Optimal transport for vector Gaussian mixture models

Anonymous Author(s)
Affiliation
Address
email

Abstract

Vector-valued Gaussian mixtures form an important special subset of vector-valued distributions. In general, vector-valued distributions constitute natural representations for physical entities, which can mutate or transit among alternative manifestations distributed in a given space. A key example is color imagery. In this note, we vectorize the Gaussian mixture model and study several different optimal mass transport related problems associated to such models. The benefits of using vector Gaussian mixture for optimal mass transport include computational efficiency and the ability to preserve structure.

1 Introduction

Finite mixture models can describe a wide range of statistical phenomena. They have been successfully applied to numerous fields including biology, economics, engineering, and the social sciences [14]. The first major use and analysis of mixture models is perhaps due to the mathematician and biostatistician Karl Pearson over 120 years ago, who explicitly decomposed a distribution into two normal distributions for the characterization of the non-normal attributes of forehead to body length ratios in female shore crab populations [16]. The literature on analyzing and applying mixture models is growing due to their simplicity, versatility and flexibility. One of the most commonly used mixture models is the Gaussian mixture model (GMM), which is a weighted sum of Gaussian distributions.

Optimal mass transport (OMT) has been a major subject of mathematical research, originating with the French civil engineer and mathematician Gaspard Monge in 1781 [19, 20]. OMT allows one to define a distance between two probability distributions, which makes it a very powerful tool to analyze the geometry of distributions. Its applications include but not limited to signal processing, machine learning, computer vision, meteorology, statistical physics, quantum mechanics, and network theory [3, 13, 17, 2]. Milestones of this subject include the seminal work of Leonid Kantorovich [19, 20], who relaxed the original problem so that it can be solved through linear programming, and Benamou and Brenier [4] who introduced a computational fluid dynamics (CFD) approach to OMT. More recent developments involve extensions of the theory to the vector-valued, matrix-valued and unbalanced cases [7, 6, 5, 9].

The problem that motivated the present work arose when the authors were working with certain medical image data. The object was to compute optimal mass transport while preserving key structures. The authors of [8] studied OMT for GMM, which however can only work on single layered data, e.g., gray scale images. The need for working directly on the original color images with the potential of capturing more information inspired us to generalize the OMT setting from the one-layered case to the three-layered case. More generally, in this note, we develop optimal transport for vector-valued Gaussian mixture models, which can have any dimension and any general connection

Submitted to the Optimal Transport and Machine Learning Workshop at NeurIPS 2023. Do not distribute.

structures among the layers. Furthermore, corresponding to unbalanced OMT, we also develop an unbalanced version for Gaussian mixture models.

There have several relevant works in the literature describing various versions of OMT to GMMs and vector-valued data as well as extending the theory to manifolds, which we would like to review here in order to put the present work in proper perspective. First of all, Fitschen, Laus, and Schmitzer [12] develop a rigorous transport theory for manifold-valued images. Delon and Desolneux [10] study a version of OMT for GMMs (with some beautiful examples), essentially equivalent to the work proposed in [8] and followed in the present work. Fitschen, Laus and Steidl [11] formulate a dynamical model of transport for discrete RGB color images inspired by the work Benamou-Brenier [4]. In the work of Thorpe *et al.* [18], a transport-based distance is defined and studied, which is directly applicable to general, non-positive and multi-channel signals.

In what follows, we will first give some background on GMM and OMT. Next, we summarize some of the work of [8], and then introduce two different approaches for the vector-valued case. We investigate the unbalanced GMM problem and conclude with some illustrative numerical results.

