Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

SIAM J. Imaging Sciences
Vol. 17, No. 3, pp. 1433–1466

© 2024 Society for Industrial and Applied Mathematics

A Wasserstein-Type Distance for Gaussian Mixtures on Vector Bundles with Applications to Shape Analysis[*]

Michael Wilson[†], Tom Needham[‡], Chiwoo Park[§], Suparteek Kundu[¶], and Anuj Srivastava[†]

Abstract. This paper uses sample data to study the problem of comparing populations on finite-dimensional parallelizable Riemannian manifolds and more general trivial vector bundles. Utilizing triviality, our framework represents populations as mixtures of Gaussians on vector bundles and estimates the population parameters using a mode-based clustering algorithm. We derive a Wasserstein-type metric between Gaussian mixtures, adapted to the manifold geometry, in order to compare estimated distributions. Our contributions include an identifiability result for Gaussian mixtures on manifold domains and a convenient characterization of optimal couplings of Gaussian mixtures under the derived metric. We demonstrate these tools on some example domains, including the preshape space of planar closed curves, with applications to the shape space of triangles and populations of nanoparticles. In the nanoparticle application, we consider a sequence of populations of particle shapes arising from a manufacturing process and utilize the Wasserstein-type distance to perform change-point detection.

Key words. Gaussian mixtures, optimal transport, shape analysis

MSC codes. 53Z15, 62P30

DOI. 10.1137/23M1620363

1. Introduction.

Modern statistical analysis increasingly involves data objects that are nonlinear and non-Euclidean. A prominent example is *directional data* [28], where data naturally lie on a unit sphere. Another example is shape analysis, where one is interested in analyzing *shapes* of imaged objects. Although several approaches have been developed for shape analysis (see [17, 25, 34, 36, 45, 2] and others), they all agree in that the representation spaces of shapes is nonlinear. Examples of nonlinear data domains are also present in covariance analysis [16, 46], functional data analysis [36], and graphical data [22, 10, 20]. Analysis of non-Euclidean data requires statistical tools adapted to the differential geometries of the underlying representation spaces. These tools include statistical modeling, parameter estimation, and inferences. Our paper is focused on a specific subproblem in this broad field, i.e., comparing probability distributions on certain nonlinear domains. Precisely, we

[*] Received by the editors November 28, 2023; accepted for publication (in revised form) April 5, 2024; published electronically July 11, 2024.
https://doi.org/10.1137/23M1620363

Funding: The work of the authors was partially supported by National Science Foundation grants CMMI-2132311, CDS&E 1953087, DMS-2107808, and DMS-2324962 and NIH grant R01 MH120299.

[†] Department of Statistics, Florida State University, Tallahassee, FL 32306 USA (mwilson5@fsu.edu, anuj@stat.fsu.edu).

[‡] Department of Mathematics, Florida State University, Tallahassee, FL 32306 USA (tneedham@fsu.edu).

[§] Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98195 USA (chiwoo.park@eng.famu.fsu.edu).

[¶] The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA (skundu2@mdanderson.org).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

will model the probability distributions as *mixtures of Gaussians,* adapted to nonlinear domains of interest, and compare them using a novel variant of the Wasserstein distance. These choices—mixtures of Gaussian models and Wasserstein metric—are driven by convenience and applicability. Gaussian mixtures [12, 39] provide a general yet parametric option for capturing population variability, and Wasserstein metrics have become a canonical choice for comparing distributions in a variety of contexts [40, 33].

We develop a general framework for domains that are trivial vector bundles. A *vector bundle* is a set of isomorphic vector spaces indexed by points on a smooth manifold [31]; a vector bundle is called *trivial* if this structure can be realized as a product of a smooth manifold with a vector space. The tangent bundle of a *parallelizable manifold* is an example of a trivial vector bundle. Examples of parallelizable manifolds include punctured spheres, spaces of symmetric positive definite matrices, Lie groups, and several commonly used shape spaces.

Distributions on vector bundles provide a natural setting for combining results from the optimal transport [12, 27] and shape analysis [17, 36] literature. The two main goals of this paper are to develop a general theory for comparing certain distributions on trivial vector bundles (section 3) and to apply this theory to study shape populations arising from images captured in a nanomanufacturing process (section 4). The main challenges in deriving a framework for comparing probability distributions on vector bundles include (1) the nonlinearity of underlying domains and the specification of convenient probability distributions for such domains, (2) efficient estimation of these distributions from data, and (3) comparisons of estimated distributions using proper metrics between distributions. We outline these choices next:

1. **Forms of probability distributions**: Our first task is to define a probability distribution on a trivial bundle. While nonparametric approaches, often based on kernel methods [23], have gained prominence due to their generality and broad applicability, they require large sample sizes to capture the population variability effectively. In contrast, parametric families, such as mixtures of Gaussians, are robust under small sample sizes and have been covered extensively in the past literature. As mentioned earlier, our situation is complicated due to the nonlinearity of targeted domains. While some parametric families have advanced from Euclidean to nonlinear domains, the choice is relatively limited. Some adaptations of Gaussians to nonlinear and compact domains include truncated Gaussians, von Mises distributions, and wrapped Gaussian distributions [28]. This paper represents the underlying distribution as a mixture of Gaussians defined appropriately for vector bundles. The primary motivation for choosing Gaussian mixtures is their generality, simplicity, and interpretability of the resulting Wasserstein distance.
2. **Estimating probability distributions**: The next issue is efficiently estimating Gaussian mixtures from given data. Depending on the chosen space and the Riemannian metric, several papers have studied the estimation of basic summary statistics from the data, such as means and covariances. However, the literature on estimating parameters of mixtures of Gaussians on nonlinear domains is relatively limited. The main issue is computational. The expectation maximization (EM) algorithm is an established approach for estimating Gaussian mixtures on vector spaces but is

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

expensive and prone to local solutions. Furthermore, its applications to nonlinear domains are costlier due to the iterative nature of mean computation [35, 8]. Some papers have adapted EM algorithms for truncated Gaussians and von Mises distributions to nonlinear spaces (see [21]). We will apply a recent method that performs clustering on manifold data by finding *modes* of the underlying distribution [14]. We shall treat these modes as estimates of Gaussian means and further estimate covariances within individual clusters. The procedure for estimating cluster memberships, means, and covariances for general metric spaces is provided in [13, 14].

3. **Metrics between probability distributions**: The final issue is defining a metric for comparing and quantifying differences between chosen distributions in vector bundles. While there are several choices for this metric, the Wasserstein metric has become popular for several reasons. When available, it provides an interpretable solution for comparing probability distributions. It is also robust to misspecifications in distributions due to estimation errors. Finally, it leads to a closed-form expression for comparing certain parametric families, in particular, Gaussians. In this paper, we will utilize a Wasserstein-type metric previously developed for Euclidean domains [12]. (These are called Wasserstein type because the couplings—joint distributions for minimizing cost function—are restricted to be mixtures of Gaussians rather than all distributions.) Specifically, we will derive this metric for comparing mixtures of Gaussians on trivial vector bundles.

   A key idea that helps us define distances between Gaussians on trivial vector bundles is that any trivialization leads to a consistent choice of basis for each of the tangent spaces of the manifold—i.e., it provides a *global frame.* Fixing consistent coordinates allows us to apply, in a coherent manner, closed-form expressions for distances between Gaussians on different tangent spaces. Later on, we give examples of simple families of trivializations which are natural from a data analysis persepective.

We will demonstrate these ideas using both simulated and real-world data sets. In the simulated study, we will generate samples from mixtures of Gaussians on the (punctured) unit sphere $\mathbb{S}^2$ and demonstrate procedures for parameter estimation and population comparisons. We will also consider an example where the shape space of planar triangles [24, 7] is identified with $\mathbb{S}^2$, so that one can compare shape populations of triangles. In the real-data study, we investigate transmission electron microscopy videos of particles in nanomanufacturing processes where each image frame contains hundreds of particles that differ in shape, size, and placement. We focus on the shapes of their contours and treat shapes in a frame as random samples from underlying shape populations. We model these shape populations as mixtures of Gaussians on the shape space of planar, closed contours. As mentioned above, we use a mode estimation procedure to infer the parameters of mixtures from the observed shape data for each frame separately. The goal is to track and compare temporal evolutions of these shape distributions and to quantify their changes over time. For instance, we use this quantification to detect change points in the nanomanufacturing process.

The salient contributions of this paper are as follows:

1. It extends the notion of Gaussians and mixtures of Gaussians to trivial vector bundles and uses them for statistical modeling and analysis. Examples of such domains include punctured spheres, tori, matrix Lie groups, spaces of symmetric positive definite matrices, shape spaces, and other domains useful in statistical analysis.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

2. It derives a convenient expression for comparing mixtures of Gaussians using a Wasserstein-type metric. This development provides useful insights into choices made for problem domains, probability models, and metrics for comparing populations.
3. It applies these tools to comparisons of populations of planar contour shapes and for finding change points in the temporal evolution of shape populations.

The paper proceeds as follows. In section 2, we cover the background information necessary to introduce the Wasserstein-type distance for Gaussian mixtures on $\mathbb{R}^n$. In section 3, we present our proposed framework for extending this Wasserstein-type distance to mixtures of Gaussians on vector bundles. In section 4, we present our experimental results involving real and simulated data. Section 5 concludes the paper with some observations.

2. Background on Wasserstein distances.

This section introduces some background material and existing results to lay the groundwork for our approach. Specifically, we focus on the Gaussian mixtures on Euclidean spaces and the expressions for Wasserstein distances between such mixtures.

2.1. Classical Wasserstein distance.

We begin with necessary background material on classical distances between probability distributions, called *Wasserstein distances,* on a general metric space.

Wasserstein distances for metric spaces. Let $(\mathcal{X}, d)$ be a metric space.

Definition 2.1. *For $p \geq 1$, the* Wasserstein space $\mathcal{P}_p(\mathcal{X})$ *is the set of probability measures on $\mathcal{X}$ with finite pth moment; i.e., for every $x_0 \in \mathcal{X}$, the integral $\int_{\mathcal{X}} d(x_0,x)^p d\mu(x)$ is finite. The* p-Wasserstein distance $W_p^{\mathcal{X}}$ *between probability measures $\mu_0, \mu_1 \in \mathcal{P}_p(\mathcal{X})$ is given by*

$$W_p^{\mathcal{X}}(\mu_0,\mu_1) := \left(\inf_{\gamma \in \Pi(\mu_0,\mu_1)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^p d\gamma(x,y) \right)^{\frac{1}{p}}, \tag{2.1}$$

where $\Pi(\mu_0,\mu_1)$ is the set of couplings *of μ_0 and μ_1, that is, the set of joint probability measures γ on $\mathcal{X}\times\mathcal{X}$ that have marginal distributions μ_0 and μ_1. A joint measure γ that achieves the infimum of* (2.1) *is called an* optimal coupling.

The field of *optimal transport* studies properties of the Wasserstein distance and related constructions; see, for example, [33, 1, 40] for overviews of the well-developed theory of optimal transport. In particular, the Wasserstein distance is a metric on $\mathcal{P}_p(\mathcal{X})$, under mild assumptions on $\mathcal{X}$ (e.g., $\mathcal{X}$ is a Polish space).

Finitely supported measures. From an applications-oriented perspective, it is most common to consider the Wasserstein distance between finitely supported distributions. In this setting, calculating the Wasserstein distance comes down to solving a constrained linear program. Indeed, for $i = 0, 1$, let

$$\mu_i = \Sigma_{k=1}^{K_i} \alpha_i^k \delta_{x_i^k}, \ \Sigma_{k=1}^{K_i} \alpha_i^k = 1, \ \text{where } \alpha_i^k > 0 \ \forall\, i,k$$

are probability measures supported on points $x_i^k \in \mathcal{X}$. By an abuse of notation, we consider μ_i as a (column) vector $\mu_i = [\alpha_i^1, \ldots, \alpha_k^{K_i}]^T \in \mathbb{R}^{K_i}$. Then the space of couplings can be identified with a set of matrices,

$$\Pi(\mu_0,\mu_1) = \{\pi \in \mathbb{R}^{K_0 \times K_1} : \pi^T \mathbf{1} = \mu_0, \mathbf{1}^T \pi = \mu_1^T\},$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

where $\mathbf{1}$ represents the column vector of all ones, whose size is inferred by context. When considering a coupling π as a matrix, we write its (i,j)-entry as π_{ij}. Then the Wasserstein p-distance is given by

$$W_p^{\mathcal{X}}(\mu_0,\mu_1)^p = \min_{\pi\in\Pi(\mu_0,\mu_1)} \Sigma_{i,j=1}^{K_0,K_1}\pi_{ij} d(x_0^i,x_1^j)^p = \min_{\pi\in\Pi(\mu_0,\mu_1)} \langle D,\pi\rangle_F,$$

where $\langle\cdot,\cdot\rangle_F$ is the Frobenius inner product on $\mathbb{R}^{K_0\times K_1}$ and $D\in\mathbb{R}^{K_0\times K_1}$ is the matrix with the (i,j)-entry given by $d(x_0^i,x_1^j)^p$. This shows that the objective of the Wasserstein distance computation is a linear function, and it is not hard to see that the constraint set $\Pi(\mu_0,\mu_1)$ is a convex polytope in $\mathbb{R}^{K_0\times K_1}$.

Gaussian distributions on $\mathbb{R}^d$. In general, calculating Wasserstein distances between continuous distributions is impossible due to the infinite-dimensional nature of the associated optimization problem. However, in the case of Gaussian distributions, there is a simple closed-form equation for the Wasserstein distance in terms of the parameters of the distributions. We use $N_d(m,\Sigma)$ to denote the Gaussian distribution on $\mathbb{R}^d$ with mean $m\in\mathbb{R}^d$ and covariance $\Sigma\in\mathsf{Sym}_d^+$, where $\mathsf{Sym}_d^+\subset\mathbb{R}^{d\times d}$ denotes the set of symmetric positive semidefinite matrices. When considering $\mathbb{R}^d$ as a metric space, we always use the standard Euclidean metric. The following result is classical.

Proposition 2.2 (see [15, 19, 30]). *Given two Gaussian distributions on $\mathbb{R}^d$, $\eta_i = N_d(m_i,\Sigma_i)$, $i\in\{0,1\}$, the squared 2-Wasserstein distance between μ_0 and μ_1 is given by*

$$W_2^{\mathbb{R}^d}(\eta_0,\eta_1)^2 = \|m_0-m_1\|^2 + \operatorname{tr}\left(\Sigma_0+\Sigma_1-2\left(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right), \tag{2.2}$$

where tr *denotes the matrix trace. Moreover, an optimal coupling is given by a Gaussian measure on $\mathbb{R}^d\times\mathbb{R}^d$. If $m_0=m_1=0\in\mathbb{R}^d$, then the optimal coupling will be a zero-mean Gaussian.*

Let $\mathsf{G}_d := \{N_d(m,\Sigma) : m\in\mathbb{R}^d, \Sigma\in\mathsf{Sym}_d^+\}$ denote the set of Gaussian measures on $\mathbb{R}^d$. The above implies that (G_d, W_2) is a metric space whose metric is explicitly computable (here, we use $W_2 = W_2|_{\mathsf{G}_d\times\mathsf{G}_d}$ by abuse of notation). Due to this computational convenience, we focus on the $p=2$ version of the Wasserstein distance for the rest of the paper.

2.2. Gaussian mixture measures on $\mathbb{R}^d$.

The closed formula (2.2) for the Wasserstein distance between Gaussians suggests that we consider a richer set of measures consisting of collections of Gaussians, given more precisely as follows.

Definition 2.3. *A measure μ on $\mathbb{R}^d$ is a* Gaussian mixture measure *(or just* Gaussian mixture*) if it can be written as*

$$\mu = \Sigma_{k=1}^K w_k\eta_k, \text{ where } \eta_k = N_d(m_k,\Sigma_k) \text{ and } \Sigma_{k=1}^K w_k = 1,\ w_k>0,\ \forall k. \tag{2.3}$$

A Gaussian mixture μ can also be considered as a discrete probability measure on G_d. We use μ^* to distinguish this representation and write

$$\mu^* = \Sigma_{k=1}^K w_k\delta_{\eta_k}, \text{ where } \eta_k\in\mathsf{G}_d \text{ and } \Sigma_{k=1}^K w_k = 1,\ w_k>0,\ \forall k.$$

We use GM_d to denote the collection of all Gaussian mixtures on $\mathbb{R}^d$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

An important property of Gaussian mixtures is that they are *identifiable* in a certain precise sense, meaning that the representation given in (2.3) is essentially unique. Let us now explain this more precisely. The representation (2.3) is not strictly unique, as one could, for example, rearrange the terms or replace a term $w_k\eta_k$ by $(w_k/2)\eta_k + (w_k/2)\eta_k$ without changing the resulting measure. If a Gaussian mixture μ is written in the form (2.3) such that all Gaussians η_k are pairwise distinct, we say that the representation is in *minimal form.* We have the following classical result from [44] (see also [12, Proposition 2]), which we record here for later use.

