

Neighbor Embeddings Using Unbalanced Optimal Transport Metrics

Muhammad Rana¹ and Keaton Hamm^{1,2}

¹Department of Mathematics, University of Texas at Arlington

²Division of Data Science, University of Texas at Arlington

Abstract

This paper proposes the use of the Hellinger–Kantorovich metric from unbalanced optimal transport (UOT) in a dimensionality reduction and learning (supervised and unsupervised) pipeline. The performance of UOT is compared to that of regular OT and Euclidean-based dimensionality reduction methods on several benchmark datasets including MedMNIST. The experimental results demonstrate that, on average, UOT shows improvement over both Euclidean and OT-based methods as verified by statistical hypothesis tests. In particular, on the MedMNIST datasets, UOT outperforms OT in classification 81% of the time. For clustering MedMNIST, UOT outperforms OT 83% of the time and outperforms both other metrics 58% of the time.

1 Introduction

With advancements in sensing systems including those in medical imaging like MRI and CT scanners, the advent of the Internet of Things, and cheap abundant data storage, the volume of data collection has grown rapidly. Oftentimes, datasets in these fields contain spatial, temporal, and contextual features, and thus datasets have grown not only in volume, but also in dimensionality and complexity. High-dimensionality can lead to challenges for access, analysis, and interpretation. Without additional assumptions on structure within the data, there is typically some form of the curse of dimensionality that appears in a given task-oriented pipeline. For this reason, many works have explored when and how low-dimensional structures appear in high-dimensional data, and have considered how to detect and utilize such structure effectively for learning [1, 2].

One structure that is abundant in theory and practice is that of a low-dimensional embedded manifold [3]. Manifold learning is an umbrella term to capture various algorithms that utilize the manifold structure to learn embeddings of the data into a lower-dimensional (typically Euclidean) space. Examples from the zoo of manifold learning algorithms are Laplacian eigenmaps [4], diffusion maps [5], multidimensional scaling (MDS) [6], t-stochastic neighbor embedding (t-SNE) [7], uniform manifold approximation and projection (UMAP) [8], and Isomap [9]. All these but t-SNE are examples of global neighbor embeddings in the terminology of [3], which work by forming a neighborhood (ε or k -nearest neighbor) graph on the data, with edge weights determined by a metric on the ambient data space. Choosing ε and k must be balanced by the manifold geometry (e.g., curvature and reach), and sampling density of the data from the manifold must be higher for highly curved manifolds [10].

Vectorizing images and using Euclidean distance to describe the geometry lacks the ability to maintain local spatial information. To mitigate this issue, several works have proposed considering

image data as probability measures on a grid in \mathbb{R}^2 [11, 12, 13], where the notion of distance between images becomes from the Wasserstein metric arising from optimal transport (OT) theory [14]. Loosely speaking, the Wasserstein distance between two images corresponds to a minimal energy morphing of one image into another (see Section 2 for a formal definition). One reason for the benefit of using transport-based metrics is that they keep local spatial information. A few works have studied manifold learning when the Wasserstein distance is the ambient metric on the data space [13, 11, 15]. In [13], MDS with the Wasserstein distance is used for classification and compared to a feature-based approach for pathological image data. They found that OT performs similar or better compared to the feature-based approach. In [15], the authors use the Wasserstein-1 distance (or Earth-mover’s distance) with Diffusion maps to discover low-dimensional patterns in gene expression networks resulting in a new, more robust way of extracting features for genetic contributions to patients developing soft-tissue sarcoma.

However, one of the limitations of using the Wasserstein metric is that it requires data to be probability measures, and so does not allow for variations in mass. But in many cases, mass naturally differs and carries meaning. The Hellinger–Kantorovich distances [16] arise from considering unbalanced optimal transport (UOT), and are well-defined metrics on the space of nonnegative measures on a Euclidean domain. These distances allow for different total masses of the data measures by creating or annihilating mass.

This work proposes the use of the UOT-based Hellinger–Kantorovich metrics when forming neighbor embeddings of image data to relax the rigidness of the Wasserstein metric and allow for more meaningful modeling of the data geometry to account for unequal mass and more complex image structures. We study this problem by studying classification and clustering task performance on embedded data with the use of three metrics: Euclidean, Wasserstein-2, and Hellinger–Kantorovich. We compute the accuracies over multiple trials, and compare results via statistical hypothesis test. We use recent medical imaging benchmark datasets from MedMNIST [17] to understand performance on diverse data as well as more classic datasets like MNIST and Coil-20. We find that UOT embeddings outperform OT embeddings a majority of the time for both clustering and classification across all datasets, and UOT embeddings yield better task performance even more often on the MedMNIST datasets.