2 Gaussian mixture models

A Gaussian mixture model is one of the most important examples of a mixture model. Mathematically, a GMM is a probability distribution which is the weighted sum of several Gaussian distributions in $\mathbb{R}^N$. Namely, an *n-component Gaussian mixture model (GMM)* is given by

$$\mu = p_1\nu_1 + p_2\nu_2 + \cdots + p_n\nu_n. \tag{1}$$

Here

$$\nu_i(x) = \frac{1}{\sqrt{(2\pi)^N|\Sigma_i|}}\exp\{-\frac{1}{2}(x-m_i)^T\Sigma_i^{-1}(x-m_i)\}, \tag{2}$$

where $m_i \in \mathbb{R}^N$ is the mean and $\Sigma_i \in \mathbb{R}^{N\times N}$ is the positive definite covariance matrix for $1 \le i \le n$. Further,

$$\sum_{i=1}^{n} p_i = 1, \quad p_i > 0, \forall i \in \{1,...,n\} \tag{3}$$

so that μ is a probability distribution.

We denote the set of all the GMMs in $\mathbb{R}^N$ by $\mathcal{G}(\mathbb{R}^N)$. It is a dense subset of the set of all the probability distributions in the sense of the weak$*$ topology [1]. Thus one can use GMM to fit a distribution with arbitrarily small error. Of course, this may involve a very large number of Gaussians.

3 Optimal mass transport

In this section we sketch the basics of optimal mass transport. See [19, 20] for all the details as well as an extensive list of references. In the present work, we only consider absolutely continuous measures, which thus have density functions representations. By slight abuse of notation and terminology, we will identify the given measure with its density function representation.

The original formulation of OMT due to Gaspard Monge may be expressed as follows:

$$\inf_T\{\int_E c(x,T(x))\rho_0(x)dx \mid T_\#\rho_0 = \rho_1\}, \tag{4}$$

where $c(x,y)$ is the cost of moving unit mass from x to y, which is a lower semi-continuous and bounded below, T is the transport map, and ρ_0, ρ_1 are two probability distributions defined on E, a subdomain of $\mathbb{R}^n$. $T_\#$ denotes the push-forward of T of corresponding measures of the distributions.

As pioneered by Leonid Kantorovich, the Monge formulation of OMT may be relaxed replacing transport maps T by couplings π:

$$\inf_{\pi\in\Pi(\rho_0,\rho_1)}\int_{E\times E} c(x,y)\pi(dx,dy), \tag{5}$$

where $\Pi(\rho_0, \rho_1)$ denotes the set of all the couplings between ρ_0 and ρ_1 (joint distributions whose marginal distributions are ρ_0 and ρ_1).

The discrete Kantorovich form may be written as follows:

$$\min_{\pi \in \Pi(\rho_0, \rho_1)} \sum_i \sum_j c(i,j)\pi(i,j), \tag{6}$$

where $\rho_0 \in \mathbb{R}_+^m, \rho_1 \in \mathbb{R}_+^n$ are two discrete probability density functions ($\sum_i^m \rho_0(i) = \sum_j^n \rho_1(j) = 1$), $\Pi(\rho_0, \rho_1)$ is the set of matrices $\{\pi \in \mathbb{R}_+^{m \times n} | \pi \vec{1}_n = \rho_0,\ \pi^T \vec{1}_m = \rho_1\}$, and $\vec{1}_m$ and $\vec{1}_n$ are vectors all 1's of length m and n, respectively. $c(\cdot, \cdot)$ is a discrete cost function. Kantorovich form is guaranteed to have a optimal solution ($\rho_0 \otimes \rho_1^T \in \Pi(\rho_0, \rho_1)$) while in some cases Monge form might admit no feasible solution.

One may show that for $c(x, y) = \|x - y\|^2$ (square of distance function), the Kantorovich and Monge formulations are equivalent in the absolutely continuous measure case; see [19, 20] and the references therein. Moreover for $c(x, y) = ||x - y||^2$, the specific infimum is called *Wasserstein-2 distance* ($\mathcal{W}_2$).

4 Optimal mass transport for Gaussian mixture models

We are interested in looking at optimal interpolation paths from GMM to another, that is geodesic paths in the space of probability distributions [15]. The problem is that for general GMMs with more than one summands, the optimal path goes out of the subspace of GMMs, that is, the GMM structure is lost. This was exactly the motivation underlying the work of [8]. There are several advantages of preserving the GMM structure including greatly saving computational cost via dimension reduction.