Proposition 2.4 (see [44]). *Let μ be a Gaussian mixture, with minimal form representations*

$$\sum_{k=1}^{K} w_k\eta_k \quad \text{and} \quad \sum_{k=1}^{K'} w'_k\eta'_k.$$

Then $K = K'$, and there exists a permutation σ of $\{1,\dots,K\}$ such that $w_k = w'_{\sigma(k)}$ and $\eta_k = \eta'_{\sigma(k)}$ for all k.

The two perspectives on Gaussian mixture measures described in Definition 2.3 lead to two candidate metrics on GM_d. On one hand, one could compute the Wasserstein distance $W_2^{\mathbb{R}^d}(\mu_0,\mu_1)$ between $\mu_0,\mu_1 \in \mathsf{GM}_d$. On the other hand, one could compute the Wasserstein distance in the metric space of discrete measures on G_d, which reads as

$$W_2^{\mathsf{G}_d}(\mu_0^*,\mu_1^*)^2 = \min_{\pi\in\Pi(w_0,w_1)} \Sigma_{k,l}\pi_{kl}W_2^{\mathbb{R}^d}(\eta_0^k,\eta_1^\ell)^2, \tag{2.4}$$

where $\mu_i = \sum_k w_i^k\eta_i^k$ and $w_i = (w_i^k)_k$ for $i \in \{0,1\}$. This latter notion of distance between Gaussian mixtures was first studied in [9].

In general, $W_2^{\mathbb{R}^d}(\mu_0,\mu_1)$ and $W_2^{\mathsf{G}_d}(\mu_0^*,\mu_1^*)$ are not equal. It turns out that this discrepancy can be reconciled by adding an extra constraint to the feasible set in the Wasserstein distance optimization problem. The following was first introduced in [12].

Definition 2.5 (see [12]). *Given $\mu_i \in \mathsf{GM}_d$, $i \in \{0,1\}$, the* mixture Wasserstein distance *is given by*

$$MW_2^{\mathbb{R}^d}(\mu_0,\mu_1)^2 := \inf_{\pi\in\Pi(\mu_0,\mu_1)\cap\mathsf{GM}_{2d}} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 \, d\pi(x,y).$$

It is shown in [12, Proposition 4] that the alteration of the Wasserstein distance given in Definition 2.5 agrees with the distance between Gaussian mixtures used in (2.4). We record this result here.

Theorem 2.6 (see [12]). *For $\mu_i \in \mathsf{GM}_d$, $i \in \{0,1\}$, we have $MW_2^{\mathbb{R}^d}(\mu_0,\mu_1) = W_2^{\mathsf{G}_d}(\mu_0^*,\mu_1^*)$.*

3. Distances between Gaussian mixtures on vector bundles.

The main contribution of this paper is to generalize the Wasserstein-type distance $MW_2^{\mathbb{R}^d}$ to Gaussian mixtures defined on trivial vector bundles. We first introduce some preliminary ideas. Proofs of all results are deferred to Appendix A.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

3.1. Preliminaries on Gaussian mixtures on vector bundles.

Gaussian measures on inner product spaces. Let $(V, \langle\cdot,\cdot\rangle_V)$ be a finite-dimensional inner product space. We wish to consider Gaussian measures on V. These could, of course, be defined by choosing an isometry to $\mathbb{R}^d$ and transferring the standard definition. It will be convenient for later computations to have a more coordinate-free description of Gaussian measures on V. We develop this point of view here.

Definition 3.1. *A Borel measure μ on V is called* Gaussian *if, for every linear functional $f : V \to \mathbb{R}$, the pushforward $f_\#\mu$ is a Gaussian measure (in the standard sense) on $\mathbb{R}$.*

Definition 3.1 is used in [5, Definition 1.2.1] to characterize Gaussian measures on Euclidean spaces. Indeed, it is straightforward to show that any Gaussian $N_d(m,\Sigma)$ on $\mathbb{R}^d$ (in the sense of the previous section) has this pushforward property. Using this definition for more general inner product spaces allows us to easily relate Gaussian measures to their Euclidean counterparts.

Proposition 3.2. *A Borel measure μ on $(V, \langle\cdot,\cdot\rangle_V)$, where $\dim(V) = d$, is Gaussian if and only if there is a Gaussian measure ν on $\mathbb{R}^d$ and a linear isometry $g : \mathbb{R}^d \to V$ such that $\mu = g_\#\nu$.*

Given a Gaussian $\nu = N_d(m,\Sigma)$ on $\mathbb{R}^d$ and a linear isometry $g : \mathbb{R}^d \to V$, let $\mu = g_\#\nu$. We define the mean and covariance of μ to be $g(m)$ and $g\Sigma g^{-1}$, respectively. Here, we consider Σ as an operator $\mathbb{R}^d \to \mathbb{R}^d$. We first show that these quantities do not depend on the choice of ν or g.

Proposition 3.3. *Let $\nu = N_d(m,\Sigma)$ and $\nu' = N_d(m',\Sigma')$ be Gaussians on $\mathbb{R}^d$, and let g and g' be linear isometries from $\mathbb{R}^d$ to V such that $g_\#\nu = g'_\#\nu'$. Then $g(m) = g'(m')$ and $g\Sigma g^{-1} = g'\Sigma'(g')^{-1}$.*

As a corollary, we get that the following is a valid definition.

Definition 3.4. *Let $(V, \langle\cdot,\cdot\rangle_V)$ be an inner product space, $g : \mathbb{R}^d \to V$ be a linear isometry, and $\nu = N_d(m,\Sigma) \in \mathsf{G}_d$. The* mean *of the Gaussian measure $g_\#\nu$ on V is $g(m)$, and the* covariance operator *is $g\Sigma g^{-1}$.*

Gaussian mixtures on vector bundles. Let $\mathcal{M}$ be a Riemannian manifold. When dealing with data valued in such a manifold, it is common to *linearize* the analysis by choosing a basepoint $m \in \mathcal{M}$ and pulling the data back to the tangent space $T_m\mathcal{M}$ via the Riemannian log map; for example, this is a standard technique in statistical shape analysis [35, 36, 4] and computational optimal transport [42, 10]. From a modeling perspective, it is often useful to fit a distribution to the linearized data, resulting in a probability distribution on $T_m\mathcal{M}$. One can consider the resulting distribution as a highly singular probability measure on the tangent bundle $T\mathcal{M}$ in the sense that it is only supported on the fiber $T_m\mathcal{M} \subset T\mathcal{M}$. This is the perspective taken in [36, 27], where the authors consider Gaussian distributions on tangent spaces as models for “wrapped Gaussians” on the underlying manifold (this terminology originates in the directional statistics literature—see [11, 28]). Observe that the well-definedness of this framework depends on technicalities such as the domain of the log map—this hints at the utility of considering parallelizable manifolds, as we do in the following.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

In this paper, we propose a data model which linearizes subsets of the data at various strategically chosen basepoints, $m_1,\dots,m_K \in \mathcal{M}$. Fitting (weighted) distributions on each $T_{m_k}\mathcal{M}$ leads to a more general singular measure on $T\mathcal{M}$ whose support is contained in $\cup_k T_{m_k}\mathcal{M} \subset T\mathcal{M}$. Arguably, the simplest such model involves fitting zero-mean Gaussian distributions in each tangent space.

We now formalize the concepts described above. It will be convenient to work, more generally, in the setting of vector bundles. Let $p:\mathcal{E}\to\mathcal{M}$ be a rank-d vector bundle over a smooth manifold $\mathcal{M}$; in what follows, we typically denote the vector bundle as $\mathcal{E}\to\mathcal{M}$, with the understanding that there is an underlying projection map that has been supressed from the notation. We denote the fiber over $m\in\mathcal{M}$ as $\mathcal{E}_m \approx \mathbb{R}^d$. Let $\langle\cdot,\cdot\rangle = \{\langle\cdot,\cdot\rangle_m\}_{m\in M}$ be a smoothly varying family of inner products on the fibers $\mathcal{E}_m$.

Definition 3.5. *A Borel measure η on the vector bundle $\mathcal{E}$ is called a* Gaussian measure *if it is a mean-zero Gaussian measure on the inner product space $(\mathcal{E}_m,\langle\cdot,\cdot\rangle_m)$ for some $m\in\mathcal{M}$ (see Definition 3.4). If the covariance operator of η is Σ, we write $\eta = N_{\mathcal{E}}(m,\Sigma)$. The collection of Gaussian measures on $\mathcal{E}$ is denoted $\mathsf{G}(\mathcal{E})$.*

A Borel measure μ on $\mathcal{E}$ is called a Gaussian mixture measure *(or just* Gaussian mixture*) if it can be written as $\mu = \Sigma_{k=1}^K w_k\eta_k$, where each $\eta_k = N_{\mathcal{E}}(m_k,\Sigma_k)$ for some $m_k\in\mathcal{M}$ and where $\Sigma_{k=1}^K w_k = 1$. We denote the collection of all Gaussian mixtures on $\mathcal{E}$ as $\mathsf{GM}(\mathcal{E})$.*

As an important example, consider a product bundle $\mathcal{E}=\mathcal{M}\times\mathbb{R}^d$, where $\mathcal{E}_m = \{m\}\times\mathbb{R}^d \approx \mathbb{R}^d$ is endowed with the standard inner product. Then a Gaussian mixture on $\mathcal{E}$ is simply a collection of mean-zero Gaussians indexed by a finite collection of points $\{m_1,\dots,m_K\}$ in $\mathcal{M}$. In particular, the following result is immediate.

Proposition 3.6. *We have $\mathsf{G}_d \approx \mathsf{G}(\mathbb{R}^d\times\mathbb{R}^d)$ and $\mathsf{GM}_d \approx \mathsf{GM}(\mathbb{R}^d\times\mathbb{R}^d)$ as sets. To make this more precise, let $\eta = N_d(m,\Sigma)$ be a Gaussian measure on $\mathbb{R}^d$, and let $\overline{\eta}$ denote the measure when considered as a Gaussian measure on the trivial bundle $\mathcal{E}=\mathbb{R}^d\times\mathbb{R}^d$, that is, $\overline{\eta} = N_{\mathcal{E}}(m,\Sigma)$. The map $\eta\mapsto\overline{\eta}$ induces a bijective correspondence between Gaussian mixture measures GM_d on $\mathbb{R}^d$ (in the sense of Definition 2.3) and Gaussian mixture measures $\mathsf{GM}(\mathcal{E})$ on $\mathcal{E}$ (in the sense of Definition 3.5). Explicitly, the bijection maps $\mu\in\mathsf{GM}_d$, written in minimal form as $\sum_k w_k\eta_k$, to $\sum_k w_k\overline{\eta}_k \in \mathsf{GM}(\mathcal{E})$.*

We now extend the discussion of identifiability of Gaussian mixture measures (see Proposition 2.4) to the setting of vector bundles. In analogy with the Euclidean setting, if a Gaussian mixture measure μ on $\mathcal{E}$ is written as $\mu = \sum_k w_k\eta_k$, we say that the representation is in *minimal form* if the measures η_k are pairwise distinct.

Proposition 3.7. *Let $\mu\in\mathsf{GM}(\mathcal{E})$ be a Gaussian mixture on a vector bundle $\mathcal{E}$ with minimal form representations*

$$\sum_{k=1}^{K} w_k\eta_k \quad \textit{and} \quad \sum_{k=1}^{K'} w_k'\eta_k'.$$

Then $K=K'$, and there exists a permutation σ of $\{1,\dots,K\}$ such that $w_k = w'_{\sigma(k)}$ and $\eta_k = \eta'_{\sigma(k)}$ for all k.

We also have the following immediate corollary, which will be useful later on.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Corollary 3.8. *Let $\mu \in \mathsf{GM}(\mathcal{E})$ with not necessarily minimal form representations $\sum_{j=1}^{K} w_j \eta_j$ and $\sum_{k=1}^{K'} w'_k \eta'_k$. Then for every $j \in \{1,\dots,K\}$, there exists $k \in \{1,\dots,K'\}$ such that $\eta'_k = \eta_j$.*

Remark 3.9. We emphasize that the context of the identifiability in Proposition 3.7 is specifically for Gaussian mixtures on vector bundles, as defined in our formalism. In applications, it is natural that the mixtures in question arise from (empirical) distributions defined over the manifold itself, which are lifted to distributions on an appropriate vector bundle (typically the tangent bundle)—see the discussion in the introduction as well as the numerical experiments in section 4. Our identifiability result applies to the lifted data rather than to the manifold distributions. The question of identifiability of Gaussian mixtures defined on the manifold is much more subtle, particularly for manifolds with finite injectivity radii. Its amenability to a clean mathematical framework was a main motivation for our vector bundle formalism, but the extent to which our identifiability result is valid for distributions over manifolds remains an interesting direction of future research.

Gaussian mixtures on trivial bundles. From now on, we restrict our attention to the especially simple case of *trivial* vector bundles; we recall the definition here. Vector bundles $\mathcal{E} \to \mathcal{M}$ and $\mathcal{E}' \to \mathcal{M}$ over the same base space are called *isomorphic* if there exists a diffeomorphism $\varphi : \mathcal{E} \to \mathcal{E}'$ such that the diagram

$$\begin{tikzcd}
\mathcal{E} \arrow[rr, "\varphi"] \arrow[dr] & & \mathcal{E}' \arrow[dl] \\
& \mathcal{M} &
\end{tikzcd}$$

commutes (i.e., such that the two possible maps from $\mathcal{E}$ to $\mathcal{M}$ agree), where the diagonal arrows are vector bundle projections, and such that the induced maps on fibers $\varphi_m := \varphi|_{\mathcal{E}_m}$ are linear isomorphisms $\mathcal{E}_m \to \mathcal{E}'_m$ for each $m \in \mathcal{M}$. The map φ is called a *bundle isomorphism.* If $\mathcal{E}$ and $\mathcal{E}'$ are both endowed with smoothly varying inner products and each φ_m is an isometry of inner product spaces, then we say that φ is a *bundle isometry.* We now consider rank-d vector bundles $\mathcal{E} \to \mathcal{M}$, which are isomorphic to the product bundle $\mathcal{M} \times \mathbb{R}^d$; such bundles are called *trivial.* In this case, an isomorphism $\varphi : \mathcal{E} \to \mathcal{M} \times \mathbb{R}^d$ is called a *trivialization* of $\mathcal{E}$. We will use the following basic result, which says that we can assume without loss of generality that trivializations are bundle isometries with respect to the standard structure on $\mathcal{M} \times \mathbb{R}^d$. Some examples of trivializations are later provided in section 3.3.

Proposition 3.10. *If $\mathcal{E} \to \mathcal{M}$ is a rank-d trivial bundle endowed with a smoothly varying inner product $\{\langle \cdot, \cdot \rangle_m\}_{m \in \mathcal{M}}$, we can choose a trivialization which is a bundle isometry with respect to the standard inner product on $\mathbb{R}^d \approx \{m\} \times \mathbb{R}^d$.*

We justify our interest in the class of trivial vector bundles by the following remarks, elaborating on the discussion in the introduction.

Remark 3.11.

1. From an applications-oriented perspective, we are especially interested in *parallelizable Riemannian manifolds*, that is, the case where $\mathcal{M}$ is a Riemannian manifold and the trivial vector bundle $\mathcal{E}$ is the tangent bundle $T\mathcal{M}$. For example, any orientable three-dimensional manifold is parallelizable [3], as is any Lie group [26].

   Although we frequently work with manifolds $\mathcal{M}$ which are not parallelizable (e.g.,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

spheres $\mathbb{S}^n$ with $n \notin \{0,1,3,7\}$—see [6]), it is typically the case in realistic data analysis applications that manifold-valued data lie in a subset $\widetilde{\mathcal{M}} \subset \mathcal{M}$ which is parallelizable, so that we may assume parallelizability without loss of generality. In the setting of the sphere $M = \mathbb{S}^n$, any proper open subset $\widetilde{\mathcal{M}} \subset M$ is parallelizable—indeed, a proper open subset must miss a point $p \in \mathcal{M}$; $\widetilde{\mathcal{M}} = \mathcal{M} \setminus \{p\}$ is parallelizable, and any open submanifold of a parallelizable manifold is also parallelizable.

Thus, it frequently suffices to consider trivial vector bundles when focused on applications.

2. In [27], a Wasserstein distance is defined between Gaussian measures on the tangent bundle $T\mathcal{M} \to \mathcal{M}$ of a Riemannian manifold $\mathcal{M}$. This is done with respect to an extended metric (i.e., a metric-like function which is allowed to take the value ∞) on $T\mathcal{M}$, which is a true (finite) metric only in the case that $\mathcal{M}$ has empty cut locus. This implies that $\mathcal{M}$ is diffeomorphic to Euclidean space and, in particular, that $T\mathcal{M}$ is a trivial vector bundle. Our setting is therefore a generalization in a formal sense, as it includes tangent bundles where the base manifold can have nontrivial topology (e.g., $\mathbb{S}^1$). In practice, the present setting and that of [27] are equivalent in the case of Gaussian measures on tangent bundles, but we further generalize to the novel case of Gaussian mixtures.
3. As shown in Proposition 3.6, Gaussian mixtures on $\mathbb{R}^d$ (in the classical sense) correspond to Gaussian mixtures on the trivial vector bundle $\mathbb{R}^d \times \mathbb{R}^d$. The setting of trivial bundles is therefore a natural generalization of the classical setting.

