Lect 4.4]. Thus, Jensen's inequality implies

$$\begin{aligned}
&\int_{\mathbb{R}^d} \phi^c(x) - \frac{1}{2}\|x\|^2\,\mathrm{d}\tilde{\mu}(x) \\
&= \int_{\mathbb{R}^d} \phi^c(\mathcal{T}_{\pi^\mu_\sigma}(s)) - \frac{1}{2}\|\mathcal{T}_{\pi^\mu_\sigma}(s)\|^2\,\mathrm{d}\sigma(s)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} h\left(\mathcal{T}_{\pi_\sigma^\mu}(s)\right)\,\mathrm{d}\sigma(x)\\
&= \int_{\mathbb{R}^d} h\left(\int_{\mathbb{R}^d} x\,\mathrm{d}\pi_s(x)\right)\,\mathrm{d}\sigma(s)\\
&\geq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h(x)\,\mathrm{d}\pi_s(x)\,\mathrm{d}\sigma(s)\\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \phi^c(x) - \frac{1}{2}\|x\|^2\,\mathrm{d}\pi_s(x)\,\mathrm{d}\sigma(s)\\
&= \int_{\mathbb{R}^d} \phi^c(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{R}^d}\frac{1}{2}\|x\|^2\,\mathrm{d}\mu(x). \qquad (16)
\end{aligned}
$$

Inserting (22) into (15), we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d}\frac{1}{2}\|s-x\|^2\,\mathrm{d}\pi_\sigma^\mu(s,x)\\
&> \sup_{\phi\in L^1_\sigma(\mathbb{R}^d)}\left\{\int_{\mathbb{R}^d}\phi(s)\,\mathrm{d}\sigma(s) + \int_{\mathbb{R}^d}\phi^c(x)\,\mathrm{d}\mu(x)\right\},
\end{aligned}
$$

which contradicts the optimality of π_σ^μ. □

B. Proof of Proposition III.1

To compute the distance (12), recall that the geodesics related to $\pi_{\mathbb{S}}^{\mathbb{X}} \in \Pi_{\mathrm{o}}(\mathbb{S},\mathbb{X})$ and $\pi_{\mathbb{S}}^{\mathbb{Y}} \in \Pi_{\mathrm{o}}(\mathbb{S},\mathbb{Y})$ are mapped to the tangents

$$
\begin{aligned}
g &\coloneqq F_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}}) = d_X - d_S \quad (\text{acting on } \mathbb{T}_g \coloneqq (S\times X, d_S, \pi_{\mathbb{S}}^{\mathbb{X}})),\\
h &\coloneqq F_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{Y}}) = d_Y - d_S \quad (\text{acting on } \mathbb{T}_h \coloneqq (S\times Y, d_S, \pi_{\mathbb{S}}^{\mathbb{Y}})),
\end{aligned}
$$

where $\mathbb{S}\sim\mathbb{T}_g\sim\mathbb{T}_h$. We next characterize the plans $\pi\in\Pi_{\mathrm{o}}(\mathbb{T}_g,\mathbb{T}_h)$ occurring in the definition of $\mathrm{GW}_{\mathbb{S}}$ in (11). Since $\mathbb{T}_g$ and $\mathbb{T}_h$ are equivalent, each plan $\pi\in\Pi_{\mathrm{o}}(\mathbb{T}_g,\mathbb{T}_h)$ satisfies

$$
\begin{aligned}
0 &= \mathrm{GW}(\mathbb{T}_g,\mathbb{T}_h)\\
&= \int_{(S\times X\times S\times Y)^2} |d_S(s_1,s_2) - d_S(s_1',s_2')|^2\\
&\qquad\qquad \times \mathrm{d}\pi(s_1,x,s_2,y)\,\mathrm{d}\pi(s_1',x',s_2',y'),\\
&= \int_{(S\times S)^2} |d_S(s_1,s_2) - d_S(s_1',s_2')|^2\,\mathrm{d}\gamma(s_1,s_2)\,\mathrm{d}\gamma(s_1',s_2'),
\end{aligned}
$$

where $\gamma \coloneqq P^{13}_{\#}\pi$. Thus, γ is an optimal self-coupling of $\mathbb{S}$ in the GW sense. As stated in [19, Lem 10.4], each self-coupling has the form $\gamma = (\mathrm{id},\psi)_{\#}\sigma$ for some measure-preserving isometry $\psi : S\to S$. This, however, implies that the mapping $P^{124}(s_1,x,s_2,y) = (s_1,x,y)$ is π-almost everywhere invertible by $(P^{124})^{-1}(s,x,y) = (s,x,\psi(s),y)$. More precisely, $(P^{124})^{-1}\circ P^{124}$ is the identity on $\mathrm{supp}(\pi)$. Therefore, we have

$$
\pi = (P^{124})^{-1}_{\#}\tilde{\pi} \quad \text{with} \quad \tilde{\pi} = P^{124}_{\#}\pi.
$$