4.1 OMT between Gaussian distributions

For two Gaussian distributions $\mu_i,\ i = 0, 1$ whose means and covariances are m_i and Σ_i, respectively, it is well-known [19, 20] that the $\mathcal{W}_2$ distance between μ_0 and μ_1 has a closed form solution:

$$\mathcal{W}_2(\mu_0, \mu_1)^2 = ||m_0 - m_1||^2 + \operatorname{trace}(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}). \tag{7}$$

For each $t \in [0, 1]$, the distribution μ_t on the geodesic path is a Gaussian whose mean and covariance matrix are defined as follows:

$$m_t = (1 - t)m_0 + tm_1 \tag{8}$$

$$\Sigma_t = \Sigma_0^{-1/2}((1 - t)\Sigma_0 + t(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2})^2\Sigma_0^{-1/2}. \tag{9}$$

4.2 OMT between GMMs

Let μ_0, μ_1 be two Gaussian mixture models of the form

$$\mu_i = p_i^1\nu_i^1 + p_i^2\nu_i^2 + \cdots + p_i^{n_i}\nu_i^{n_i},\ i = 0, 1.$$

Following [8, 10], the distance between μ_0, μ_1 is defined as

$$d(\mu_0, \mu_1)^2 = \min_{\pi \in \Pi(p_0, p_1)} \sum_{i,j} c(i,j)\pi(i,j), \tag{10}$$

where

$$c(i,j) = \mathcal{W}_2(\nu_0^i, \nu_1^j)^2. \tag{11}$$

As ν_0^i and ν_1^j are Gaussian distributions, the $\mathcal{W}_2$ distance may be computed as in (7). In [8, 10], it is proven that $d(\cdot, \cdot)$ is indeed a metric on $\mathcal{G}(\mathbb{R}^N)$. Further, the geodesic on $\mathcal{G}(\mathbb{R}^N)$ connecting μ_0 and μ_1 is given by

$$\mu_t = \sum_{i,j} \pi^*(i,j)\nu_t^{ij}, \tag{12}$$

where ν_t^{ij} is the displacement interpolation in (8) between ν_0^i and ν_1^j. $\pi^*(\cdot, \cdot)$ is the optimal solution of (10).

5 Vector-valued GMM

In this section, we extend the definition of GMM to the vector-valued case, based on which we will formulate generalizations of the work of [8].

5.1 Vector-valued distributions

A *vector-valued distribution* has a corresponding density function which is vector-valued. Formally, a *vector-valued distribution*, $\rho = [\rho_1, ..., \rho_M]$ on $\mathbb{R}^N$, is a map from $\mathbb{R}^N$ to $\mathbb{R}_+^M$ such that

$$\sum_{i=1}^{M} \int_{\mathbb{R}^N} \rho_i(x) dx = 1,$$

with the connections among its M channels, defined by a connected graph $G = (V, E)$, which has M nodes and whose edges determine the connections. Thus, ρ may be considered as a general distribution on $\mathbb{R}^N \times G$, where $V = \{1, 2, \cdots, M\}$ with E defining the connections among the channels (layers). As described in [7], it may represent a physical entity that may mutate or be transported among several alternative manifestations with certain relationships among its M channels.

The Euclidean structure of $\mathbb{R}^N$ and graph structure of E together give a complete metric structure for $\mathbb{R}^N \times G$,

$$d^p((x, u), (y, w)) = ||x - y||^p + \gamma d_G^p(u, w),$$

where $(x, u), (y, w) \in \mathbb{R}^N \times G$, are two points in the space, $p > 0$, $|| \cdot ||$ is the norm of $\mathbb{R}^N$ and $d_G(\cdot, \cdot)$ is the graph distance which is defined as the length of shortest path on G. The vector-valued OMT problem deals with transport on such a metric space.