The next result shows that the property of a measure being a Gaussian mixture is well-defined on bundle isometry classes.

Proposition 3.12. *Let $\mathcal{E}, \mathcal{E}' \to \mathcal{M}$ be vector bundles endowed with inner products. If $\varphi : \mathcal{E} \to \mathcal{E}'$ is a bundle isometry, then a measure μ on $\mathcal{E}$ is a Gaussian mixture if and only if its pushforward $\varphi_\# \mu$ is a Gaussian mixture on $\mathcal{E}'$.*

Explicitly, if $\mu = \sum_{k=1}^K w_k \eta_k$ with $\eta_k = N_{\mathcal{E}}(m_k, \Sigma_k)$, then $\varphi_\# \mu = \sum_{k=1}^K w_k \varphi_\# \eta_k$, where $\varphi_\# \eta_k = N_{\mathcal{E}'}(m_k, \varphi_{m_k} \Sigma_k \varphi_{m_k}^{-1})$.

3.2. Distances between Gaussian mixtures on trivial bundles.

We now extend Theorem 2.6 to the setting of Gaussian mixtures on trivial vector bundles.

Distance between Gaussians on trivial bundles. Next, we characterize distances between Gaussian measures on a trivial bundle with respect to a family of metrics. In the following, suppose that we have some fixed metric $d_{\mathcal{M}}$ on our base manifold $\mathcal{M}$ (e.g., geodesic distance with respect to a Riemannian metric).

For a product bundle $\mathcal{M} \times \mathbb{R}^d \to \mathcal{M}$, let $d_{\mathcal{M}\times\mathbb{R}^d} : (\mathcal{M} \times \mathbb{R}^d) \times (\mathcal{M} \times \mathbb{R}^d) \to \mathbb{R}$ be the ℓ_2-product metric with respect to $d_{\mathcal{M}}$ and Euclidean distance, that is,

$$d_{\mathcal{M}\times\mathbb{R}^d}((m_0, v_0), (m_1, v_1))^2 = d_{\mathcal{M}}(m_0, m_1)^2 + \|v_0 - v_1\|^2.$$

Let $W_2^{\mathcal{M}\times\mathbb{R}^d}$ denote the associated 2-Wasserstein distance. We will use the following function, which is clearly a smooth bijection.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Definition 3.13. *The* coordinate permutation map *is*

$$\sigma : (\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d) \to (\mathcal{M}\times\mathbb{R}^d)\times(\mathcal{M}\times\mathbb{R}^d)$$
$$((p_0,p_1),(v_0,v_1)) \mapsto ((p_0,v_0),(p_1,v_1)).$$

Lemma 3.14. *Let $\eta_i = N_{\mathcal{M}\times\mathbb{R}^d}(m_i,\Sigma_i)$ be Gaussian measures on the product bundle $\mathcal{M}\times\mathbb{R}^d$ for $i\in\{0,1\}$. Then*

$$W_2^{\mathcal{M}\times\mathbb{R}^d}(\eta_0,\eta_1)^2 = d_{\mathcal{M}}(m_0,m_1)^2 + \operatorname{tr}\left(\Sigma_0+\Sigma_1-2\left(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right).$$

Moreover, there is an optimal coupling of the form $\sigma_{\#}\overline{\pi}$ for some $\overline{\pi}\in \mathsf{G}((\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d))$, where σ is the coordinate permutation map.

This result extends to the setting of a general trivial bundle (not just a product bundle) $\mathcal{E}\to\mathcal{M}$. Let $\varphi:\mathcal{E}\to\mathcal{M}\times\mathbb{R}^d$ be a trivialization, and define a metric d_φ on $\mathcal{E}$ as the pullback of the product metric, that is,

$$d_\varphi(q_0,q_1) := \varphi^* d_{\mathcal{M}\times\mathbb{R}^d}(q_0,q_1) = d_{\mathcal{M}\times\mathbb{R}^d}(\varphi(q_0),\varphi(q_1)).$$

Let W_2^φ denote the Wasserstein distance associated to this metric.

Proposition 3.15. *Let $\mathcal{E}\to\mathcal{M}$ be a trivial bundle with φ and W_2^φ as above, and let $\eta_i = N_{\mathcal{E}}(m_i,\Sigma_i)$ for $i\in\{0,1\}$. Then*

$$W_2^\varphi(\eta_0,\eta_1)^2 = d_{\mathcal{M}}(m_0,m_1)^2 + \operatorname{tr}\left(\Sigma_0+\Sigma_1-2\left(\Sigma_0^{\frac{1}{2}}\Phi_{m_0,m_1}^{-1}\Sigma_1\Phi_{m_0,m_1}\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right),$$

where

$$\Phi_{m_0,m_1} := \varphi_{m_1}^{-1}\circ\varphi_{m_0} : \mathcal{E}_{m_0}\to\mathcal{E}_{m_1}.$$

Distance between Gaussian mixtures on trivial bundles. Finally, we extend the results above to the setting of Gaussian mixture measures on trivial bundles.

We begin with the product bundle $\mathcal{M}\times\mathbb{R}^d$, with metric $d_{\mathcal{M}\times\mathbb{R}^d}$ and associated Wasserstein metric $W_2^{\mathcal{M}\times\mathbb{R}^d}$ defined as above. Let $\mu_i = \sum_{k=1}^{K_i} w_i^k\eta_i^k$, $i\in\{0,1\}$ be elements of $\mathsf{GM}(\mathcal{M}\times\mathbb{R}^d)$. As in the classical setting of Gaussian mixtures on $\mathbb{R}^d$, we can consider alternative metrics on the space of Gaussian mixtures. First, let $W_2^{\mathsf{G}(\mathcal{M}\times\mathbb{R}^d)}(\mu_0^*,\mu_1^*)$ denote the Wasserstein distances between the measures when they are considered as discrete distributions on the space of Gaussians $\mathsf{G}(\mathcal{M}\times\mathbb{R}^d)$; as in the Euclidean setting, we write

$$\mu_i^* = \sum_{k=1}^{K_i} w_i^k \delta_{\eta_i^k} \in \mathcal{P}_2(\mathsf{G}(\mathcal{M}\times\mathbb{R}^d)).$$

Second, consider the *mixture Wasserstein distance*

$$\begin{aligned} &MW_2^{\mathcal{M}\times\mathbb{R}^d}(\mu_0,\mu_1)^2 \\ &\quad := \inf_{\pi\in\Pi(\mu_0,\mu_1)\cap\mathsf{GM}_\sigma(\mathcal{M}\times\mathbb{R}^d)} \int_{(\mathcal{M}\times\mathbb{R}^d)\times(\mathcal{M}\times\mathbb{R}^d)} d_{\mathcal{M}\times\mathbb{R}^d}((p_0,v_0),(p_1,v_1))^2 d\pi((p_0,v_0),(p_1,v_1)), \end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

where

$$\mathsf{GM}_\sigma(\mathcal{M}\times\mathbb{R}^d) := \{\pi = \sigma_\#\overline{\pi} \,|\, \overline{\pi}\in\mathsf{GM}((\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d))\} \subset \mathcal{P}_2((\mathcal{M}\times\mathbb{R}^d)\times(\mathcal{M}\times\mathbb{R}^d)),$$

with σ denoting the coordinate permutation map from Definition 3.13.

We have the following generalization of Theorem 2.6 (Theorem 2.6 is recovered by setting $\mathcal{M}=\mathbb{R}^d$ and $d_\mathcal{M}$ to be Euclidean distance), whose proof follows similar ideas as the proof in [12].

Lemma 3.16. *Let $\mu_0,\mu_1\in\mathsf{GM}(\mathcal{M}\times\mathbb{R}^d)$, with $\mu_i=\sum_{k=1}^{K_i} w_i^k\eta_i^k$ and $\eta_i^k = N_{\mathcal{M}\times\mathbb{R}^d}(m_i^k,\Sigma_i^k)$. Then*

$$MW_2^{\mathcal{M}\times\mathbb{R}^d}(\mu_0,\mu_1) = W_2^{\mathsf{G}(\mathcal{M}\times\mathbb{R}^d)}(\mu_0^*,\mu_1^*).$$

Now consider an arbitrary trivial vector bundle $\mathcal{E}\to\mathcal{M}$. For a trivialization $\varphi:\mathcal{E}\to\mathcal{M}\times\mathbb{R}^d$, define the *associated mixture Wasserstein distance* between Gaussian mixtures $\mu_0,\mu_1\in\mathsf{GM}(\mathcal{E})$ as

$$MW_2^\varphi(\mu_0,\mu_1)^2 := \inf_{\pi\in\Pi(\mu_0,\mu_1)\cap\mathsf{GM}_\sigma^\varphi(\mathcal{E})} \int_{\mathcal{E}\times\mathcal{E}} d_\mathcal{E}(w_0,w_1)^2 d\pi(w_0,w_1), \tag{3.1}$$

where

$$\mathsf{GM}_\sigma^\varphi(\mathcal{E}) := \{(\varphi^{-1}\circ\sigma)_\#\overline{\pi} \,|\, \overline{\pi}\in\mathsf{GM}((\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d))\}.$$

Theorem 3.17. *With the notation above, $MW_2^\varphi(\mu_0,\mu_1) = W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*)$, where $\mathsf{G}(\mathcal{E})$ is considered as a metric space with (the restriction of) the Wasserstein distance W_2^φ.*

We immediately obtain the following corollary from Theorems 2.6 and 3.17 together with Proposition 3.6.

Corollary 3.18. *The mixture Wasserstein distance MW_2^φ defines a metric on $\mathsf{GM}(\mathcal{E})$. When $\mathcal{E}=\mathbb{R}^d\times\mathbb{R}^d$ and the trivialization φ is the identity map, MW_2^φ agrees with the mixture Wasserstein distance $MW_2^{\mathbb{R}^d}$ of [12].*

Next, we bound the mixture Wasserstein distance in terms of the (unconstrained) Wasserstein distance. This is an extension of [12, Proposition 6], but we give a different proof which yields a tighter upper bound.

Proposition 3.19. *Let $\mathcal{E}\to\mathcal{M}$ be a trivial vector bundle with trivialization φ, and let $\mu_0,\mu_1\in\mathsf{GM}(\mathcal{E})$ with $\mu_i=\sum_{k=1}^{K_i} w_i^k\eta_i^k$, $\eta_i^k=N_\mathcal{E}(m_i^k,\Sigma_i^k)$, presented in minimal form. Then*

$$W_2^\varphi(\mu_0,\mu_1)\le MW_2^\varphi(\mu_0,\mu_1)\le W_2^\varphi(\mu_0,\mu_1)+\sum_{i=0,1}\left(\sum_{k=1}^{K_i} w_i^k\operatorname{tr}(\Sigma_i^k)\right)^{1/2}.$$

Summary of the mixture Wasserstein distance. For the convenience of the reader, we summarize the notation and constructions of the previous subsections here. The mixture Wasserstein

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

distance for a trivial vector bundle $\mathcal{E} \to \mathcal{M}$ with trivialization $\varphi : \mathcal{E} \to \mathcal{M} \times \mathbb{R}^d$, MW_2^φ, between Gaussian mixtures

$$\mu_i = \sum_{k=1}^{K_i} w_i^k \eta_i^k \in \mathsf{GM}(\mathcal{E}), \quad \eta_i^k = N_{\mathcal{E}}(m_i^k, \Sigma_i^k), \quad i \in \{0,1\},$$

can be expressed as

$$MW_2^\varphi(\mu_0,\mu_1)^2 = \min_{\omega \in \Pi(\mu_0,\mu_1)} \sum_{k,\ell} \omega_{k\ell} W_2^\varphi(\eta_0^k, \eta_1^\ell)^2, \tag{3.2}$$

where

$$\begin{aligned} &W_2^\varphi(N_{\mathcal{E}}(m_0,\Sigma_0), N_{\mathcal{E}}(m_1,\Sigma_1))^2 \\ &\quad = \sum_{k,\ell} d_{\mathcal{M}}(m_0,m_1)^2 + \operatorname{tr}\left(\Sigma_0 + \Sigma_1 - 2\left(\Sigma_0^{\frac{1}{2}} \Phi_{m_0,m_1}^{-1} \Sigma_1 \Phi_{m_0,m_1} \Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \end{aligned}$$

and

$$\Phi_{m_0,m_1} = \varphi_{m_1}^{-1} \circ \varphi_{m_0} : \mathcal{E}_{m_0} \to \mathcal{E}_{m_1}.$$

3.3. Choosing trivializations.

In this subsection, we address the question of how to choose an appropriate trivialization to fit into the pipeline described in the preceding subsection. We focus on the case that the vector bundle is of the form $T\mathcal{M} \to \mathcal{M}$, where $\mathcal{M}$ is a parallelizable Riemannian manifold. We will make the simplifying assumption that the manifold is endowed with a distinguished point $p \in \mathcal{M}$ and a distinguished isometry $P_m : T_p\mathcal{M} \to T_m\mathcal{M}$ for every point $m \in \mathcal{M}$ such that P_p is the identity map and such that the isometries vary smoothly in m.

Example 3.20. We now provide some examples of where the data described above might arise:

1. In the numerical experiments in the following section, the underlying manifold $\mathcal{M}$ is a *punctured sphere*—that is, a sphere $\mathbb{S}^n$ with a point removed. In this case, the distinguished point p is taken to be antipodal to the removed point. For every $m \in \mathcal{M}$, there is a unique geodesic joining p to m, and the isometry $P_m : T_p\mathcal{M} \to T_m\mathcal{M}$ is given by parallel transporting along the geodesic. In practice, one begins with a Gaussian mixture on the sphere $\mathbb{S}^n$ and chooses a distinguished point $p \in \mathbb{S}^n$, say, by always taking the north pole or by choosing a Fréchet mean of the basepoints of the Gaussians. Generically, the antipodal point $-p$ will not be a basepoint of any Gaussian in the mixture, so that we can restrict our analysis to the setting of the punctured sphere $\mathcal{M} = \mathbb{S}^n \setminus \{-p\}$.
2. More generally, if there is some point $p \in \mathcal{M}$ such that there is a unique geodesic to any other point $m \in \mathcal{M}$, then the parallel transport approach described above defines the desired collection of isometries. This is the setting used in [27] for comparing wrapped Gaussians.
3. If $\mathcal{M}$ is a Lie group with a left- (or right)-invariant Riemannian metric, then the natural choice of basepoint is the identity $e \in \mathcal{M}$, and the isometry P_m is the derivative of the left (or right) multiplication map.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Given $p \in \mathcal{M}$ and $\{P_m\}_{m\in\mathcal{M}}$ as above, we define a bundle isometry $T\mathcal{M} \to \mathcal{M} \times \mathbb{R}^d$ as follows. Choosing any orthonormal basis $F = (f_1,\dots,f_d)$ for $T_p\mathcal{M}$, one can extend this to a global frame for $\mathcal{M}_p$ by defining $F(m) = \{f_j(m) = P_m(f_j), j = 1,\dots,d\}$. Let $\varphi^{p,F} : T\mathcal{M} \to \mathcal{M} \times \mathbb{R}^d$ be defined by

$$\varphi^{p,F}(v) = \Big(m, (\langle v, f_j(m)\rangle_m)_{j=1}^d\Big) \qquad \text{for } v \in T_m\mathcal{M}.$$

We now observe that, while this construction depends on choices of p, $\{P_m\}_{m\in\mathcal{M}}$ and F, the induced distance $MW_2^{\varphi^{p,F}}$ is actually invariant under the choice of F. Since p and $\{P_m\}_{m\in\mathcal{M}}$ arise naturally in several examples of interest (Example 3.20), the following result shows that the mixture Wasserstein distance is a natural tool in practice.

Proposition 3.21. *With the notation defined above, the mixture Wasserstein distance $MW_2^{\varphi^{p,F}}$ does not depend on the choice of orthonormal basis F. That is, for any orthonormal bases $F = (f_1,\dots,f_d)$ and $G = (g_1,\dots,g_d)$ for $T_p\mathcal{M}$, we have $MW_2^{\varphi^{p,F}}(\mu_0,\mu_1) = MW_2^{\varphi^{p,G}}(\mu_0,\mu_1)$ for any Gaussian mixtures μ_0,μ_1.*

Remark 3.22. The result and its proof show that, in the situation described in this subsection, the part of $W_2^{\mathsf{G}(\mathcal{E})}$ that involves comparing Gaussians on $\mathcal{E}$ can be computed by moving each Gaussian to the basepoint; that is, $\eta = N_{\mathcal{E}}(m,\Sigma)$ can be replaced with $(P_m^{-1})_{\#}\eta$. Independence of choice of frame in constructing the trivialization then just amounts to the fact that the Wasserstein distance on a Euclidean space is invariant under isometries.