Considering the marginals of $\tilde{\pi}$, we conclude that every 4-plan $\pi\in\Pi_{\mathrm{o}}(\mathbb{T}_g,\mathbb{T}_h)$ can be uniquely identified by the 3-plan $\tilde{\pi}\in\Pi_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}},\pi_{\mathbb{S}}^{\mathbb{Y}})$, and vice versa. This identification finally allows us to rewrite the metric (11) on the tangent space using the substitution $\pi = (P^{124})^{-1}_{\#}\tilde{\pi}$ to obtain

$$
\begin{aligned}
&\mathrm{GW}_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}},\pi_{\mathbb{S}}^{\mathbb{Y}})\\
&= \inf_{\pi\in\Pi_{\mathrm{o}}(\mathbb{T}_g,\mathbb{T}_h)}\int_{(S\times X\times S\times Y)^2}\big|d_X(x,x') - d_S(s_1,s_1')\\
&\qquad\qquad - d_Y(y,y') + d_S(s_2,s_2')\big|^2\\
&\qquad\qquad \times \mathrm{d}\pi(s_1,x,s_2,y)\,\mathrm{d}\pi(s_1',x',s_2',y'),\\
&= \inf_{\tilde{\pi}\in\Pi_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}},\pi_{\mathbb{S}}^{\mathbb{Y}})}\int_{(S\times X\times Y)^2}\big|d_X(x,x') - d_Y(y,y')\\
&\qquad\qquad + d_S(\psi(s),\psi(s')) - d_S(s,s')\big|^2\\
&\qquad\qquad \times \mathrm{d}\tilde{\pi}(s,x,y)\,\mathrm{d}\tilde{\pi}(s',x',y'),\\
&= \inf_{\tilde{\pi}\in\Pi_{\mathbb{S}}(\pi_{\mathbb{S}}^{\mathbb{X}},\pi_{\mathbb{S}}^{\mathbb{Y}})}\int_{(S\times X\times Y)^2}|d_X(x,x') - d_Y(y,y')|^2\\
&\qquad\qquad \times \mathrm{d}\tilde{\pi}(s,x,y)\,\mathrm{d}\tilde{\pi}(s',x',y'),
\end{aligned}
$$

which establishes the assertion. □

REFERENCES

[1] W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, and G. K. Rohde, "A linear optimal transportation framework for quantifying and visualizing variations in sets of images," *Int. J. Comput. Vis.*, vol. 101, no. 2, pp. 254–269, 2013.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, Basel, 2005.

[3] W. Wang, J. A. Ozolek, D. Slepčev, A. B. Lee, C. Chen, and G. K. Rohde, "An optimal transportation approach for nuclear structure-based pathology," *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 621–631, 2011.

[4] A. H. M. Rubaiyat, K. M. Hallam, J. M. Nichols, M. N. Hutchinson, S. Li, and G. K. Rohde, "Parametric signal estimation using the cumulative distribution transform," *IEEE Trans. Signal Process.*, vol. 68, no. 68, pp. 3312–3324, 2020.

[5] S. Kolouri, A. Tosun, J. Ozolek, and G. Rohde, "A continuous linear optimal transport approach for pattern analysis in image datasets," *Pattern Recognition*, vol. 51, pp. 453–462, 2016.

[6] S. Park, S. Kolouri, S. Kundu, and G. Rohde, "The cumulative distribution transform and linear pattern classification," *Appl. Comput. Harmon. Anal.*, 2017.

[7] T. H. Emerson and J. M. Nichols, "Fitting local, low-dimensional parameterizations of optical turbulencemodeled from optimal transport velocity vectors," *Pattern Recognition Lett.*, vol. 133, pp. 123–128, 2020.

[8] S. Basu, S. Kolouri, and G. Rohde, "Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry," *Proc. Natl. Acad. Sci. USA*, vol. 111, no. 9, pp. 3448–3453, 2014.

[9] J. Ozolek, A. Tosun, W. Wang, C. Chen, S. Kolouri, S. Basu, H. Huang, and G. Rohde, "Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning," *Med. Image. Anal.*, vol. 18, no. 5, pp. 772–780, 2014.

[10] A. B. Tosun, O. Yergiyev, S. Kolouri, J. F. Silverman, and G. K. Rohde, "Detection of malignant mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens," *Cytometry Part A*, vol. 87, no. 4, pp. 326–333, 2015.