5.2 Vector GMMs as a subset of vector-valued distributions

Vector-valued GMMs are those vector-valued distributions such that the distribution in each layer is a weighted sum of Gaussians and the weights of the Gaussians sum up to 1. Formally,

$$\rho = p^1 \nu^1 \vec{\delta}_{q^1} + p^2 \nu^2 \vec{\delta}_{q^2} + \cdots + p^n \nu^n \vec{\delta}_{q^n}, \tag{13}$$

where $\vec{\delta}_k$ is a column vector which is the k^{th} column of the M by M identity matrix and q^i is the index of channel where the i^{th} Gaussian lies in. We will always assume that the latter is a probability distribution, i.e.,

$$\sum_{i=1}^{n} p^i = 1. \tag{14}$$

6 Generalization of the OMT GMM framework to vector-valued GMMs

Consider two vector-valued GMMs ρ_0 and ρ_1:

$$\rho_i = p_i^1 \nu_i^1 \vec{\delta}_{q_i^1} + p_i^2 \nu_i^2 \vec{\delta}_{q_i^2} + \cdots + p_i^{n_i} \nu_i^{n_i} \vec{\delta}_{q_i^{n_i}}, \; i = 0, 1.$$

We want to compute an OMT based distance and a displacement interpolation between these two vector-valued distributions with the requirement that the vector GMM structure is preserved along the interpolation path. In short, we want to construct the analogous framework of [8], but replace scalar-valued GMMs with vector-valued GMMs.

As above, let $\Pi(p_0, p_1)$ denote the set of joint probabilities with given marginals p_0 and p_1. Given a graph structure, the most straightforward approach is to require only certain parts of Π to be nonzero, namely only when the the source and target Gaussians are in the same channel or when they are located in adjacent channels. A more detailed description is given in Appendix A.

Unfortunately, this natural (and perhaps most straightforward) generalization may not admit a solution. Indeed, the newly added constraints on Π may not work for general graph structures (See Appendix B for detailed explanations of why the basic generalization doesn't work). Thus, the only other choice left in (10) is to modify the cost matrix $c(\cdot, \cdot)$.

7 Continuous version of $\mathcal{G}(\mathbb{R}^N \times G)$

For vector-valued GMMs, one cannot directly apply the same OMT framework as in the scalar case [8] since the last index is taken discretely. Therefore, we will generalize the framework by making the last index continuous as well. The basic idea is to consider a continuous problem and view the vector-valued distribution as a projection of the continuous solution onto the original discrete space.

More precisely, we propose to extend each point on the edges of the given graph G instead of only taking values on vertices of the graph. Moreover, we extend each edge half-way from both ends, so that newly added points are centered at the original vertices of the graph. Thus, we consider the following point set of a continuous version of the graph G:

$$G^c = \{u + t(w - u) | u, w \in V(G), u \sim w, t \in [-0.5, 0.5]\}. \tag{15}$$

Here $u, w \in V(G)$ are taken as abstract vertices, not as integers. In addition, we assign a length to each edge, γ, so that we are able to perform integration on that set. (We may consider the use of nonuniform edge lengths, in case we are given specific edge weights.)

In fact, we do not need to realize the global structure of the complicated space $\mathbb{R}^N \times G^c$ as a whole. Instead, we can just consider the local structure. A natural and simple way to do that is to impose a manifold structure, which we will now elucidate. We denote the manifold by $\mathcal{M}$.

In order to define $\mathcal{M}$, we need to specify its atlas:

$$A = \{\mathbb{R}^N \times p | p \in [G^c]_0\}, \tag{16}$$

where $[G^c]_0$ is a subset of all continuous paths on G^c which have no cycles (no recurring vertices of G on the path). We can characterize the charts as we stack layers (like "bricks"), where we follow the order of the path on G^c. It is clear that each p is homeomorphic to $\mathbb{R}$, so that each chart is homeomorphic to $\mathbb{R}^{N+1}$.

We want to define a distribution on $\mathcal{M}$ in such a manner that the original distribution is the projection of each layer's range. The projection is defined as the integral of the last index:

$$P_u(f(x, z)) = \sum_{w \sim u} \int_{-0.5}^{0.5} f(x, u + t(w - u))dt, \tag{17}$$

where $P_u(\cdot)$ is the projection of the range of layer u, and f is a distribution on the manifold. The integral range of the last index is the intersection of a ball centered at u which has half-edge radius with G^c (layer u's range). Note that for different w's which are connected to u, the ranges are like different orbits centered at u.