4. Experimental studies.

In this section, we demonstrate applications of the proposed Wasserstein-type distances on some nonlinear domains. We first introduce a general setup and present procedures for estimating Gaussian mixture parameters from sample data. Then we move to examples involving simulated data on a unit sphere and the Kendall shape space of planar triangles. Finally, we present an application on real data in the preshape space of planar closed shapes. We note that the Python package Geomstats (see [32]) was used for geometric computations in these experiments.

4.1. Basic experimental setup.

Let $\mathcal{M}$ be a d-dimensional parallelizable Riemannian manifold. One can use the following procedure to define a global frame on $\mathcal{M}$: (1) choose a reference point $p \in \mathcal{M}$, (2) randomly generate d linearly independent tangent vectors in $T_p\mathcal{M}$, and (3) calculate a principal component analysis of those vectors, and take the principal directions as an orthonormal basis F. Given a reference point p and a reference frame F, we calculate an element of the global frame $F(m)$ by parallel transporting the tangent vectors in F to $T_m(\mathcal{M})$ along the unique geodesic from p to m in $\mathcal{M}$.

The next issue is the estimation of Gaussian mixture parameters from given samples on $\mathcal{M}$. As mentioned earlier, there exist computational solutions for estimating mean, covariance, and weight parameters (see, for example, [21]) for mixture components, but they can become costly, especially on nonlinear domains. Computing a Gaussian mean on manifolds is already an iterative procedure, and when combined with an outer loop over mixture components, it becomes prohibitive. Instead of EM-type solutions, we make use of clustering-based approaches, such as Riemannian K-means (see [32]) and K-modes kernel mixtures clustering (see [14]). Note that the Riemannian K-means can be computationally expensive, and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

we restrict its use to low-dimensional examples. For estimation on high-dimensional shape manifolds, we utilize the K-mode clustering method.

Riemannian K-means is an extension of the well-known K-means algorithm to nonlinear manifolds. Given a data set $X = \{x_i \in \mathcal{M},\ i = 1, \dots, n\}$, the K-means algorithm returns a cluster index $\mathcal{I} = \{\ell_i \in (1, \dots, K), i = 1, \dots, n\}$ representing the cluster assignments for each x_i. In order to obtain estimates for Gaussian mixture parameters, we treat the clusters found by K-means as samples from Gaussian mixture components. Let $\hat{m}^k$ denote the Fréchet mean [18] for the sample points in cluster k, and let n^k denote the number of sample points in cluster k. We estimate the weight for component k as $\hat{w}^k = n^k/n$ and calculate the tangent space covariance in the coordinates of the global frame $F(m) = \{f_j(m), j = 1, \dots, d\}$ as $\hat{\Sigma}^k = \frac{1}{n^k-1} V^t V$, where $V_{j,i} = \langle \exp^{-1}{}_{\hat{m}^k}(x_i), f_j(\hat{m}^k)\rangle$, where $\exp^{-1}$ is the inverse exponential map on $\mathcal{M}$. Our estimate of the Gaussian mixture is then $\hat{\mu} = \Sigma_{k=1}^K \hat{w}^k \hat{\eta}^k$, where $\hat{\eta}^k = N_{\mathcal{E}}(\hat{m}^k, \hat{\Sigma}^k)$.

In the second approach for estimating model parameters, we use a K-modal kernel-mixture clustering algorithm presented in [14, 13]. This metric-based nonparametric procedure takes in the matrix of pairwise distances between all data points and uses it to cluster individual points around estimated modes. Modes are significant local maxima of the underlying probability distribution and are computed as the points with the most neighbors. An important strength of this approach is that neighborhoods are determined automatically from the data, avoiding the need for any manual selection of hyperparameters. The output of this procedure are (1) cluster labels for most of the input points, (2) points that are modes of each cluster, and (3) outliers that do not belong to any cluster. We treat each cluster as a mixture component, with its mode as the estimated mean $\hat{m}^k$. Furthermore, we compute a covariance matrix $\hat{\Sigma}^k$ in the tangent space of this estimated mean with respect to the global frame. These steps are the same as in the previous item and result in a Gaussian mixture $\hat{\mu}$.

Given estimates of Gaussian mixtures $\hat{\mu}_i = \sum_{k=1}^{K_i} \hat{w}_i^k \hat{\eta}_i^k$, $i = 0, 1$, parameterized with respect to a global frame (p, F), we calculate the Wasserstein-type distance between them by using linear programming to solve

$$MW_2^{\varphi_{p,F}}(\hat{\mu}_0, \hat{\mu}_1)^2 = \min_{w \in \Pi(\hat{w}_0, \hat{w}_1)} \sum_{k,\ell}^{K_0,K_1} w_{k\ell} W_2^{\varphi}(\hat{\eta}_0^k, \hat{\eta}_1^\ell)^2. \tag{4.1}$$

In the following, we take specific examples of $\mathcal{M}$ and demonstrate some applications of MW_2.

4.2. Simulated data on punctured 2-sphere.

As the first example, we consider the domain $\mathbb{S}^2$ and simulate several Gaussian mixtures on $\mathbb{S}^2$ to demonstrate the MW metric. In this experiment, we select either $p = [0, 0, 1]^T$ or $-p = [0, 0, -1]^T$ as a reference point, and we set $F = [[-1, 0, 0]^T, [0, -1, 0]^T] \in \mathbb{R}^{3\times 2}$ as a reference frame—we then assume that Gaussian mixtures have no component in the tangent space of $-p$ (resp., in the tangent space of p). Figure 1 (left) shows the reference frame transported to several points on the $\mathbb{S}^2$. The right side shows a similar plot when the reference point is $[0, 0, -1]$ instead, illustrating the dependence of the global frame on the basepoint.

To generate sample data X from a Gaussian with mean $m \in \mathbb{S}^2$ and an associated covariance Σ, we use the following steps. Recall that the global frame at m is given by $F(m)$. We generate a set $V = \{v_i \sim N(0, \Sigma), \Sigma \in \mathsf{Sym}_d^+, i = 1, \dots, N\}$. Then we compute

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

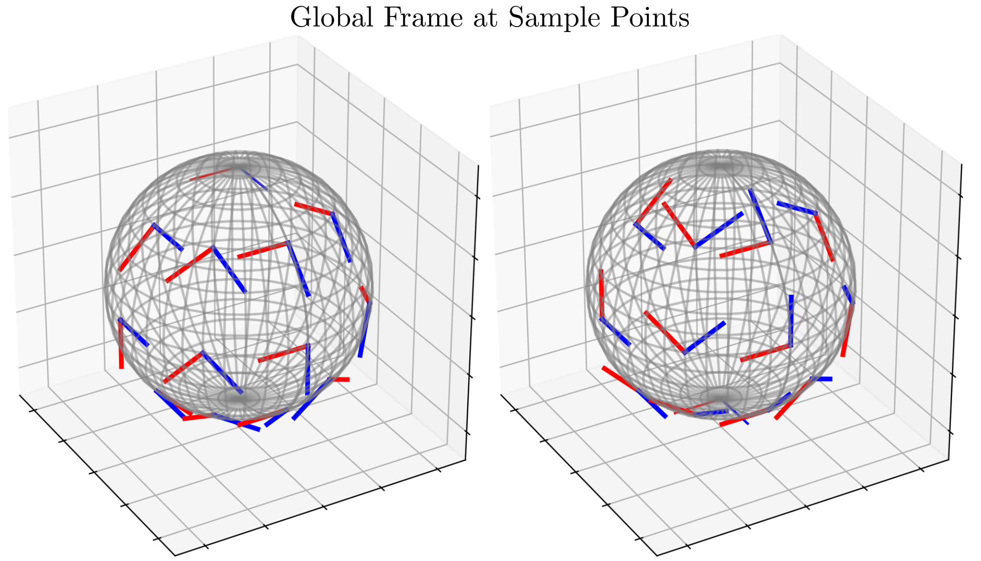


Figure 1. *Examples of two reference frames transported to a select number of points on the punctured sphere, providing a consistent coordinate system over the entire manifold. Left: global frames transported from reference pointx* $[0,0,1]$. *Right: global frames transported from reference point* $[0,0,-1]$.

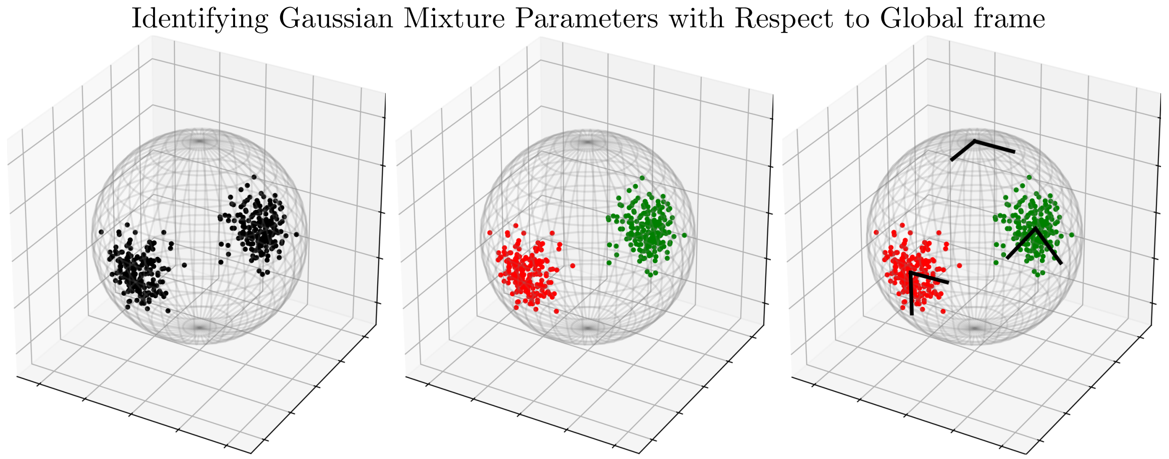

Identifying Gaussian Mixture Parameters with Respect to Global frame

Figure 2. *Steps for estimating Gaussian mixture parameters with respect to a global frame on a punctured* $\mathbb{S}^2$. *Left: sphere with sample data. Middle: sample data colored by* K*-means cluster assignment. Right: global frame transported to Fréchet means of clusters.*

$X_i = \exp_m((\langle v_i, f_j(m)\rangle_m)_{j=1}^d)$ to obtain samples from this Gaussian. Further, to generate samples from a Gaussian mixture, we just generate samples from individual Gaussians with frequencies proportional to their weights. Given a set of points $\{X_i \in \mathbb{S}^2\}$, represented in terms of their extrinsic coordinates $\{X_i \in \mathbb{R}^3\}$, one can fit a Gaussian mixture model to those points using the Riemannian K-means approach described above. Figure 2 illustrates this process for a chosen global frame. The left panel shows samples from a Gaussian mixture with two components, the middle panel shows a data clustering using colors, and the right

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

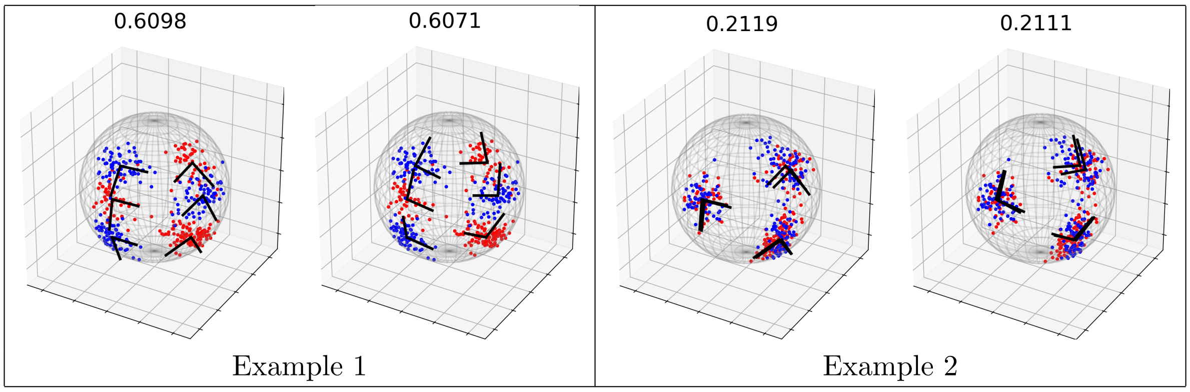


Figure 3. *Examples of Wasserstein-type distances between Gaussian mixtures on the tangent bundle of a punctured* $\mathbb{S}^2$ *computed using two different global frames. Color denotes a Gaussian mixture; plot titles show calculated distances. Example* 1*: samples from two Gaussian mixtures with different means, covariances, and weights. Example* 2*: samples from two Gaussian mixtures with the same means and covariances but different weights.*

panel shows a global frame at p transported to the Fréchet means of the clusters. We use these global frames to express the tangent-space covariance of the data in each cluster, as discussed in section 4.1. The relative weights of the components $\{w^k\}$ are estimated using the relative frequencies. Letting $\hat{\eta}^k = N_{\mathcal{E}}(\hat{m}^k, \hat{\Sigma}^k)$, the resulting Gaussian mixture is denoted by $\hat{\mu} = \Sigma_{k=1}^K \hat{w}^k \hat{\eta}^k$.

Given two such estimated Gaussian mixtures $\hat{\mu}_i$, $i = 0, 1$, parameterized with respect to a global frame (p, F), we calculate the Wasserstein-type distance between them using linear programming to solve (4.1). Figure 3 shows several examples of evaluations of the Wasserstein distance between Gaussian mixtures estimated from the same sets of data but with different global frames. In particular, the two examples use the reference points $[0,0,1]^T$ and $[0,0,-1]^T$, respectively. The titles of the plots contain the Wasserstein-type distance calculated with respect to the global frame, and color represents a Gaussian mixture. Example 1 shows a case where the Gaussian mixtures differ significantly in terms of means, covariances, and weights. Consequently, the Wasserstein-type distance between them is large: 0.6098. If we use a different reference point p, the distance can potentially change, as illustrated by the second panel, where we obtain a value of 0.6071. In example 2, the two Gaussian mixtures are relatively similar: They have the same means and covariances and differ only in weights. In this case, changing the reference point results in a small change in distance.

4.3. Comparing shape populations of triangles.

In this paper, we are interested in shape spaces of objects as the domains for imposing and comparing probability distributions. In other words, we want to compare shape populations, modeled as Gaussian mixtures, using Wasserstein-type distances. Recall that shape is a geometric property that is invariant to rotation, translation, and scaling. Before we consider shapes of planar contours, we analyze a simpler case analyzing shapes of planar triangles. The shape space of planar triangles is denoted by the quotient space $\mathbb{S}^3/SO(2)$, which can be further identified with $\mathbb{S}^2$ [24]. Hence, our analysis of triangle shapes is performed on a (punctured) $\mathbb{S}^2$. The steps of the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

computation—establishing a global frame, estimating Gaussian mixture parameters, and calculating Wasserstein-type distances between Gaussian mixtures—are similar to the previous subsection.

Now we provide details for the $\mathbb{S}^2$ representation of planar triangles. Let $\{x_i \in \mathbb{R}^2, i = 1,2,3\}$ be the set of all planar triangles. We identify x_i with elements $z_i \in \mathbb{C}$ such that $z_i = (x_{i,1} + jx_{i,2})$, $j = \sqrt{-1}$. After we remove rigid translations and global scaling, we obtain the set $\mathcal{P}_T = \{z \in \mathbb{C}^3 | \frac{1}{3}\sum_{i=1}^{3} z_i = 0, \|z\| = 1\}$. $\mathcal{P}_T$ is referred to as the *preshape space* because we have not yet removed rigid rotations. The shape space of two-dimensional triangles is thus $\mathcal{S}_T = \{[z] = \{e^{j\varphi}z | \varphi \in \mathbb{S}^1, z \in \mathcal{P}_T\}\}$. An element $[z] \in \mathcal{S}_T$ corresponds to a unique triangular shape, with the parameter φ denoting its rotation with respect to a chosen coordinate system. Any $[z] \in \mathcal{S}_T$ can be isometrically mapped to a point on $\mathbb{S}^2$ using the Hopf fibration presented in Appendix B.2. We use this mapping to estimate mixture parameters and compute Wasserstein-type distances between Gaussian mixtures in $\mathcal{S}_T$.

Similar to the previous section, we arbitrarily select a reference point $p \in \mathcal{S}_T$, use the Hopf fibration to map it to $\tilde{p} \in \mathbb{S}^2$, and generate an arbitrary basis F for the tangent space $T_{\tilde{p}}\mathbb{S}^2$. One such global frame is shown in the left panel of Figure 4, along with the triangular representations of the reference point and the tangent vectors in the reference frame, shown in the middle and left panels of Figure 4, respectively. Next, we generate random samples from Gaussian mixtures on $\mathbb{S}^2$ using the procedure outlined before. Given sample data and a global frame for $\mathbb{S}^2$, we use the Riemannian K-means algorithm described in section 4.1 to estimate Gaussian mixture parameters from the sample data. Plots of the sample data, colored by cluster assignment, are presented in the first and third plots of Figure 5. The accompanying panels show these colored points as planar triangles to visualize clustered shapes. Given parameter estimates, we can calculate the Wasserstein-type distance using (4.1).