[11] M. Eckermann, B. Schmitzer, F. van der Meer, J. Franz, O. Hansen, C. Stadelmann, and T. Salditt, "Three-dimensional virtual histology of the human hippocampus based on phase-contrast computed tomography," *PNAS*, vol. 118, no. 48, p. e2113835118, 2021.

[12] S. Guan, B. Liao, Y. Du, and X. Yin, "Vehicle type recognition based on Radon-CDT hybrid transfer learning," *IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 1–4, 2019.

[13] S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde, "De-multiplexing vortex modes in optical communications using transport-based pattern recognition," *Opt. Express*, vol. 26, no. 4, pp. 4004–4022, 2018.

[14] S. Kolouri, S. Park, and G. Rohde, "The Radon cumulative distribution transform and its application to image classification," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 920–934, 2016.

[15] P. C. Álvarez-Esteban, E. del Barrioa, J. A. Cuesta-Albertos, and C. Matrán, "A fixed-point approach to barycenters in Wasserstein space," *J. Appl. Math. Anal. Appl.*, vol. 441, pp. 744–762, 2016.

[16] A. Aldroubi, S. Li, and G. K. Rohde, "Partitioning signal classes using transport transformations for data analysis and machine learning," *arXiv:2008.03452v2*, 2021.

[17] C. Moosmüller and A. Cloninger, "Linear optimal transport embedding: Provable Wasserstein classification for certain rigid transformations and perturbations," *arXiv:2008.09165*, 2021.

[18] T. Cai, J. Cheng, B. Schmitzer, and M. Thorpe, "The linearized Hellinger-Kantorovich distance," *arXiv:2102.08807*, 2021.

[19] F. Mémoli, "Gromov–Wasserstein distances and the metric approach to object matching," *Found. Comput. Math.*, vol. 11, no. 4, pp. 417–487, 2011.
[20] K.-T. Sturm, "The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces," *arXiv:1208.0434*, 2020.
[21] F. Mémoli and T. Needham, "Distance distributions and inverse problems for metric measure spaces," *arXiv:1810.09646*, 2021.
[22] V. Titouan, R. Flamary, N. Courty, R. Tavenard, and L. Chapel, "Sliced Gromov-Wasserstein," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/a9cc6694dc40736d7a2ec018ea566113-Paper.pdf
[23] A. Salmona, J. Delon, and A. Desolneux, "Gromov-Wasserstein distances between Gaussian distributions," *arXiv:2104.07970*, 2021.
[24] C. Villani, *Optimal Transport: Old and New*. Springer, 2008, vol. 338.
[25] Y. Brenier, "Décomposition polaire et réarrangement monotone des champs de vecteurs," *Comptes rendus de l'Académie des Sciences, Paris, Série I*, vol. 305, pp. 805–808, 1987.
[26] L. Ambrosio, E. Brué, and D. Semola, *Lectures on Optimal Transport*, ser. Unitext. Cham: Springer, 2021, no. 130.
[27] T. Vayer, "A contribution to optimal transport on incomplete spaces," *PhD Thesis, Université Bretagne*, 2020.
[28] G. Peyré, M. Cuturi, and J. Solomon, "Gromov-Wasserstein averaging of kernel and distance matrices," in *International Conference on Machine Learning*, 2016, pp. 2664–2672.
[29] ——, "Gromov-Wasserstein learning for graph matching and node embedding," in *International Conference on Machine Learning*, 2019, pp. 6932–6941.
[30] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer, "POT: Python optimal transport," *J. Mach. Learn. Res.*, vol. 22, no. 78, pp. 1–8, 2021. [Online]. Available: http://jmlr.org/papers/v22/20-451.html
[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[32] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., 2008, pp. 11–15.
[33] A. Carlier, K. Leonard, S. Hahmann, G. Morin, and M. Collins, "The 2D shape structure dataset," https://2dshapesstructure.github.io.
[34] T. Teuber, G. Steidl, P. Gwosdek, C. Schmaltz, and J. Weickert, "Dithering by differences of convex functions," *SIAM J. Imaging Sci.*, vol. 4, no. 1, pp. 79–108, 2011.
[35] M. Ehler, M. Gräf, S. Neumayer, and G. Steidl, "Curve based approximation of measures on manifolds by discrepancy minimization," *Foundations in Computational Mathematics*, vol. 21, no. 6, pp. 1595–1642, 2021.
[36] R. W. Sumner and J. Popovic, "Mesh data from deformation transfer for triangle meshes," http://people.csail.mit.edu/sumner/research/deftransfer/data.html.
[37] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, "Kantorovich duality for general transport costs and applications," *J. Funct. Anal.*, vol. 273, no. 11, pp. 3327–3405, 2017.
[38] F. Beier, J. von Lindheim, S. Neumayer, and G. Steidl, "Unbalanced multi-marginal optimal transport," *arXiv:2103.10854*, 2021.