One of the simplest choices for lifting the original distribution to the manifold is a "Gaussian cylinder," i.e., a product of a Gaussian distribution and a uniform distribution within the range of the layer. Thus, we accordingly thicken each Gaussian.

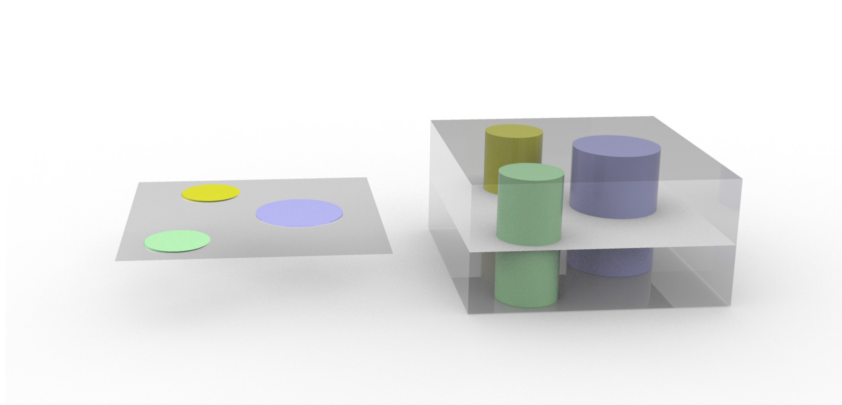

Figure 1: Left hand side is one of the layers of vector GMM. Right hand side is the chart centered at that layer. Gaussians in the original layer become "Gaussian cylinders" on the manifold.

If a layer has more than one edge connected to it, then the original distribution may be lifted to multiple "Gaussian cylinders" (located at all the possible orbits that are centered at the given layer) with combined weights. Notice that even though the "Gaussian cylinders" project to be the same vector-valued distribution within the given layer, they may have different potentials to transport to different directions on the graph.

Let us briefly summarize the optimal transport problem we are going to solve on the manifold $\mathcal{M}$ with the approach we just introduced. Given the projection of each layer's range for the source (starting) and target (terminal) distributions, we want to find corresponding source and target distributions on $\mathcal{M}$ such that the transport cost is optimally low. As before, we first consider the sub-problem where the starting and terminal vector-valued distributions are two Gaussian distributions, which may be located on different layers.

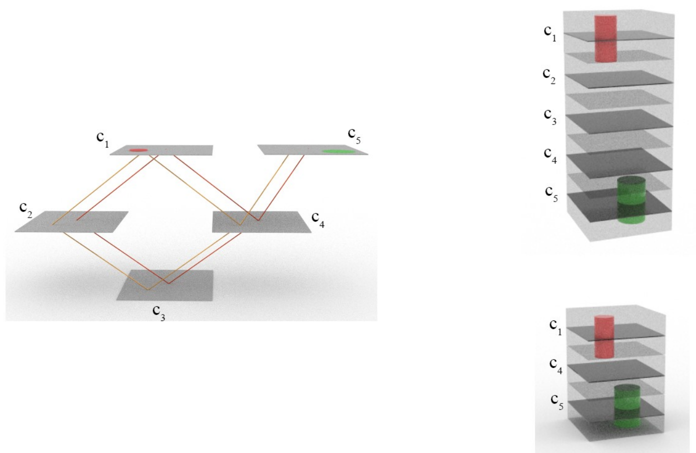

Figure 2: When we consider the transport map from the red Gaussian distribution to the green Gaussian, we consider the transport problem on all the charts that cover both Gaussian distributions. The above figure gives two of the charts.