Figure 6 presents some examples of comparing populations of planar triangles using our Wasserstein metric. The plot titles on the top state the Wasserstein-type distance calculated with respect to the chosen global frames, and the point colors (red versus blue) label the Gaussian mixtures. The two right panels display the triangle shapes of these points in these

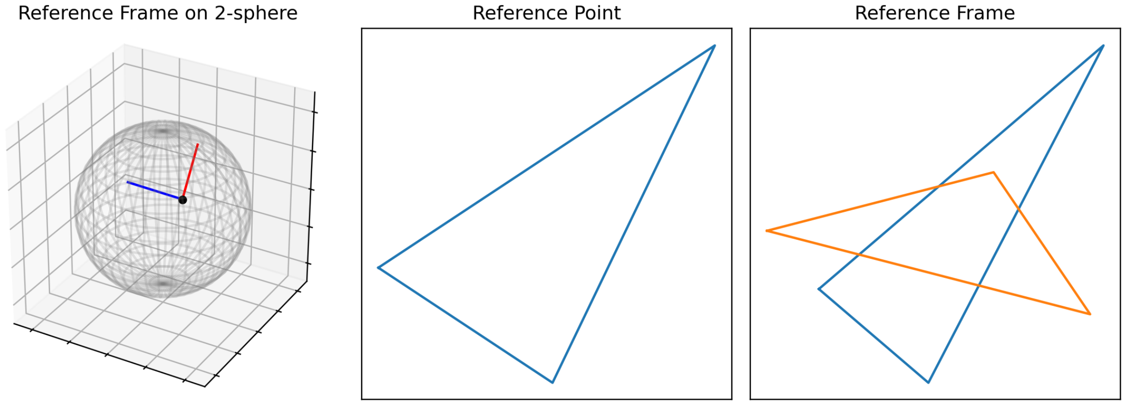


Figure 4. *A reference point and a reference frame for Kendall shape space of two-dimensional triangles. Left: representation of a global frame under the Hopf map on $\mathbb{S}^2$. Middle: triangular representation of the reference point. Right: triangular representations of tangent vectors in the reference frame.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

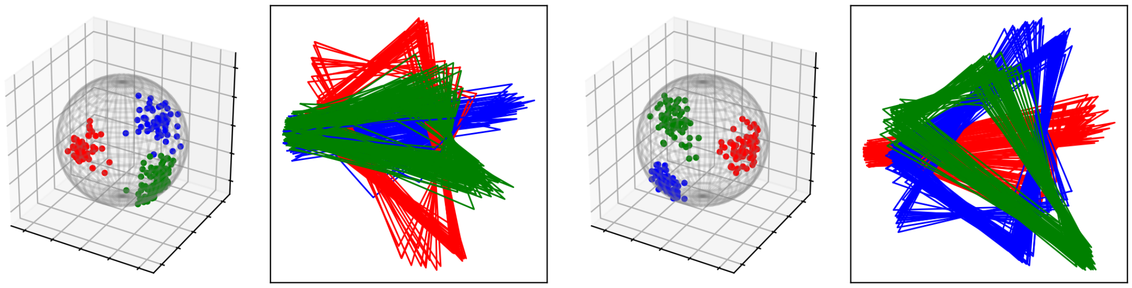


Figure 5. *Samples from two Gaussian mixtures defined on the tangent bundle of $\mathcal{S}_T$, colored by cluster assignment. The triangles in panel* 2 *(resp.,* 4*) correspond to the points on the sphere in panel* 1 (resp., 3).

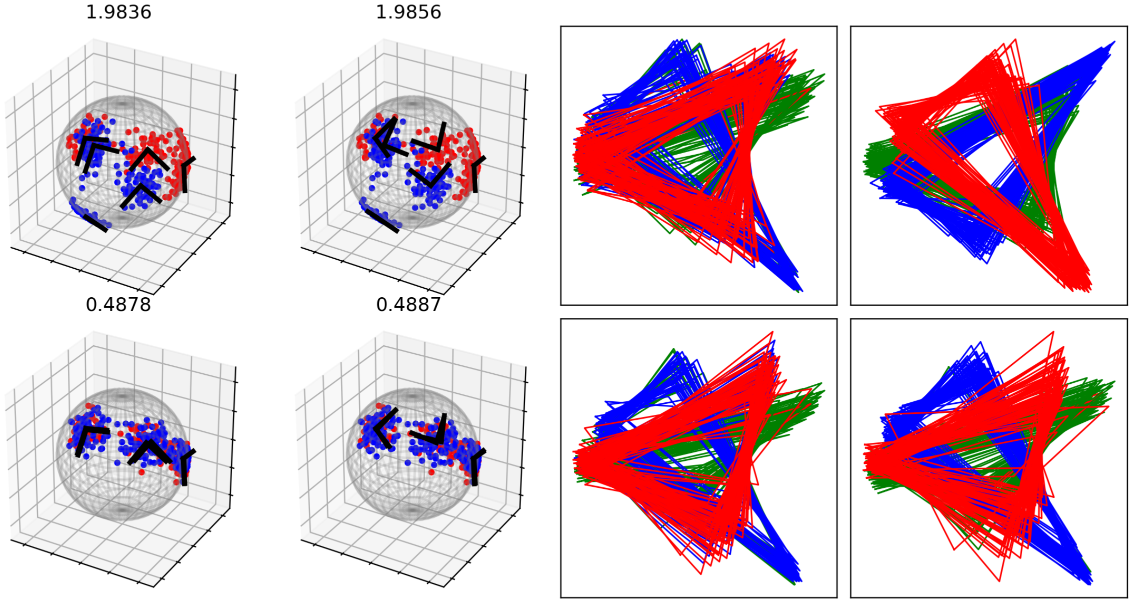


Figure 6. *Comparison of Wasserstein-type distances between Gaussian mixtures on $\mathcal{S}_T$, estimated with the same procedure from the same data, using two different global frames. Top left: Gaussian mixtures with different means, covariances, and weights. Bottom left: Gaussian mixtures with the same means and covariances but different weights. Top right: triangular representations of sample points from Gaussian mixtures for the experiment in the top row (left corresponds to blue, right corresponds to red); Bottom right: triangular representations of sample points from Gaussian mixtures for the experiment in the bottom row (left corresponds to blue, right corresponds to red).*

Gaussian mixtures. The top rows present an example where the Gaussian mixtures differ significantly in terms of means, covariances, and weights, while the example in the bottom row has two Gaussian mixtures that are relatively similar: They have the same means and covariances and differ only in weights. In the first case, the Wasserstein-type distances are relatively large and relatively stable with respect to choice of global frame. The distances are naturally smaller in the second example.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

4.4. Comparing shape populations of nanoparticles. In this section, we focus on capturing, quantifying, and comparing shapes of silver nanoparticles observed in industrial manufacturing. Silver nanoparticles are produced through a solution phase process, leveraging the radiochemistry of electron beam–induced nanoparticle growth (additional information is available in the paper [43]). Over the course of the synthesis, the shapes of these nanoparticles evolve due to chemical reactions, such as atomic addition to particles and particle merging. This solution phase process is captured using in situ transmission electron microscopy over a span of 62 seconds, with images taken at a rate of one image per second. Each image, on average, displays around 100 silver nanoparticles. The outlines of these nanoparticles are extracted using segmentation methodology presented in [41]. Each image in the video is preprocessed (including a step which involves removing particles below a certain size threshold) and segmented, returning a set of planar closed curves denoting the outlines of the individual nanoparticles in that frame. The left panel of Figure 7 shows some examples of extracted contours in imaged frames from the data set.

In nanomanufacturing, the shapes of nanoparticles are indicators of the material properties. One hypothesis is that constituent nanoparticle shapes can control the resulting material's physical properties. Thus, a vital tool is to model and quantify the particle shape populations associated with individual images and compare them across images. In any image, we treat extracted closed curves as samples from a probability distribution on a shape space, with each curve's location, orientation, and scale treated as nuisance variables. A brief introduction to the shape space $\mathcal{S}_c = \mathbb{S}^{(2T-1)}/SO(2)$ is presented in Appendix B.1. Here, each contour is represented by an array $\mathbf{q} \in \mathbb{R}^{2\times T}$ made up of equispaced points on the square-root velocity function curve of the contour. The appendix also defines a shape metric d_s and the computation of sample statistics (mean and covariance) of a set of shapes under d_s. A set of shapes rotationally aligned to their mean can be treated as points on the preshape space, the unit sphere $\mathbb{S}^{(2T-1)}$. On this unit sphere, we take the reference point $p = [1, 0, \ldots, 0]^\dagger$ and select $F = [[0, 1, 0, \ldots, 0]^\dagger, \ldots, [0, 0, \ldots, 0, 1]^\dagger]$ (where † denotes the transpose) in order to define the global frame (p, F). We estimate and compare shape distributions with respect to this global frame.

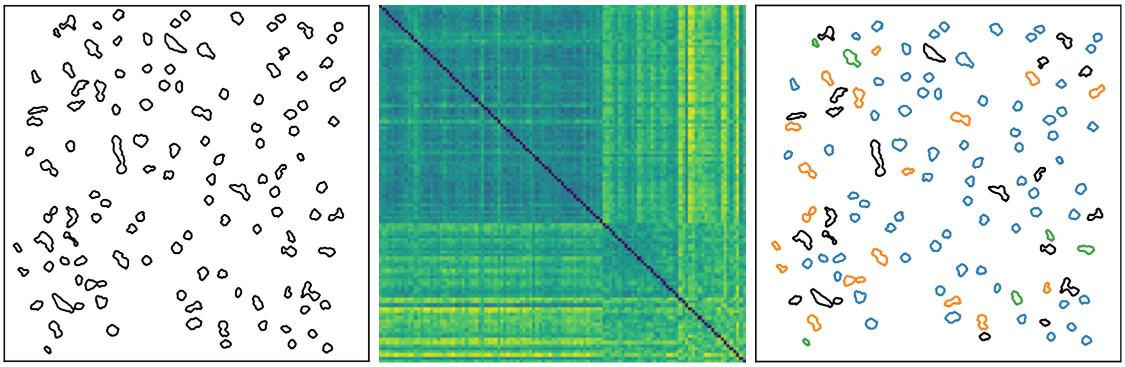


Figure 7. *K-modes clustering. Left: particles in image at $t = 38$ before clustering. Middle: pairwise shape distance matrix for all particles that image, sorted by clusters. Right: particles in image at $t = 38$ colored according to their cluster assignment.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Estimating Gaussian mixture parameters. Our approach is to model the distribution of shapes in a given video image as a mixture of Gaussians on the preshape space $\mathbb{S}^{(2T-1)}$. There are several steps that make up this approach.

Given a set of particle contours extracted from a video image, the first step is to cluster them according to their shapes. Figure 7 shows the process of applying the mode-based clustering process discussed in section 4.1. The left panel shows the video image corresponding to time 38 in the data set, prior to clustering. The middle panel shows the within-frame pairwise shape distance matrix, sorted by clusters. Green denotes smaller distances, and yellow denotes larger distances. One can see that more than half of the particles fall into the largest cluster. The algorithm automatically selects three clusters and labels the remaining particles as outliers. The right panel shows these particles colored according to their assigned clusters, in blue, orange, and green. The outliers are drawn in black.

Given a clustering of the shapes in video image t, we compute the shape mean m_t^k and tangent space covariance Σ_t^k of shapes in cluster k, as described in Appendix B.1, and estimate the Gaussian component corresponding to cluster k as $\mu_t^k = N_{\mathcal{E}}(m_t^k, \Sigma_t^k)$. Letting n_t^k be the number of shapes in cluster k and setting $n_t = \sum_k n_t^k$, we estimate component weights as $w_t^k = \frac{n_t^k}{n_t}$. Thus, for each video image, indexed by time t, we obtain a Gaussian mixture $\mu_t = \Sigma_{k=1}^{K_t} w_t^k N_{\mathcal{E}}(m_t^k, \Sigma_t^k)$.

We then calculate the pairwise Wasserstein-type distances between distributions associated with all images in the video using (4.1). The resulting 62 × 62 distance matrix is presented in the leftmost panel of Figure 8. The computational cost associated with this experiment is $O(Rn^2) + O(R^2K^3)$, where $R = 62$ is the number of images, n is the approximate number of shapes per image, and $K = 3$ is the number of clusters. The first term corresponds to the cost all pairwise distances between all shapes within images and the second to the cost of finding optimal couplings across images.

Change-point detection in time series of shape populations. Inspection of the Wasserstein-type distance matrix suggests that the shape distributions in the latter part of the process

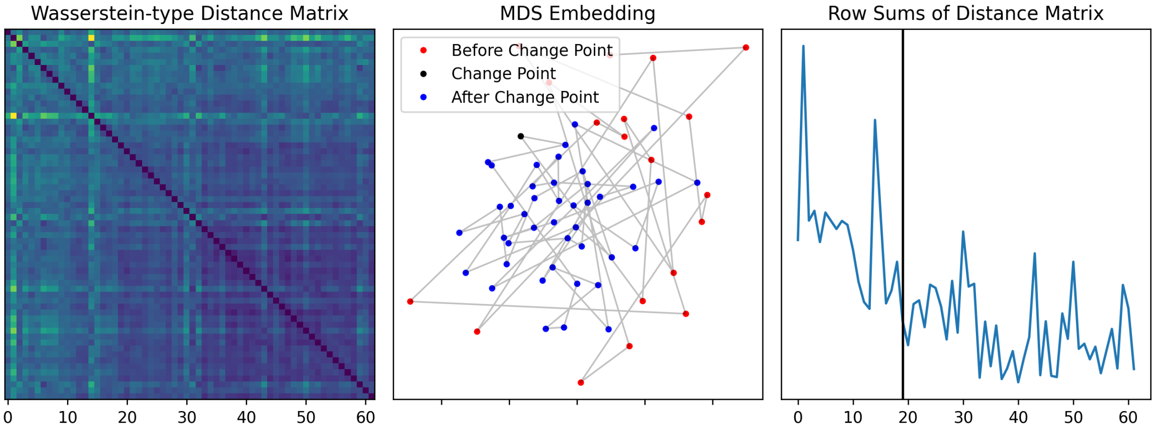


Figure 8. *Left: pairwise Wasserstein-type distance matrix between time-indexed populations, μ_t. Middle: two-dimensional multidimensional scaling (MDS) plot based on the Wasserstein-type distance matrix, colored by the relationship of t to the estimated change point. Right: sums of rows of the distance matrix plotted versus t, with the change point marked by a black line.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

appear to be closer to each other than to those in the earlier part. In order to test this statistically, we use the *E-divisive* procedure for change point detection [29]. This method is particularly well suited to our situation, as it only depends on distances between populations, requires minimal assumptions, and provides a straightforward method for testing the hypothesis of no additional change points.

The *E-divisive* algorithm is an iterative procedure where candidate change points are selected as the time point which maximizes the two-sample energy statistic [38] produced by splitting the data at that time point, and the statistical significance of the candidate change point is inferred on the basis of a permutation test based on the same two-sample energy statistic. The algorithm has several hyperparameters: (1) ρ, the number of permutations; (2) p_0, the p-value for each permutation test; (3) $\alpha \in (0,2)$, the power of distance in the test statistics; and (4) min_size, the minimum segment length to be considered for bisection. We applied the *E-divisive* algorithm to our Wasserstein-type distance matrix, with parameters $p_0 = 0.0125$, $\rho = 499$, $min\_size = 12$, and $\alpha = 1$. The algorithm found one statistically significant change point at time point $t = 19$, with a p-value of 0.002. The next candidate change point occurs at time point $t = 37$ but is rejected with a p-value of 0.034, thereby terminating the algorithm. These findings lead us to reject the hypothesis of no change points and provide support for the original observation that the distributions of shapes in the latter part of the manufacturing process differ from those in the earlier part.