2 Background

Let $\mathcal{P}_2(\mathbb{R}^n)$, be the set of all probability measures on \mathbb{R}^n with finite 2-nd moment: $\int_{\mathbb{R}^n} |x|^2 d\mu(x) < \infty$. The Wasserstein-2 distance is defined by the Kantorovich problem

$$W_2(\mu, \nu) := \inf_{\pi \in \Gamma(\mu, \nu)} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\pi(x, y) \right)^{\frac{1}{2}}, \quad (1)$$

where $\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \gamma(A \times \mathbb{R}^n) = \mu(A), \gamma(\mathbb{R}^n \times A) = \nu(A) \text{ for all Borel measurable } A \subset \mathbb{R}^n\}$ are couplings, i.e., joint measures whose marginals are μ and ν . For discrete measures, one can exactly compute an optimal coupling via a linear program [14] or use entropic regularization to approximate the distance quickly [18].

To allow for unequal mass, we will use the Hellinger–Kantorovich metric on $\mathcal{M}_+(\Omega)$, the space of nonnegative measures on a bounded convex domain $\Omega \subset \mathbb{R}^n$ [16, 19], which for a given length scale $\kappa > 0$ is given by

$$\text{HK}_\kappa^2(\mu, \nu) := \inf_{\gamma \in \mathcal{M}_+(\Omega \times \Omega)} \left\{ \int_{\Omega \times \Omega} c_\kappa(x, y) d\gamma(x, y) + \text{KL}(P_1 \gamma \mid \mu) + \text{KL}(P_2 \gamma \mid \nu) \right\}, \quad (2)$$

where KL is the Kullback–Leibler divergence, $P_i\gamma$, $i = 1, 2$ denotes the i -th marginal of the coupling measure γ , and the cost c_κ is given by

$$c_\kappa := \begin{cases} -2 \log \cos(|x - y|/\kappa), & |x - y| \leq \kappa\pi/2, \\ +\infty, & \text{otherwise.} \end{cases} \quad (3)$$

The cost function is such that within a length of $\kappa\pi/2$, mass transportation is prioritized, whereas no mass is transported outside that scale, but is rather created and destroyed.

Beyond accounting for mass difference, unbalanced OT can reveal surprisingly low-dimensional structures that are invisible in Euclidean space. For example, consider a dataset in $\mathcal{M}_+(\Omega)$, $\Omega \subset \mathbb{R}^2$, given by translated indicator functions of the unit disk $\{\mathbb{1}_D(\cdot - t_i)\}_{i=1}^N$. As a subset of $(\mathcal{M}_+(\Omega), \text{HK}_\kappa)$, this is a 2-dimensional manifold (when κ is suitably large) that is nearly isometric to the translation set $\{t_i\}_{i=1}^N \subset \mathbb{R}^2$ via the MDS embedding (see [11] for similar reasoning for W_2) due to the fact that $\text{HK}_\kappa(\mathbb{1}_D(\cdot - t_i), \mathbb{1}_D(\cdot - t_j)) \approx c|t_i - t_j|$ for large κ . On the other hand, if these were objects imaged by a camera or similar imaging device by being mapped to an n -pixel grid, the collection of images $\{x_i\}_{i=1}^N \subset \mathbb{R}^n$ would be typically full rank (depending on the spacing of the translation set), and would not be concentrated near a 2-dimensional manifold. Examples like this suggest that modeling data using the HK geometry can yield a better low-dimensional model, and potentially lead to improved task performance, which is the main purpose of our study here.

3 Datasets

We use 5 benchmark datasets for both classification and clustering experiments: MNIST, Coil-20 [20], and for more sophisticated image types, we use BloodMNIST, OrganCMNIST, and RetinaMNIST from the MedMNIST v2 benchmark [17]. These are more varied in terms of texture, morphology, and complexity compared to the basic benchmarks, but are nonetheless curated, so they provide a fruitful testing ground for the use of UOT metrics in neighbor embeddings.