Theorem 1. *For any two Gaussian cylinder-shaped distributions whose projections on each layer's range are simple Gaussian distributions denoted by ν_0 and ν_1 and located on layers u and w, respectively, the optimal transport $\mathcal{W}_2$ distance between them on $\mathcal{M}$ is given by $d_{\mathcal{M}} = \mathcal{W}_2(\nu_0, \nu_1)^2 + \gamma \tilde{d}_G(u, w)^2$*

Proof. We consider all couplings on the manifold $\mathcal{M}$ denoted by $\Pi(\mathcal{M})$. More precisely, we consider all the possible transports on the charts in A which can cover the supports of both Gaussian cylinders lifted from the two original Gaussians. Namely, we consider all the charts in $\{\mathbb{R}^N \times p | p \in [G^c]_0^{uw}\}$ where $[G^c]_0^{uw}$ denotes the subset of $[G^c]_0$ of those paths contain both layer u and layer w. With this definition, we can explicitly formulate the optimization problem:

$$
\begin{aligned}
d_{\mathcal{M}} &= \inf_{\pi \in \Pi(\mathcal{M})} \int_{\mathcal{M} \times \mathcal{M}} ||\tilde{x} - \tilde{y}||^2 \pi(d\tilde{x}, d\tilde{y}) \\
&= \inf_{p \in [G^c]_0^{uw}} \inf_{\pi \in \Pi^p(\mathbb{R}^{N+1})} \int_{\mathbb{R} \times \mathbb{R}} \int_{\mathbb{R}^N \times \mathbb{R}^N} ||x - y||^2 + |z_1 - z_2|^2 \pi(dx dz_1, dy dz_2)
\end{aligned}
$$

Here, $\Pi^p(\mathbb{R}^{N+1})$ denotes the couplings in $\mathbb{R}^{N+1}$ (which is homeomorphic to $\mathbb{R}^N \times p$) whose marginals are the source and target Gaussian cylinders, respectively.

Further, because of the special structure of "Gaussian cylinders," the first N indices and the last index may be treated separately. If we denote by $\Pi_1(\mathbb{R}^N)$ the set of couplings in $\mathbb{R}^N$ for which the two marginals are the original source and target Gaussians (which does not depend on the path p), and denote by $\Pi_2^p(\mathbb{R})$ the set of couplings whose two marginals are two uniform distributions located in their corresponding layers, the distance expression may be divided into two parts:

$$\inf_{p\in[G^c]_0^{uw}}\inf_{\pi_1\in\Pi_1(\mathbb{R}^N)}\int_{\mathbb{R}^N\times\mathbb{R}^N}||x-y||^2\pi_1(dx,dy)+\inf_{p\in[G^c]_0^{uw}}\inf_{\pi_2\in\Pi_2^p(\mathbb{R})}\int_{\mathbb{R}\times\mathbb{R}}|z_1-z_2|^2\pi_2(dz_1,dz_2)$$

The second term is a simple 1D optimal transport problem between two uniform distributions which are centered at u and w, respectively, with the same radius of thickness of each layer. The optimal transport distance between them is simply the distance between their respective centers, which is easy to calculate since both centers are located on the path p. To be specific, the distance is the length of the path connecting u and w times the thickness of each layer. Hence,

$$\begin{aligned} d_{\mathcal{M}} &= \inf_{\pi_1\in\Pi_1(\mathbb{R}^N)}\int_{\mathbb{R}^N\times\mathbb{R}^N}||x-y||^2\pi_1(dx,dy)+\inf_{p\in[G^c]_0^{uw}}\Delta_p z^2 \\ &= \mathcal{W}_2(\nu_0,\nu_1)^2+\gamma\tilde{d}_G(u,w)^2. \end{aligned}$$

Here the relative distance $\Delta_p z$ is determined by the path p. Moreover, γ is introduced as a parameter for the thickness of each layer's range. We assume that the thickness of each layer is $\sqrt{\gamma}$. The minimum among all the possible paths is just $\tilde{d}_G(u,w)$, the shortest distance on the graph G between vertices u and w. $\square$

Using the latter theorem, we can compute the minimum $\mathcal{W}_2$ cost of moving a source Gaussian distribution to a Gaussian target distribution. Indeed, for the i^{th} and j^{th} Gaussian cylinders on $\mathcal{M}$, we set

$$c_2(i,j)=\mathcal{W}_2(\nu_0^i,\nu_1^j)^2+\gamma\tilde{d}_G(q_0^i,q_1^j)^2. \tag{18}$$

If we take $c_2(\cdot,\cdot)$ in (18) as the cost matrix and compute the Kantorovich formulation of OMT, we can derive a distance:

$$d_{V_2}(\rho_0,\rho_1)^2=\min_{\pi\in\Pi(p_0,p_1)}\sum_{i,j}c_2(i,j)\pi(i,j). \tag{19}$$

This distance is derived from the continuous manifold, but it may be shown to be a metric for our original vector-valued distributions. See the proof in Appendix E.