Modeling shape dynamics. In addition to testing *if* a change point exists, it may also be desirable to describe *how* the distributions change over time. The optimal transport plans between shape distributions associated with successive video images provide a natural way to quantify the transitions, and the preshape representation of the means combined with the mixture assumption can make the transport plans simple to interpret. Recall that, for Gaussian mixtures $\mu_t = \Sigma_{k=1}^{K_t} w_t^k N_{\mathcal{E}}(m_t^k, \Sigma_t^k)$ and $\mu_{t+1} = \Sigma_{\ell=1}^{K_{t+1}} w_{t+1}^\ell N_{\mathcal{E}}(m_{t+1}^\ell, \Sigma_{t+1}^\ell)$ associated with images at times t and $t+1$, the optimal transport plan is given by

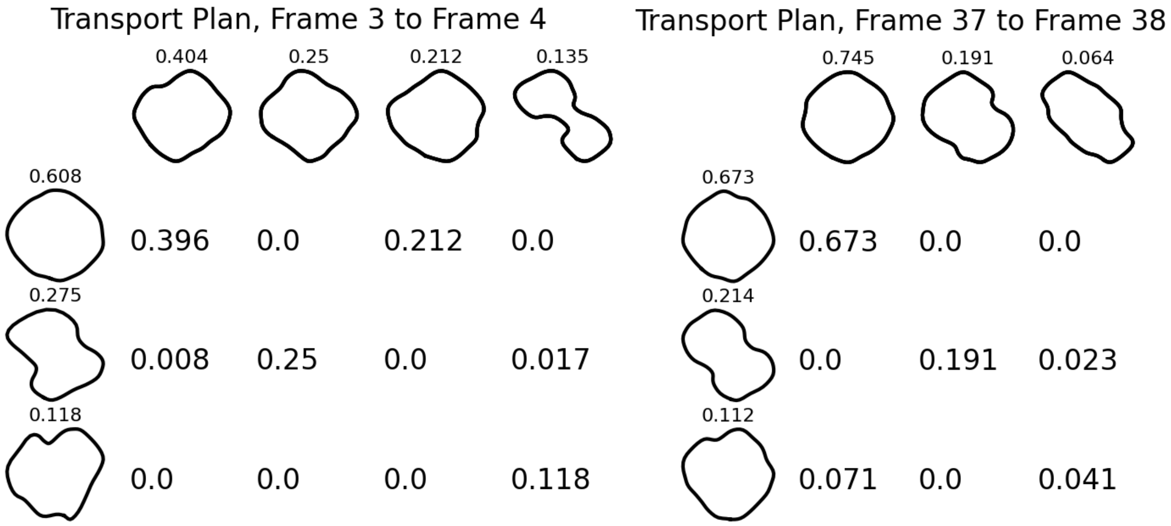
$$\pi = \arg \min_{w \in \Pi(w_0, w_1)} \sum_{k,\ell}^{K_t, K_{t+1}} w_{k\ell} W_2^{\varphi}(\mu_t^k, \mu_{t+1}^\ell)^2.$$

For example, in the left panel of Figure 9, we see the optimal transport plan π between shape distributions of associated with video images at times $t = 3$ and $t = 4$. The mean shapes for the source distribution (frame 3) are drawn on the left column and mean shapes for the target distribution (frame 4) along the top. The weights for the Gaussian mixture components are written above their corresponding mean shape. The optimal transport plan is presented in the rows/columns of the plot. This matrix shows how much mass is transported from each component in the source distribution to each component in the target distribution.

These transition matrices can be used to analyze the dynamics of nanoparticle shapes during the manufacturing process. For example, the mass in the cluster with the most circular mean in image $t = 3$ (with weight 0.608) ends up being split between two clusters in the transition to image at $t = 4$. On the other hand, the mass in the cluster with the most circular mean at $t = 37$ (with weight 0.673) is all transported to a single cluster at $t = 38$, which also gains most of the mass from another component as well. This dynamic seems to hold for the process in general; transport plans between consecutive frames show a tendency toward

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy


Transport Plan, Frame 3 to Frame 4
Transport Plan, Frame 37 to Frame 38

Transport Plan, Frame 3 to Frame 4

	0.404	0.25	0.212	0.135
0.608	0.396	0.0	0.212	0.0
0.275	0.008	0.25	0.0	0.017
0.118	0.0	0.0	0.0	0.118

Transport Plan, Frame 37 to Frame 38

	0.745	0.191	0.064
0.673	0.673	0.0	0.0
0.214	0.0	0.191	0.023
0.112	0.071	0.0	0.041

Figure 9. *Visual representation of interframe population transport plans. The means of Gaussian mixture components are displayed in the margins. Numbers above mean shapes correspond to the estimated weight for that component; the array of numbers in the center of plot are the transport plans; they represent the amount of mass transported between the corresponding marginal components.*

distributions with more mass being centered around more circular shapes, especially in the later part of the process.

5. Conclusion and discussion.

This paper develops a framework for representing and comparing populations (probability distributions) on certain nonlinear domains. The domains of interest are trivial vector bundles, with a focus on (finite-dimensional) parallelizable Riemannian manifolds. The populations are represented by mixtures of Gaussians on tangent bundles of these manifolds, and the populations are compared using a convenient expression for a Wasserstein-type distance. This distance is called Wasserstein type because the search for optimal couplings is restricted to joint mixtures of Gaussians. The paper demonstrates this framework for several examples involving simulated and real data. It uses simulated populations on a unit sphere $\mathbb{S}^2$ to explain how one can compare distributions. The process also involves steps for modeling populations using mixtures of Gaussians and estimating mixture parameters using clustering methods.

An important application of this framework is in comparing populations of shapes using image data. This paper uses videos of nanoparticles during a manufacturing process to pursue this application. One associates particles in an imaged frame as samples from a shape population and compares different image frames using the Wasserstein-type metric between associated shape populations. It further develops a procedure for detecting a change point in the temporal evolution of a shape population during manufacturing.

In the future, we would like to adapt this framework for solving shape regression problems. In these problems, the shape populations of objects serve as response variables with some Euclidean input variable influencing the outcomes. The goal is to develop statistical models capturing the relationships between input variables and output shape populations.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Appendix A. Proofs of theoretical results.

A.1. Proof of Proposition 3.2.

Proof. First, assume that μ is Gaussian, let $h : V \to \mathbb{R}^d$ be an arbitrary linear isometry, and set $\nu = h_\# \mu$. Then for any linear functional $f : \mathbb{R}^d \to \mathbb{R}$, we have $f_\# \nu = f_\#(h_\# \mu) = (f \circ h)_\# \mu$, and $f \circ h : V \to \mathbb{R}$ is a linear functional. It follows that $f_\# \nu$ is Gaussian on $\mathbb{R}$, so that ν must be Gaussian on $\mathbb{R}^d$. Setting $g = h^{-1}$, we have that $\mu = g_\# \nu$.
Conversely, suppose that $\mu = g_\# \nu$ for a Gaussian ν on $\mathbb{R}^d$. Then for any linear functional $f : V \to \mathbb{R}$, we have that $f_\# \mu = f_\# g_\# \nu = (f \circ g)_\# \nu$ is a Gaussian on $\mathbb{R}$. It follows that μ is Gaussian on V. ■

A.2. Proof of Proposition 3.3.

Proof. We will show that the quantities $\tilde{m} := g(m)$ and $\widetilde{\Sigma} := g^{-1}\Sigma g$ are intrinsic to the measure $\mu := g_\# \nu = g'_\# \nu'$, from which the claims will follow. We have

$$\int_V v \, d\mu(v) = \int_V v \, d(g_\# \nu)(v) = \int_{\mathbb{R}^d} g(w) \, d\nu(w) = g\left(\int_{\mathbb{R}^d} w \, d\nu(w)\right) = g(m),$$

where the second equality is the change-of-variables formula and third follows by the assumption that g is an isometry. Next, we have

$$\int_V \langle u, x - \tilde{m}\rangle_V \langle v, x - \tilde{m}\rangle_V \, d\mu(x) = \int_{\mathbb{R}^d} \langle g(s), g(y) - g(m)\rangle_V \langle g(t), g(y) - g(m)\rangle_V \, d\nu(y) \tag{A.1}$$

$$= \int_{\mathbb{R}^d} \langle s, y - m\rangle \langle t, y - m\rangle \, d\nu(y) \tag{A.2}$$

$$= \langle \Sigma s, t\rangle \tag{A.3}$$

$$= \langle g^{-1}\widetilde{\Sigma} g g^{-1} u, g^{-1} v\rangle = \langle \widetilde{\Sigma} u, v\rangle_V.$$

(A.1) is the change-of-variables formula with $s := g^{-1}(u)$ and $t := g^{-1}(v)$, (A.2) uses the fact that g is a linear isometry, (A.3) is [5, Corollary 1.2.3], and the remaining equalities follow by definition and the fact that g^{-1} is an isometry. These identities give the desired intrinsic characterizations. ■

A.3. Proof of Proposition 3.7.

Proof. Because μ is a Gaussian mixture measure, its support is of the form $\cup_j^J \mathcal{E}_{m_j}$ for some pairwise distinct points $m_j \in \mathcal{M}$. Fix $m = m_j$, and consider the restriction of $\mu|_{\mathcal{E}_m}$ of μ to $\mathcal{E}_m$. There are some subcollections $\{\eta_{k_i}\}_{i=1}^I$ and $\{\eta'_{k_i}\}_{i=1}^{I'}$ of measures which are Gaussians on $\mathcal{E}_m$. Let us assume without loss of generality that $(\mathcal{E}_m, \langle\cdot,\cdot\rangle_m) = (\mathbb{R}^d, \langle\cdot,\cdot\rangle)$ (the latter endowed with the standard inner product)—the two inner product spaces are isometric, so this assumption can be made without loss of generality, allowing us to suppress the isometry from the notation. Then the measures

$$\sum_{i=1}^{I} \left(\frac{w_{k_i}}{\mu(\mathcal{E}_m)}\right) \eta_{k_i} \quad \text{and} \quad \sum_{i=1}^{I'} \left(\frac{w'_{k_i}}{\mu(\mathcal{E}_m)}\right) \eta'_{k_i}$$

are representations of the Gaussian mixture measure $\frac{1}{\mu(\mathcal{E}_m)}\mu|_{\mathcal{E}_m}$ on $\mathbb{R}^d$ which are in minimal form. It follows from Proposition 2.4 that $I = I'$ and that the w_{k_i} and η_{k_i} agree with the w'_{k_i}

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

and η'_{k_i} up to a permutation of $\{1,\dots,I\}$. Running the same argument on each m_j completes the proof. ■

A.4. Proof of Proposition 3.10.

Proof. First, observe that any smooth inner product on $\mathcal{M}\times\mathbb{R}^d$ is bundle isometric to the standard one. Indeed, this is achieved by choosing a smoothly- varying orthonormal basis (with respect to the arbitrary inner product) for each fiber $\{m\}\times\mathbb{R}^d$ and then defining the bundle isomorphism by sending this to the standard orthonormal basis. Now, for an arbitrary trivialization $\varphi:\mathcal{E}\to\mathcal{M}\times\mathbb{R}^d$, define an inner product on $\mathcal{M}\times\mathbb{R}^d$ by pulling back each $\langle\cdot,\cdot\rangle_m$ via φ^{-1}. Choose a bundle isometry $\psi:\mathcal{M}\times\mathbb{R}^d\to\mathcal{M}\times\mathbb{R}^d$ sending this pullback family of inner products to the standard one. Then $\psi\circ\varphi:\mathcal{E}\to\mathcal{M}\times\mathbb{R}^d$ is a bundle isometry. ■

A.5. Proof of Proposition 3.12.

Proof. We have

$$\varphi_\#\mu=\sum_k w_k\varphi_\#\eta_k$$

by the linearity of pushforwards. Moreover, each $\varphi_\#\eta_k$ is a Gaussian on $\mathcal{E}'$ with mean m_k. Indeed, since φ is a bundle isometry, it must be that $\varphi_\#\eta_k$ is supported on $\mathcal{E}_{m_k}$. Also, for any linear functional $f:\mathcal{E}'_{m_k}\to\mathbb{R}$, we have

$$f_\#(\varphi_\#\eta_k)=(f\circ\varphi_{m_k})_\#\eta_k,$$

and $f\circ\varphi_{m_k}$ is a linear functional on $\mathcal{E}_{m_k}$, so the result must be a Gaussian on $\mathbb{R}$. It remains to derive the formula for the covariance operator of the pushforward. Choose a linear isometry $g:\mathbb{R}^d\to\mathcal{E}_{m_k}$ and a covariance operator $\widetilde{\Sigma}\in\mathsf{Sym}_d^+$ such that $\Sigma=g\widetilde{\Sigma}g^{-1}$ (see Proposition 3.3 and Definition 3.4). Then $\varphi_{m_k}\circ g:\mathbb{R}^d\to\mathcal{E}'_{\varphi(m_k)}$ is a linear isometry taking the covariance operator $\widetilde{\Sigma}$ to

$$(\varphi_{m_k}\circ g)\widetilde{\Sigma}(\varphi_{m_k}\circ g)^{-1}=\varphi_{m_k}\Sigma\varphi_{m_k}^{-1}.$$

The converse follows by the same argument, using φ^{-1} in place of φ. ■

A.6. Proof of Proposition 3.15.

Proof. By Proposition 3.12, $\varphi_\#\eta_i$ is a Gaussian on the product bundle $\mathcal{M}\times\mathbb{R}^d$, specifically,

$$\varphi_\#\eta_i=N_{\mathcal{M}\times\mathbb{R}^d}(m_i,\varphi_{m_i}\Sigma\varphi_{m_i}^{-1}).$$

By definition of d_φ, φ is an isometry of metric spaces and so induces an isometry at the level of Wasserstein spaces through the pushforward map, that is,

$$W_2^\varphi(\eta_0,\eta_1)=W_2^{\mathcal{M}\times\mathbb{R}^d}(\varphi_\#\eta_0,\varphi_\#\eta_1).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Applying Lemma 3.14, we have

$$\begin{aligned} W_2^{\varphi}(\eta_0,\eta_1) &= W_2^{\mathcal{M}\times\mathbb{R}^d}(\varphi_{\#}\eta_0,\varphi_{\#}\eta_1)\\ &= d_{\mathcal{M}}(m_0,m_1)^2\\ &\quad + \operatorname{tr}(\varphi_{m_0}\Sigma_0\varphi_{m_0}^{-1})\\ &\quad + \operatorname{tr}\left(\varphi_{m_1}\Sigma_1\varphi_{m_1}^{-1} - 2\left(\left(\varphi_{m_0}\Sigma_0\varphi_{m_0}^{-1}\right)^{\frac{1}{2}}\varphi_{m_1}\Sigma_1\varphi_{m_1}^{-1}\left(\varphi_{m_0}\Sigma_0\varphi_{m_0}^{-1}\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)\\ &= d_{\mathcal{M}}(m_0,m_1)^2 + \operatorname{tr}(\Sigma_0+\Sigma_1-2(\Sigma_0^{\frac{1}{2}}\Phi_{m_0,m_1}^{-1}\Sigma_1\Phi_{m_0,m_1}\Sigma_0^{\frac{1}{2}})^{\frac{1}{2}}), \end{aligned}$$

where the last equality follows by linearity and cyclic permutation invariance of trace, together with the fact that the square root of a symmetric positive definite matrix is invariant under change of basis, in the sense that $(\varphi\Sigma\varphi^{-1})^{\frac{1}{2}} = \varphi\Sigma^{\frac{1}{2}}\varphi^{-1}$. ■

A.7. Proof of Lemma 3.14.