Each of the datasets we use from MedMNIST are from the set of 2D images, all of which have resolution 28×28 after preprocessing [17]. BloodMNIST is a collection of 17,092 preprocessed images of 8 types of blood cells (basophils, eosinophils, erythroblasts, immature granulocytes (promyelocytes, myelocytes, and metamyelocytes), lymphocytes, monocytes, neutrophils and platelets), where the classification task is to determine what type of cell the image contains. OrganCMNIST comprises 23,660 2D coronal plane images from 3D CT scans from 11 different organs, and the task is to determine which images belong to which organ. RetinaMNIST consists of 1,600 images of human retinas with 4 different grades of retinopathy caused by diabetes, where the task is to classify the grade of retinopathy.

4 Methodology

In all experiments, we utilize a manifold learning pipeline of the form data \rightarrow neighbor embedding \rightarrow task, where the tasks are classification and clustering. We form embeddings via MDS, Isomap, t-SNE, and Laplacian eigenmaps using Euclidean, W_2 , and HK_1 metrics. We exactly compute W_2 distances using the python optimal transport package [21], and use the code from [22] to compute HK distances.

For classification, the embedded data in the low-dimensional Euclidean space is split into 80%-20% training and testing data, and we use the following classification algorithms on the train-test

split in the embedded space: Linear Discriminant Analysis (LDA), k-Nearest Neighbors (KNN) for $k = 1, 3, 5$, linear Support Vector Machine (SVM(L)), RBF kernel SVD (SVM(R)), Random Forest, and Multinomial Logistic Regression (MLR). To cluster embedded data, we use k -means and spectral clustering. To compute the clustering accuracy, linear sum assignment is used to match the ground truth labels with the cluster labels.

To determine which embedding outperforms the others (if any), we use statistical hypothesis tests to compare the mean of 10 replicate experiments for each of the Euclidean, OT and UOT-based embeddings. This is done by carrying out 6 hypothesis tests formulated with **Null Hypothesis**: The task accuracy using the embedding metric A is less than or equal to that using metric B , and **Alternative Hypothesis**: The task accuracy using the embedding metric A is better than that using metric B . We use a significance level $p = 0.05$ for all hypothesis tests, and run two one-sided t-tests for each pair $A, B \in \{\text{Euclidean}, W_2, \text{HK}_1\}$ (where the roles of A and B are reversed in the second test).

The embedding dimension is a hyperparameter for neighbor embeddings, and for our experiments, we follow the heuristic of [23], based on the fact that local Singular Value Decomposition (SVD) provides a good estimation for the dimension of a tangent plane to data in a small neighborhood, and use the SVD to compute the embedding dimension as the integer that accounts for a fixed proportion of the variance of the spectrum, i.e., choose the minimum integer n so that

$$\frac{\sum_{i=1}^n \sigma_i^2}{\sum_{j=1}^N \sigma_j^2} \geq a, \quad (4)$$

for fixed $a \in (0, 1)$.

We used 1,000 data points for each experiment for each dataset. We compute the embedding dimension via (4), typically with the value of $a = .97$, however, for some datasets, this leads to a very small embedding dimension (e.g., 1), so for these we augment a somewhat. For MNIST, Coil-20, and OrganCMNIST, we use $a = .97$ corresponding to embedding dimensions 129, 66, and 120, respectively. For BloodMNIST and RetinaMNIST, we use $a = .99$ and $a = .999$, respectively, corresponding to embedding dimensions 21 and 80. Note that in using the sklearn implementation of t-SNE, data is embedded into dimension 3 for all experiments.

5 Clustering experiment results

Table 1 reports the clustering accuracy with columns being datasets and rows being the clustering algorithms (k -means or spectral clustering) applied to data embedded by MDS, Isomap, t-SNE or Eigenmaps. Bold entries mean that the given metric outperformed all others at confidence level $p = .05$, and bars over Euclidean or UOT values mean they outperformed the other, but did not significantly outperform OT.

Performance of most methods is relatively similar on MNIST and Coil-20, though UOT tends to perform better for spectral clustering on MDS embeddings. Additionally, UOT frequently outperforms OT in these datasets. Indeed, UOT outperforms OT 70% (28/40) of the time. For the MedMNIST datasets, UOT outperforms the other metrics much of the time (58% of the trials), and more interestingly, most of the time (83% of trials) outperforms OT by a large magnitude.

The MedMNIST clustering results seem promising in that UOT almost always outperforms OT. Given that these datasets involve various textures and morphologies, it seems that the ability of UOT to create and destroy mass allows for the manifold learning methods to find a more concrete structure, whereas for less texture-based data like MNIST and Coil-20, OT sometimes performs comparably to UOT.