In addition to the latter distance, we can derive the optimal transport plan $\tilde{\pi}_2^*(\cdot,\cdot)$ from the optimal solution of (19). The transport plan gives the combination of weights for how the "Gaussian cylinders" are arranged at different orbits within each layer's range.

Based on the optimal transport plan, a geodesic (proof in Appendix F) on the manifold $\mathcal{M}$ may be expressed in the following manner:

$$\rho_t=\sum_{i,j}\tilde{\pi}_2^*(i,j)\nu_t^{ij}U_z(path_G(q_0^i,q_1^j,t)), \tag{20}$$

where $U_z(z_0)$ is the 1D uniform distribution density function centered at z_0 on a path of the graph G. The distribution $\nu_t^{ij}U_z(path_G(q_0^i,q_1^j,t))$ at time t is supported on the chart defined by the shortest path that connects q_0^i to q_1^j on the graph, expressed in its own local continuous coordinates for the z index along that path.

For each pair of Gaussians, source and target, a deformation Gaussian cylinder moves across layers following the shortest path on the graph. When it moves across a layer boundary, the Gaussian cylinder is cut into two parts belonging to the respective ranges of two adjacent layers. Each part remains a Gaussian cylinder. Hence after projection of each layer's range, the projected distribution is still a vector GMM distribution. Now if we project (20) onto the range of each layer, we get the following displacement interpolation in the original space:

$$\rho_t=\sum_{i,j}\tilde{\pi}_2^*(i,j)\nu_t^{ij}\vec{\delta}_{path_G(q_0^i,q_1^j,t)}. \tag{21}$$

Note this form of displacement interpolation is very similar to (26) just with slightly different weights.

We also extended our model to the unbalanced case where the mass conservation constraint is relaxed (see Appendix G).

Remark 1: We should note that the way in which we define vector-valued GMM, already makes it a manifold (each layer is its own chart). However, it is impossible to define charts that contain all the possible paths, which is to say the atlas contains only the geometric information within each layer. Our manifold $\mathcal{M}$ on the other hand, has an atlas that contains all the possible paths which encode global geometric information for which we can solve the OMT problem.

Remark 2: This approach gives a geometric intuitive understanding of the optimal transport for vector-valued GMMs. Comparing to the cost matrix (24) in Approach 1, (18) just uses the sum of squares instead of the direct sum, in analogy to the difference between $\mathcal{W}_1$ and $\mathcal{W}_2$. It may seem complicated to consider all the possible paths on the graph in that approach. But for actual computations, we only need to consider the shortest path on the G that connects two layers. Moreover, from the parameter γ that also appears in Approach 1, we find a clear geometric meaning: it represents the thickness of each layer.

8 Numerical results

We applied our method on several real-world data. Figure 3 gives the interpolation between two fitted GMMs of different moons. Figure 4 shows the geodesic path between two fitted nebulae. In Figure 5, a transformation between two different fonts of the word "MATH" is computed via our GMM-based optimal transport method. Some additional numerical examples which further illustrate the properties of our model are included in Appendix H.