Proof. Any coupling $\pi \in \Pi(\eta_0,\eta_1)$ must be supported on the product of the supports of the η_i, that is,

$$\operatorname{supp}(\pi) \subset (\{m_0\}\times\mathbb{R}^d)\times(\{m_1\}\times\mathbb{R}^d) \approx \mathbb{R}^d\times\mathbb{R}^d.$$

We then have

$$\begin{aligned} W_2^{\mathcal{M}\times\mathbb{R}^d}(\eta_0,\eta_1)^2 &= \inf_{\pi\in\Pi(\eta_0,\eta_1)}\int_{(\mathcal{M}\times\mathbb{R}^d)\times(\mathcal{M}\times\mathbb{R}^d)} d_{\mathcal{M}\times\mathbb{R}^d}((p_0,v_0),(p_1,v_1))^2 d\pi((p_0,v_0),(p_1,v_1))\\ &= \inf_{\pi\in\Pi(\eta_0,\eta_1)}\int_{(\{m_0\}\times\mathbb{R}^d)\times(\{m_1\}\times\mathbb{R}^d)} d_{\mathcal{M}}(m_0,m_1)^2\\ &\qquad + \int_{(\{m_0\}\times\mathbb{R}^d)\times(\{m_1\}\times\mathbb{R}^d)} \|v_0-v_1\|^2 d\pi((m_0,v_0),(m_1,v_1))\\ &= d_{\mathcal{M}}(m_0,m_1)^2\\ &\qquad + \inf_{\pi\in\Pi(\eta_0,\eta_1)}\int_{(\{m_0\}\times\mathbb{R}^d)\times(\{m_1\}\times\mathbb{R}^d)} \|v_0-v_1\|^2 d\pi((m_0,v_0),(m_1,v_1)). \end{aligned} \tag{A.4}$$

Now consider the bijection

$$\begin{aligned} \rho: (\{m_0\}\times\mathbb{R}^d)\times(\{m_1\}\times\mathbb{R}^d) &\to \mathbb{R}^d\times\mathbb{R}^d\\ ((m_0,v_0),(m_1,v_1)) &\mapsto (v_0,v_1). \end{aligned}$$

We claim that this induces a correspondence:

$$\rho_{\#}: \Pi(\eta_0,\eta_1) \to \Pi(N_d(0,\Sigma_0),N_d(0,\Sigma_1)).$$

Indeed, for $\pi \in \Pi(\eta_0,\eta_1)$ and any measurable subset $A \subset \mathbb{R}^d$, we have

$$\rho_{\#}\pi(A\times\mathbb{R}^d) = \pi(\rho^{-1}(A\times\mathbb{R}^d)) = \pi((\{m_0\}\times A)\times\{m_1\}\times\mathbb{R}^d) = \eta_0(\{m_0\}\times A) = N_d(0,\Sigma_0)(A),$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

and the computation for the other marginal is similar. Combining this observation with the change-of-variables formula yields

$$\begin{aligned}&\inf_{\pi\in\Pi(\eta_0,\eta_1)}\int_{(\{m_0\}\times\mathbb{R}^d)\times(\{m_1\}\times\mathbb{R}^d)}\|v_0-v_1\|^2 d\pi((m_0,v_0),(m_1,v_1))\\ &=\inf_{\pi\in\Pi(\eta_0,\eta_1)}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|v_0-v_1\|^2 d(\rho_\#\pi)(v_0,v_1)\\ &=\inf_{\xi\in\Pi(N_d(0,\Sigma_0),N_d(0,\Sigma_1))}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|v_0-v_1\|^2 d\xi(v_0,v_1).\end{aligned}$$

Thus, the second term of (A.4) is equivalent to computing the Wasserstein distance between mean-zero Gaussians in $\mathbb{R}^d$, so Proposition 2.2 implies that the claimed formula for $W_2^{\mathcal{M}\times\mathbb{R}^d}(\eta_0,\eta_1)$ holds. Moreover, the proposition tells us that there is an optimal coupling ξ which is a Gaussian on $\mathbb{R}^d\times\mathbb{R}^d$ with mean zero. The pushforward $(\rho^{-1})_\#\xi\in\Pi(\eta_0,\eta_1)$ is an optimal coupling for the original Wasserstein distance. To prove the last statement of the proposition, first consider the measurable map

$$\begin{aligned}\tau:\mathbb{R}^d\times\mathbb{R}^d&\to(\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d)\\ (v_0,v_1)&\mapsto((m_0,m_1),(v_0,v_1)).\end{aligned}$$

It is easy to see that $\tau_\#$ takes zero-mean Gaussians on $\mathbb{R}^d\times\mathbb{R}^d$ to Gaussians on the product bundle $(\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d)$, explicitly,

$$\tau_\# N_{2d}(0,\Sigma)=N_{(\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d)}((m_0,m_1),\Sigma).$$

Next, observe that $\rho^{-1}=\sigma\circ\tau$, so it follows that the optimal coupling $(\rho^{-1})_\#\xi$ can be expressed as $\sigma_\#\overline{\pi}$, where $\overline{\pi}:=\tau_\#\xi\in\mathsf{G}((\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d))$. ■

A.8. Proof of Lemma 3.16.

Proof. For notational convenience, let $\mathcal{E}=\mathcal{M}\times\mathbb{R}^d$ and $\mathcal{F}=(\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d)$ for the rest of the proof. Let $\omega\in\Pi(\mu_0^*,\mu_1^*)$ be an optimal coupling for $W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*)$, so that

$$W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*)^2=\sum_{k,\ell}\omega_{k\ell}W_2^{\mathcal{E}}(\eta_0^k,\eta_1^\ell)^2.$$

For each pair of indices (k,ℓ), choose an optimal coupling $\pi^{k\ell}\in\Pi(\eta_0^k,\eta_1^\ell)$ for the Wasserstein distance $W_2^{\mathcal{E}}(\eta_0^k,\eta_1^\ell)$, and set $\pi=\sum_{k,\ell}\omega_{k\ell}\pi^{k\ell}$ (the superscript notation $\pi^{k\ell}$ is intended to distinguish from the notation $\pi_{k\ell}$, which we used for an entry in a matrix). Then $\pi\in\Pi(\mu_0,\mu_1)$. We also claim that $\pi\in\mathsf{GM}_\sigma(\mathcal{E})$. Indeed, from Lemma 3.14, each $\pi^{k\ell}$ is of the form $\sigma_\#\overline{\pi}^{k\ell}$ for some $\overline{\pi}^{k\ell}\in\mathsf{G}(\mathcal{F})$, and therefore $\pi=\sigma_\#\overline{\pi}$, where $\overline{\pi}:=\sum_{k,\ell}\omega_{k\ell}\overline{\pi}^{k\ell}\in\mathsf{GM}(\mathcal{F})$. Moreover,

$$\begin{aligned}&\int_{\mathcal{E}\times\mathcal{E}}d_{\mathcal{E}}((p_0,v_0),(p_1,v_1))^2 d\pi((p_0,v_0),(p_1,v_1))\\ &\quad=\sum_{k,\ell}\omega_{k\ell}\int_{\mathcal{E}\times\mathcal{E}}d_{\mathcal{E}}((p_0,v_0),(p_1,v_1))^2 d\pi^{k\ell}((p_0,v_0),(p_1,v_1))\\ &\quad=\sum_{k,\ell}\omega_{k\ell}W_2^{\mathcal{E}}(\eta_0^k,\eta_1^\ell)^2=W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*)^2.\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Since π is not necessarily optimal for the mixture Wasserstein distance, it follows that $MW_2^{\mathcal{E}}(\mu_0,\mu_1) \leq W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*)$.

To see the reverse inequality, suppose that π is an arbitrary coupling in the feasible set for $MW_2^{\mathcal{E}}(\mu_0,\mu_1)$. Let $\pi = \sigma_\#\overline{\pi}$ for some $\overline{\pi} \in \mathsf{GM}(\mathcal{F})$ of the form $\overline{\pi} = \sum_{j=1}^K \omega_j \overline{\pi}^j$, where the ω_j are positive real numbers satisfying $\sum_j \omega_j = 1$ and where each $\overline{\pi}^j$ is a Gaussian on the vector bundle $\mathcal{F}$. Therefore, $\pi = \sum_{j=1}^K \omega_j \pi^j$, where $\pi^j := \sigma_\# \overline{\pi}^j$. Let $\rho_i : \mathcal{E}\times\mathcal{E} \to \mathcal{E}$ be a coordinate projection onto the left, for $i=0$, or right, for $i=1$, component. The marginal condition on π implies that

$$\sum_{k=1}^{K_0} w_0^k \eta_0^k = \mu_0 = (\rho_0)_\#\pi = \sum_{j=1}^{K} \omega_j (\rho_0)_\#\pi^j.$$

Corollary 3.8 then tells us that for each j, there is some k such that $(\rho_0)_\#\pi^j = \eta_0^k$. Moreover, it must be that $(\rho_1)_\#\pi^j = \eta_1^\ell$ for some ℓ by the second marginal condition on π. Putting these observations together and reindexing with double indices, for convenience, we write $\pi = \sum_{k,\ell} \omega_{k\ell}\pi^{k\ell}$, where the marginals of $\pi^{k\ell}$ are η_0^k and η_1^ℓ, respectively. It follows that $\omega := (\omega_{k\ell})_{k,\ell}$ defines a coupling of μ_0^* and μ_1^*, so a computation very similar to the one in the previous paragraph shows that the value of the mixture Wasserstein distance objective on the coupling π is bounded below by $W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*)$. Since π was an arbitrary element of $\Pi(\mu_0,\mu_1)\cap \mathsf{GM}_\sigma(\mathcal{E})$, this completes the proof. ■

A.9. Proof of Theorem 3.17.

Proof. We claim that

$$MW_2^{\varphi}(\mu_0,\mu_1) = MW_2^{\mathcal{M}\times\mathbb{R}^d}(\varphi_\#\mu_0,\varphi_\#\mu_1).$$

Indeed, for any $\overline{\pi} \in \mathsf{GM}((\mathcal{M}\times\mathcal{M})\times(\mathbb{R}^d\times\mathbb{R}^d))$, the equality

$$\int_{(\mathcal{M}\times\mathbb{R}^d)\times(\mathcal{M}\times\mathbb{R}^d)} d_{\mathcal{M}\times\mathbb{R}^d}((p_0,v_0),(p_1,v_1))^2 d(\sigma_\#\overline{\pi})((p_0,v_0),(p_1,v_1))$$
$$= \int_{\mathcal{E}\times\mathcal{E}} d_\varphi(w_0,w_1)^2 d((\varphi^{-1}\circ\sigma)_\#\overline{\pi})(w_0,w_1)$$

holds by the change-of-variables formula, the definition of d_φ, and the functoriality of push-forwards. Minimizing over couplings (of $\varphi_\#\mu_0$ and $\varphi_\#\mu_1$) of the form $\sigma_\#\overline{\pi}$ on the left-hand side recovers $MW_2^{\mathcal{M}\times\mathbb{R}^d}(\mu_0,\mu_1)$, and this is equivalent to minimizing over couplings (of μ_0 and μ_1) of the form $(\varphi^{-1}\circ\sigma)_\#\overline{\pi}$ on the right-hand side, which recovers $MW_2^{\varphi}(\mu_0,\mu_1)$.

Next, we show that

$$W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*) = W_2^{\mathsf{G}(\mathcal{M}\times\mathbb{R}^d)}((\varphi_\#\mu_0)^*,(\varphi_\#\mu_1)^*).$$

Observe that

$$\varphi_\#\mu_i = \sum_k w_i^k \varphi_\#\eta_i^k \Rightarrow (\varphi_\#\mu_i)^* = \sum_k w_i^k \delta_{\varphi_\#\eta_i^k}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

It follows that $\Pi(\mu_0^*,\mu_1^*) = \Pi((\varphi_\#\mu_0)^*,(\varphi_\#\mu_1)^*)$ (considered as elements of $\mathbb{R}^{K_1\times K_2}$). For any coupling ω, the fact that $\varphi_\#$ is an isometry from $W_2^{\mathcal{E}}$ to $W_2^{\mathcal{M}\times\mathbb{R}^d}$ implies that

$$\sum_{k,\ell}\omega_{k\ell}W_2^{\varphi}(\eta_0^k,\eta_1^\ell)^2 = \sum_{k,\ell}\omega_{k\ell}W_2^{\mathcal{M}\times\mathbb{R}^d}(\varphi_\#\eta_0^k,\varphi_\#\eta_1^\ell)^2,$$

and this proves the claim.

Now the result follows from Lemma 3.16:

$$\begin{aligned}W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*,\mu_1^*) &= W_2^{\mathsf{G}(\mathcal{M}\times\mathbb{R}^d)}((\varphi_\#\mu_0)^*,(\varphi_\#\mu_1)^*) = MW_2^{\mathcal{M}\times\mathbb{R}^d}(\varphi_\#\mu_0,\varphi_\#\mu_1)\\ &= MW_2^{\varphi}(\mu_0,\mu_1).\end{aligned}$$ ■

A.10. Proof of Proposition 3.19.

Proof. Throughout the proof, we assume without loss of generality that $\mathcal{E}=\mathcal{M}\times\mathbb{R}^d$ and that φ is the identity map. The inequality on the left is trivial by the definition of MW_2^{φ}, (3.1). To see the inequality on the right, let $\tilde{\mu}_i = \sum_{k=1}^{K_i} w_i^k\delta_{m_i^k}$; that is, $\tilde{\mu}_i$ is a sum of Dirac measures located at the means of the Gaussian mixture μ_i—each Dirac $\delta_{m_i^k}$ is considered as a measure on $\mathcal{E}_{m_i^k}$ given by a degenerate Gaussian whose covariance operator is identically zero. By Corollary 3.18, we are able to apply the triangle inequality to deduce

$$MW_2^{\varphi}(\mu_0,\mu_1) \le MW_2^{\varphi}(\mu_0,\tilde{\mu}_0) + MW_2^{\varphi}(\tilde{\mu}_0,\tilde{\mu}_1) + MW_2^{\varphi}(\tilde{\mu}_1,\mu_1).$$

The first term satisfies

$$\begin{aligned}MW_2^{\varphi}(\mu_0,\tilde{\mu}_0)^2 &= \min_{\omega\in\Pi(w_0,w_1)}\sum_{k,\ell}\omega_{k\ell}W_2^{\varphi}(\eta_0^k,\delta_{m_1^\ell})^2\\ &= \min_{\omega\in\Pi(w_0,w_1)}\sum_{k,\ell}\omega_{k\ell}\left(d_{\mathcal{M}}(m_0^k,m_1^\ell)^2 + \mathrm{tr}(\Sigma_0^k)\right)\\ &= \min_{\omega\in\Pi(w_0,w_1)}\sum_{k,\ell}\omega_{k\ell}d_{\mathcal{M}}(m_0^k,m_1^\ell)^2 + \sum_k w_0^k\mathrm{tr}(\Sigma_0^k)\\ &= \sum_k w_0^k\mathrm{tr}(\Sigma_0^k),\end{aligned}$$

where the last line follows because the first term is made zero by taking ω to be the diagonal coupling. Likewise, $MW_2^{\varphi}(\tilde{\mu}_1,\mu_1)^2 = \sum_k w_1^k\mathrm{tr}(\Sigma_1^k)$, so that

$$MW_2^{\varphi}(\mu_0,\mu_1) \le MW_2^{\varphi}(\tilde{\mu}_0,\tilde{\mu}_1) + \sum_{i=0,1}\left(\sum_{k=1}^{K_i} w_i^k\mathrm{tr}(\Sigma_i^k)\right)^{1/2}.$$

We claim that $MW_2^{\varphi}(\tilde{\mu}_0,\tilde{\mu}_1)\le W_2^{\varphi}(\mu_0,\mu_1)$, and this will complete the proof. First, observe that

$$\begin{aligned}MW_2^{\varphi}(\tilde{\mu}_0,\tilde{\mu}_1)^2 &= \min_{\omega\in\Pi(w_0,w_1)}\sum_{k,\ell}\omega_{k\ell}W_2^{\varphi}(\delta_{m_0^k},\delta_{m_1^\ell})^2\\ &= \min_{\omega\in\Pi(w_0,w_1)}\sum_{k,\ell}\omega_{k\ell}d_{\mathcal{M}}(m_0^k,m_1^\ell)^2 = W_2^{\varphi}(\tilde{\mu}_0,\tilde{\mu}_1)^2.\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Next, we show that $W_2^\varphi(\tilde{\mu}_0, \tilde{\mu}_1) \leq W_2^\varphi(\mu_0, \mu_1)$. Let $\pi \in \Pi(\mu_0, \mu_1)$ be an arbitrary coupling. Define a coupling $\omega \in \Pi(w_0, w_1)$ by

$$\omega_{k\ell} = \pi(\mathcal{E}_{m_0^k} \times \mathcal{E}_{m_1^\ell}).$$

This is indeed a coupling,

$$\sum_\ell \omega_{k\ell} = \sum_\ell \pi(\mathcal{E}_{m_0^k} \times \mathcal{E}_{m_1^\ell}) = \pi(\mathcal{E}_{m_0^k} \times \mathcal{E}) = w_0^k,$$

and the other marginal condition follows similarly. The cost of this coupling with respect to the W_2^φ-distance from $\tilde{\mu}_0$ to $\tilde{\mu}_1$ satisfies

$$\begin{aligned}
\sum_{k,\ell} \omega_{k\ell} d_{\mathcal{M}}(m_0^k, m_1^\ell)^2 &= \sum_{k,\ell} d_{\mathcal{M}}(m_0^k, m_1^\ell)^2 \pi(\mathcal{E}_{m_0^k} \times \mathcal{E}_{m_1^\ell}) \\
&= \sum_{k,\ell} d_{\mathcal{M}}(m_0^k, m_1^\ell)^2 \int_{((m_0^k, v_0),(m_1^\ell, v_1)) \in \mathcal{E}_{m_0^k} \times \mathcal{E}_{m_1^\ell}} \pi(d(m_0^k, v_0) \times d(m_1^\ell, v_1)) \\
&= \int_{((m_0, v_0),(m_1, v_1)) \in \mathcal{E} \times \mathcal{E}} d_{\mathcal{M}}(m_0, m_1)^2 \pi(d(m_0, v_0) \times d(m_1, v_1)) \\
&\leq \int_{((m_0, v_0),(m_1, v_1)) \in \mathcal{E} \times \mathcal{E}} d_\varphi(m_0, m_1)^2 \pi(d(m_0, v_0) \times d(m_1, v_1)),
\end{aligned}$$

and the last quantity is the cost of the coupling π with respect to the W_2^φ-distance between μ_0 and μ_1. Since this construction and bound hold for any choice of π, this shows that $W_2^\varphi(\tilde{\mu}_0, \tilde{\mu}_1) \leq W_2^\varphi(\mu_0, \mu_1)$, and the proof is complete. ■

A.11. Proof of Proposition 3.21.