Method	Algorithm	MNIST	Coil-20	BloodMNIST	OrganCMNIST	RetinaMNIST
		Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT
MDS	<i>k</i> -Means	0.51 0.43 0.49	0.64 0.50 0.56	0.45 0.33 0.45	0.39 0.31 0.36	0.34 0.28 0.32
	Spectral	0.60 0.49 0.64	0.67 0.68 0.69	0.48 0.34 0.52	0.34 0.31 0.38	0.36 0.29 0.31
Isomap	<i>k</i> -means	0.57 0.59 0.59	0.69 0.61 0.65	0.45 0.36 0.43	0.34 0.33 0.29	0.30 0.28 0.38
	Spectral	0.59 0.50 0.58	0.52 0.56 0.61	0.35 0.34 0.30	0.38 0.34 0.44	0.35 0.29 0.40
t-SNE	<i>k</i> -means	0.57 0.62 0.62	0.72 0.70 0.72	0.44 0.37 0.44	0.41 0.38 0.45	0.30 0.26 0.32
	Spectral	0.59 0.62 0.58	0.48 0.42 0.41	0.42 0.36 0.45	0.41 0.36 0.45	0.42 0.37 0.42
Eigenmaps	<i>k</i> -means	0.38 0.49 0.49	0.44 0.65 0.57	0.42 0.32 0.46	0.36 0.32 0.40	0.32 0.32 0.37
	Spectral	0.38 0.49 0.44	0.51 0.52 0.44	0.36 0.31 0.41	0.37 0.33 0.38	0.31 0.36 0.33

Table 1: Clustering results for all datasets applying *k*-means and spectral clustering algorithms on data embedded via MDS, Isomap, t-SNE, or Laplacian Eigenmaps. In each cell, the metric is Euclidean (left), OT (middle) and UOT (right).

6 Classification experiment results

Synthesizing the classification results in Tables 2–6, we see the following trends. For metric MDS embeddings, UOT typically outperforms both Euclidean and OT by a substantial margin, with an exception of RetinaMNIST. For Isomap embeddings, which are designed to capture global nonlinear manifolds, UOT typically outperforms both other metrics on OrganCMNIST and RetinaMNIST, while results are mixed for the rest of the datasets. That said, UOT typically performs better than OT on BloodMNIST.

For t-SNE embeddings, the results vary quite a lot from dataset to dataset. In some instances Euclidean embeddings are better, including Coil-20, while the other datasets typically show no overall winner. However, oftentimes, the UOT results are still much better than the OT results. Eigenmap embeddings also yield varied outcomes, but notably the majority of the time, UOT performs as well or better than OT.

	MDS	Isomap	t-SNE	Eigenmaps
Metric	Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT
LDA	0.81 0.83 0.79	0.84 0.86 0.86	0.70 0.55 0.68	0.87 0.90 0.92
1NN	0.84 0.79 0.90	0.84 0.86 0.87	0.87 0.76 0.88	0.69 0.87 0.87
3NN	0.83 0.80 0.89	0.84 0.87 0.86	0.87 0.77 0.88	0.56 0.89 0.88
5NN	0.84 0.81 0.89	0.83 0.86 0.85	0.85 0.76 0.87	0.49 0.89 0.88
SVM (L)	0.86 0.20 0.87	0.84 0.83 0.88	0.76 0.63 0.73	0.12 0.12 0.12
SVM (R)	0.85 0.90 0.90	0.87 0.90 0.89	0.82 0.72 0.80	0.85 0.90 0.90
RF	0.82 0.82 0.88	0.83 0.87 0.87	0.87 0.75 0.87	0.86 0.88 0.89
MLR	0.80 0.41 0.82	0.83 0.81 0.85	0.72 0.59 0.68	0.12 0.24 0.17

Table 2: Classification results for MNIST on Euclidean (left), OT (middle) and UOT (right) neighbor embeddings.

All in all, these tables represent 160 classification experiments. The Euclidean metric outperforms both others in 25/160 trials (or 16% of the time); UOT outperforms both other metrics in 56/160 trials (or 35% of the time), and OT outperforms the others 3/160 (or 2% of the time). Interestingly, we find that UOT outperforms OT in 110/160 trials, or 69% of the time, and UOT outperforms the Euclidean metric in 100 trials (or 45% of the time). Taken together, these results show that using UOT metrics for neighbor embeddings appear capable of finding more useful low-dimensional embeddings of image data.