Figure 3: Image example: vector-valued GMM geodesic path

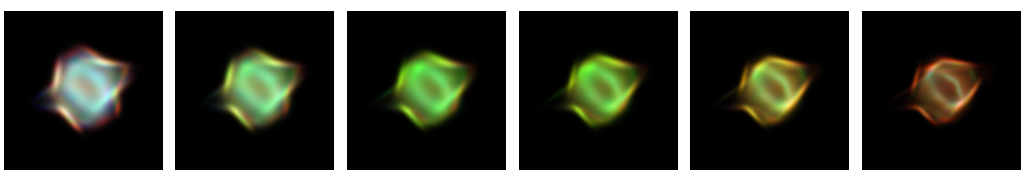

Figure 4: Image example: vector-valued GMM geodesic path

Figure 5: Font transformation

9 Conclusion

This work focuses on the optimal transport for vector-valued GMMs, which is a structured version of vector-valued OMT. As an extension of [8], we defined a distance and geodesic path in the vector-valued case. To the best of our knowledge, the present work is the ***first*** to employ a manifold-based approach to the problem of GMM vector-valued data. Simply applying manifold-valued OMT to vector-valued distributions while preserving the vector GMM structure is not completely straightforward. In fact, just combining the vector-valued case in which the layers are connected by a general graph structure [7] and the GMM metric [8, 10] via adding constraints to the appropriate set of joint probability distributions may not work. See Appendix A below for all of the details. Thus one needs a manifold-based approach in the present situation, which we have shown easily extends to the unbalanced case. In particular, we have extended the approach of transforming the unbalanced scalar OMT problem to the balanced vector-valued problem from CFD [21] to a Kantorovich formulation in this work. Preserving the GMM structure along a transport path both in the balanced and unbalanced cases may have broad applications given the prevalence of such models in many areas of engineering, computer science, and machine learning [14].

The proposed transport is useful, because of its speed advantage and unique ability to preserve structure. This paper just investigates Gaussian mixture case, but it is quite straightforward to apply our framework for other mixture models. We are planning on applying our methodology to the analysis of medical imagery and other appropriate vector-valued distributions including multi-omic data.

References

[1] *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar, and Medical Imaging Real Time Systems*. CRC Press, 0 edition, Dec 2000.

[2] Functional network analysis reveals an immune tolerance mechanism in cancer. *Proceedings of the National Academy of Sciences*, 117(28):16339–16345, jul 2020.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. *arxiv.org*, 1701.07875, 2017.

[4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[5] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Interpolation of density matrices and matrix-valued measures: The unbalanced case. *Euro. Jnl of Applied Mathematics*, 30(3):458–480, 2018.

[6] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. *IEEE Trans. Automatic Control*, 63(8):2612 – 2619, 2018.

[7] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Vector-valued optimal mass transport. *SIAM Journal Applied Mathematics*, 78(3):1682–1696, 2018.

[8] Yongxin Chen, Tryphon T. Georgiou, and Allen Tannenbaum. Optimal mass transport for Gaussian mixture models. *IEEE Access*, 7:6269 – 6278, 2019.

[9] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and Francois-Xavier Vialard. An interpolating distance between optimal transport and Fisher-Rao metrics. *Foundations of Computational Mathematics*, 10:1–44, 2016.

[10] J. Delon and A. Desolneux. A Wasserstein-type distance in the space of Gaussian mixture models. siam journal on imaging sciences, 13(2), 936-970. *SIAM Journal on Imaging Sciences*, 13(2):936–970, 2020.

[11] F. Fitschen, J.H.and Laus and G. Steidl. Transport between RGB images motivated by dynamic optimal transport. *J. Math. Imaging and Vision*, 58:1–21, 2016.

[12] J. H. Fitschen, F. Laus, and B. Schmitzer. Optimal transport for vector-valued images. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 460–472. Springer, 2017.

[13] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.

[14] Geoffrey McLachlan and David Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, sep 2000.

[15] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 2001.

[16] Karl Pearson. III. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. (A.)*, 185:71–110, dec 1894.

[17] Svetlozar T Rachev and Ludger Rüschendorf. *Mass Transportation Problems: Volumes I and II*. Springer Science & Business Media, 1998.

[18] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepcev. A transportation l^p distance for signal analysis. *J Math Imaging and Vision*, 59(2):187–210, 2017.

[19] Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Soc., 2003.

[20] Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008.

[21] Jiening Zhu, Rena Elkin, Jung Hun Oh, Joseph O. Deasy, and Allen Tannenbaum. A vectorial approach to unbalanced optimal mass transport. *IEEE Access*, 8:209224–209231, 2020.