Proof. By Theorem 3.17, $MW_2^{\varphi^{p,F}}(\mu_0, \mu_1) = W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*, \mu_1^*)$, where for $\mu_i = \sum_{k=1}^{K_i} w_i^k \eta_i^k$,

$$W_2^{\mathsf{G}(\mathcal{E})}(\mu_0^*, \mu_1^*)^2 = \min_{\pi \in \Pi(w_0, w_1)} \Sigma_{k,l} \pi_{kl} W_2^{\varphi^{p,F}}(\eta_0^k, \eta_1^\ell)^2.$$

By Proposition 3.15, for $\eta_j = N_{\mathcal{E}}(m_j, \Sigma_j)$,

$$W_2^{\varphi^{p,F}}(\eta_0, \eta_1) = d_{\mathcal{M}}(m_0, m_1)^2 + \operatorname{tr}\left(\Sigma_0 + \Sigma_1 - 2\left(\Sigma_0^{\frac{1}{2}} \left(\Phi_{m_0,m_1}^{p,F}\right)^{-1} \Sigma_1 \Phi_{m_0,m_1}^{p,F} \Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right),$$

with $\Phi_{m_0,m_1}^{p,F} = (\varphi_{m_1}^{p,F})^{-1} \circ \varphi_{m_0}^{p,F} : \mathcal{E}_{m_0} \to \mathcal{E}_{m_1}$. We claim that

$$(\varphi_{m_1}^{p,F})^{-1} \circ \varphi_{m_0}^{p,F} = P_{m_1} \circ P_{m_0}^{-1};$$

since the map on the right-hand side does not depend on F, the result follows. The claim is verified by a direct calculation: For $v \in T_{m_0}\mathcal{M}$, write $v = \sum_j v_j f_j(m_0)$. Then

$$P_{m_1} \circ P_{m_0}^{-1}(v) = P_{m_1}\left(\sum_j v_j f_j\right) = \sum_j v_j f_j(m_1)$$

and

$$(\varphi_{m_1}^{p,F})^{-1} \circ \varphi_{m_0}^{p,F}(v) = (\varphi_{m_1}^{p,F})^{-1}((v_j)_j) = \sum_j v_j f_j(m_1).$$ ■

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Appendix B. Details of experimental setups.

B.1. Brief introduction to the shape space of planar contours. Here, we describe a mathematical representation of shape of closed, planar contours as elements of a finite-dimensional unit sphere. Let $\mathcal{AC}_0$ denote the set of all absolutely continuous curves of the type $\beta:\mathbb{S}^1\to\mathbb{R}^2$ such that $\beta(t)=0$ for some $t\in\mathbb{S}^1$. An element $\beta\in\mathcal{AC}_0$ represents a parameterized planar, closed curve passing through the origin. We are interested in quantifying the shape of β in a manner that is invariant to its rotation, translation, scale, and reparameterization. Taking an elastic approach to shape analysis of curves [36, 37], we represent β using its square-root velocity function (SRVF) $q(t)=\frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}}$. The mapping $\beta\mapsto q$ is a bijection from $\mathcal{AC}_0$ to $\mathbb{L}^2(\mathbb{S}^1,\mathbb{R}^2)$. Representing a curve β by its SRVF q removes the effect of its translations. Further, we rescale β to have unit length, so that $\|q\|^2=\text{length}(\beta)=1$. The set of all scaled SRVFs forms a unit Hilbert sphere $\mathbb{S}_\infty\subset\mathbb{L}^2$. To remove reparametrizations, we select a representative shape (this can be same as the reference point p on the manifold $\mathcal{M}$ needed to build a global frame). We reparameterize this representative curve to be arc-length and then register, through reparameterization, all individual curves (in a given data set) to this curve. Now we have removed translation, scaling, and reparameterization.

Next, we consider a discretized representation of curves as follows. We sample an SRVF q using T uniformly spaced sample points $\{t_i\in\mathbb{S}^1, i=1,\dots,T\}$ and denote the samples by an array $\mathbf{q}\in\mathbb{R}^{2\times T}$, where $\mathbf{q}_i=q(t_i)$. To ensure unit scale, we rescale the array $\mathbf{q}$ to have Frobenious norm one, and thus we have $\mathbf{q}\in\mathbb{S}^{2T-1}$. To remove rotation, we use Procrustes alignment in a pairwise fashion as follows. Define the action of $SO(2)$ on $\mathbb{S}^{2T-1}$ as $(O,\mathbf{q})=Oq$, and form equivalence classes $[\mathbf{q}]=\{O\mathbf{q}|O\in SO(2)\}$. The shape space of discrete contours, $\mathcal{S}_c$, is the set of all equivalence classes and denoted by the quotient space $\mathbb{S}^{2T-1}/SO(2)$. Given any two curves $\beta_1,\beta_2\in\mathcal{AC}_0$ and their corresponding discrete SRVFs $\mathbf{q}_1,\mathbf{q}_2\in\mathbb{S}^{2T-1}$, $O^*=\arg\min_{O\in SO(2)}\|\mathbf{q}_1-O\mathbf{q}_2\|^2$.

The shape metric is then given by $d_s([\mathbf{q}_1],[\mathbf{q}_2])=\cos^{-1}(\langle\mathbf{q}_1,O^*\mathbf{q}_2\rangle)$. Similarly, given a number of contours $\beta_1,\beta_2,\dots,\beta_n$, one can compute the sample mean of their shapes $[\mathbf{q}_1],[\mathbf{q}_2],\dots,[\mathbf{q}_n]$ according to

$$[\hat{m}]=\underset{[\mathbf{q}]\in\mathcal{S}}{\arg\min}\sum_{i=1}^{n}d_s([\mathbf{q}],[\mathbf{q}_i])^2\ .$$

An iterative algorithm for finding the minimizer is presented in several places, including [37]. In this paper, we use a mode-based procedure [14, 13] to reach estimates of mean shape more efficiently. Once we have computed the sample mean, we can rotationally align individual curves to the mean and express them in a preferred orientation according to

$$\mathbf{q}_i^*=O_i^*\mathbf{q}_i,\quad\text{where}\quad O^*=\underset{O\in SO(2)}{\arg\min}\|\hat{m}-O\mathbf{q}_i\|^2\ .$$

These aligned shapes can be treated as elements of $\mathbb{S}^{2T-1}$ for the purpose of statistical modeling and comparisons. Furthermore, we can compute the shooting vectors $\mathbf{v}_i^*=\exp_{\hat{m}}^{-1}(\mathbf{q}_i^*)\in T_{\hat{\mu}_n}(\mathbb{S}^{2T-1})$ (on the unit sphere) and define a covariance matrix $\hat{\Sigma}=\frac{1}{n-1}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{v}_i^T\in$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

$\mathbb{R}^{(2T-1)\times(2T-1)}$. This gives us a way to represent contour shapes as elements of a finite-dimensional unit sphere and to define their sample statistics, such as means and covariance. One can use these statistics to impose a Gaussian model $\eta = N_{\mathcal{E}}(m,\Sigma)$ on $\mathbb{S}^{2T-1}$.

B.2. Mapping between $\mathcal{S}_T$ and $\mathbb{S}^2$. A planar triangle is represented by a matrix $x \in \mathbb{R}^{3\times 2}$ or a complex vector $z \in \mathbb{C}^3$. Let the ith element of z be $z_i = x_{i,1} + j x_{i,2}$. The bijective mappings between the Kendall shape space of triangles $\mathcal{S}_T$ and $\mathbb{S}^2$ (using Hopf fibration) are as follows. The forward map from $\mathcal{S}_T$ to $\mathbb{S}^2$ is given by

$$(x_{1,1},\dots,x_{3,2}) \;\; \mapsto \left(\theta = \cos^{-1}\left(\frac{y_3}{r}\right),\ \varphi = \tan^{-1}(\frac{y_2}{y_1})\right), \quad \text{where}$$
$$y_1 = 2(x_{1,2}x_{2,2} + x_{1,1}x_{2,1}), \quad y_2 = 2(x_{2,1}x_{2,2} - x_{1,1}x_{1,3}), \quad y_3 = 1 - 2(x_{1,2}^2 + x_{2,1}^2),$$

and $r = \sqrt{y_1^2 + y_2^2 + y_3^2}$. The backward map from $(\varphi,\theta) \in \mathbb{S}^2$ to $\mathcal{S}_T$ is given by

$$x_{1,1} = \cos\left(\frac{\psi+\varphi}{2}\right)\sin(\theta/2), \;\; x_{1,2} = \sin\left(\frac{\psi+\varphi}{2}\right)\sin(\theta/2), \quad x_{2,1} = \cos\left(\frac{\psi-\varphi}{2}\right)\cos(\theta/2),$$
$$x_{2,2} = \sin\left(\frac{\psi-\varphi}{2}\right)\cos(\theta/2), \quad x_{3,1} = -(x_{1,1} + x_{2,1}), \quad x_{3,2} = -(x_{1,2} + x_{2,2}),$$

where $\theta \in [0,\pi]$, $\varphi \in [0,\pi]$, and $\psi \in [0,2\pi]$. The angle ψ here is arbitrary and controls the rotation of the resulting triangle.

REFERENCES

[1] L. Ambrosio and N. Gigli, *A User's Guide to Optimal Transport*, Springer-Verlag, Berlin, 2013, pp. 1–155, https://doi.org/10.1007/978-3-642-32160-3_1.

[2] M. Bauer, N. Charon, E. Klassen, S. Kurtek, T. Needham, and T. Pierron, *Elastic metrics on spaces of cuclidean curves: Theory and algorithms*, J. Nonlinear Sci., 34 (2024), pp. 1–37.

[3] R. Benedetti and P. Lisca, *Framing 3-manifolds with bare hands*, Enseign. Math., 64 (2019), pp. 395–413.

[4] K. Bharath, S. Kurtek, A. Rao, and V. Baladandayuthapani, *Radiologic image-based statistical shape analysis of brain tumours*, J. R. Stat. Soc. Ser. C. Appl. Stat., 67 (2018), p. 1357.

[5] V. Bogachev, *Gaussian Measures*, American Mathematical Society, Providence, RI, 1998.

[6] R. Bott and J. Milnor, *On the parallelizability of the spheres*, Bull. Amer. Math. Soc., 64 (1958), pp. 87–89.

[7] J. Cantarella, T. Needham, C. Shonkwiler, and G. Stewart, *Random triangles and polygons in the plane*, Amer. Math. Monthly, 126 (2019), pp. 113–134.

[8] X. Chen and Y. Yang, *Diffusion K-means clustering on manifolds: Provable exact recovery via semidefinite relaxations*, Appl. Comput. Harmon. Anal., 52 (2021), pp. 303–347.

[9] Y. Chen, T. Georgiou, and A. Tannenbaum, *Optimal transport for Gaussian mixture models*, IEEE Access, 7 (2018), pp. 6269–6278.

[10] S. Chowdhury and T. Needham, *Gromov-Wasserstein averaging in a Riemannian framework*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 842–843.

[11] D. Collett and T. Lewis, *Discriminating between the Von Mises and wrapped normal distributions*, Aust. J. Stat., 23 (1981), pp. 73–79.

[12] J. Delon and A. Desolneux, *A Wasserstein-type distance in the space of Gaussian mixture models*, SIAM J. Imaging Sci., 13 (2020), pp. 936–970, https://doi.org/10.1137/19M1301047.

[13] X. Deng, R. Sarkar, E. Labruyère, J.-C. Olivo-Marin, and A. Srivastava, *Characterizing cell populations using statistical shape modes*, in 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

[14] X. Deng, A. Srivastava, R. Sarkar, E. Labruyère, and J.-C. Olive-Marin, *Characterizing cell shape populations using k-mode kernel mixtures*, in International Conference on Pattern Recognition (ICPR), 2022, pp. 2517–2523.

[15] D. Dowson and B. Landau, *The Fréchet distance between multivariate normal distributions*, J. Multivariate Anal., 12 (1982), pp. 450–455, https://doi.org/10.1016/0047-259X(82)90077-X, https://www.sciencedirect.com/science/article/pii/0047259X8290077X.

[16] I. L. Dryden, A. Koloydenko, and D. Zhou, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Appl. Stat., 3 (2009), pp. 1102–1123, https://doi.org/10.1214/09-AOAS249.

[17] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, John Wiley, New York, 1998.

[18] M. Fréchet, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Ann. Inst. H. Poincaré, 10 (1948), pp. 215–310.

[19] C. Givens and R. Shortt, *A class of Wasserstein metrics for probability distributions*, Mich. Math. J., 31 (1984), pp. 231–240, https://doi.org/10.1307/mmj/1029003026.

[20] X. Guo, A. B. Bal, T. Needham, and A. Srivastava, *Statistical shape analysis of brain arterial networks (BAN)*, Ann. Appl. Stat., 16 (2022), pp. 1130–1150.

[21] S. Hauberg, *Directional statistics with the spherical normal distribution*, in 21st International Conference on Information Fusion (FUSION), 2018, pp. 704–711, https://api.semanticscholar.org/CorpusID:52160858.

[22] B. Jain, *On the geometry of graph spaces*, Discrete Appl. Math., 214 (2016), pp. 126–144.

[23] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, *Kernel methods on Riemannian manifolds with Gaussian RBF kernels*, IEEE Trans. Pattern Anal. Mach. Intell., 37 (2015), pp. 2464–2477.

[24] D. G. Kendall, *Shape manifolds, procrustean metrics, and complex projective spaces*, Bull. Lond. Math. Soc., 16 (1984), pp. 81–121, https://doi.org/10.1112/blms/16.2.81, https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/blms/16.2.81.

[25] D. G. Kendall, D. Barden, T. K. Carne, and H. Le, *Shape and Shape Theory*, John Wiley, New York, 1999.

[26] J. Lee, *Smooth Manifolds*, Springer-Verlag, Berlin, 2012.

[27] A. Mallasto and A. Feragen, *Optimal transport distance between wrapped Gaussian distributions*, in MaxEnt 2018—38th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 2018, pp. 1–10.

[28] K. Mardia and P. Jupp, *Directional Statistics*, Vol. 2, Wiley Online Library, New York, 2000.

[29] D. Matteson and N. James, *A nonparametric approach for multiple change point analysis of multivariate data*, J. Amer. Statist. Assoc., 109 (2014), pp. 334–345.

[30] R. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179, https://doi.org/10.1006/aima.1997.1634, https://www.sciencedirect.com/science/article/pii/S0001870897916340.

[31] J. Milnor and J. Stasheff, *Characteristic Classes*, Princeton University Press, Princeton, NJ, 1974.

[32] N. Miolane, N. Guigui, A. Brigant, J. Mathe, B. Hou, Y. Thanwerdas, S. Heyder, O. Peltre, N. Koep, H. Zaatiti, H. Hajri, Y. Cabanes, T. Gerald, P. Chauchat, C. Shewmake, D. Brooks, B. Kainz, C. Donnat, S. Holmes, and X. Pennec, *Geomstats: A Python package for Riemannian geometry in machine learning*, J. Mach. Learn. Res., 21 (2020), pp. 1–9, http://jmlr.org/papers/v21/19-027.html.

[33] G. Peyre and M. Cuturi, *Computational optimal transport*, Found. Trends Mach. Learn., 11 (2019), pp. 355–607.

[34] C. G. Small, *The Statistical Theory of Shape*, Springer-Verlag, Berlin, 1996.

[35] A. Srivastava, S. H. Joshi, W. Mio, and X. Liu, *Statistical shape anlaysis: Clustering, learning and testing*, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), pp. 590–602.

[36] A. Srivastava and E. Klassen, *Functional and Shape Data Analysis*, Vol. 1, Springer-Verlag, Berlin, 2016.

[37] A. Srivastava, E. Klassen, S. Joshi, and I. Jermyn, *Shape analysis of elastic curves in Euclidean spaces*, IEEE Trans. Pattern Anal. Mach. Intell., 33 (2011), pp. 1415–1428.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[38] G. Szekely and M. Rizzo, *Energy statistics: A class of statistics based on distances*, J. Statist. Plann. Inference, 8 (2013), https://doi.org/10.1016/j.jspi.2013.03.018.

[39] A. Takatsu, *On Wasserstein geometry of the space of Gaussian measures*, Probab. Approach Geom., 57 (2010), pp. 463–472.

[40] C. Villani, *Optimal Transport: Old and New*, Vol. 338, Springer-Verlag, Berlin, 2009.

[41] G. Vo and C. Park, *Robust regression for image binarization under heavy noise and nonuniform background*, Commun. Comput. Inf. Sci., 81 (2018), pp. 224–239.

[42] W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, and G. Rohde, *A linear optimal transportation framework for quantifying and visualizing variations in sets of images*, Int. J. Comput. Vis., 101 (2013), pp. 254–269.

[43] T. Woehl, J. Evans, I. Arslan, W. Ristenpart, and N. Browning, *Direct in situ determination of the mechanisms controlling nanoparticle nucleation and growth*, ACS Nano, 6 (2012), pp. 8599–8610.

[44] S. Yakowitz and J. Spragins, *On the identifiability of finite mixtures*, Ann. Math. Stat., 39 (1968), pp. 209–214.

[45] L. Younes, *Shapes and Diffeomorphisms*, Springer-Verlag, Berlin, 2010.

[46] Z. Zhang, J. Su, E. Klassen, H. Le, and A. Srivastava, *Rate-invariant analysis of covariance trajectories*, J. Math. Imaging Vision, 60 (2018), pp. 1306–1323.

Downloaded 08/01/24 to 98.230.63.150 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.