Of particular note, we see that on the MedMNIST datasets, which are more in line with types of data that previous work on OT embeddings have been done, UOT appears to give much better results consistently. Indeed, the Euclidean metric outperforms the rest only 9% of the time, while

Classifier	MDS	Isomap	t-SNE	Eigenmaps
Metric	Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT
LDA	0.93 0.77 0.90	0.95 0.95 0.95	0.79 0.74 0.75	0.98 0.10 0.09
1NN	0.94 0.90 0.98	0.94 0.96 <u>0.96</u>	0.99 0.90 0.90	0.97 0.91 0.89
3NN	0.92 0.86 0.95	0.89 0.93 0.94	0.97 0.88 0.90	0.92 0.89 0.86
5NN	0.89 0.83 0.93	0.86 0.92 <u>0.90</u>	0.95 0.86 0.87	0.89 0.86 0.86
SVM(L)	0.97 0.07 0.99	0.99 0.84 0.99	0.85 0.82 0.85	0.04 0.07 0.05
SVM(R)	0.97 0.87 0.95	0.91 0.94 0.91	0.86 0.78 0.80	0.97 0.88 0.87
RF	0.92 0.87 0.98	0.97 0.97 0.98	0.98 0.90 0.90	0.97 0.92 0.90
MLR	0.95 0.24 0.94	0.98 0.84 0.98	0.80 0.74 0.77	0.04 0.15 0.12

Table 3: Classification results for Coil-20 on Euclidean (left), OT (middle) and UOT (right) neighbor embeddings.

Classifier	MDS	Isomap	t-SNE	Eigenmaps
Metric	Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT
LDA	0.62 0.52 0.62	0.59 0.57 0.61	0.54 0.49 0.56	0.66 0.57 0.67
1NN	0.54 0.43 0.62	0.54 0.47 0.56	0.61 0.46 0.65	0.55 0.49 0.60
3NN	0.53 0.44 0.62	0.56 0.50 0.57	0.63 0.47 0.64	0.58 0.50 0.62
5NN	0.55 0.46 0.63	0.58 0.51 0.58	0.63 0.48 0.63	0.60 0.52 0.62
SVM(L)	0.63 0.21 0.63	0.58 0.32 0.58	0.57 0.51 0.58	0.21 0.19 0.20
SVM(R)	0.63 0.56 0.72	0.65 0.58 0.64	0.59 0.53 0.61	0.68 0.56 0.68
RF	0.62 0.51 0.70	0.63 0.57 0.63	0.63 0.50 0.65	0.66 0.56 0.67
MLR	0.64 0.21 0.65	0.58 0.44 0.61	0.55 0.50 0.57	0.21 0.20 0.22

Table 4: Classification results for BloodMNIST on Euclidean (left), OT (middle) and UOT (right) neighbor embeddings.

Classifier	MDS	Isomap	t-SNE	Eigenmaps
Metric	Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT
LDA	0.67 0.55 0.61	0.68 0.61 0.71	0.48 0.44 0.55	0.75 0.66 0.76
1NN	0.52 0.53 0.75	0.66 0.65 0.71	0.75 0.53 0.72	0.65 0.65 0.75
3NN	0.47 0.52 0.70	0.63 0.62 0.70	0.70 0.52 0.67	0.53 0.64 0.71
5NN	0.51 0.53 0.70	0.64 0.62 0.71	0.70 0.53 0.67	0.47 0.64 0.67
SVM(L)	0.63 0.23 0.67	0.68 0.38 0.71	0.52 0.44 0.59	0.25 0.26 0.25
SVM(R)	0.63 0.61 0.72	0.71 0.66 0.71	0.64 0.50 0.60	0.72 0.70 0.76
RF	0.66 0.60 0.76	0.69 0.64 0.74	0.70 0.53 0.70	0.75 0.67 0.77
MLR	0.63 0.23 0.63	0.71 0.38 0.73	0.49 0.44 0.54	0.26 0.25 0.26

Table 5: Classification results for OrganCMNIST on Euclidean (left), OT (middle) and UOT (right) embeddings.

Classifier	MDS	Isomap	t-SNE	Eigenmaps
Metric	Euc OT UOT	Euc OT UOT	Euc OT UOT	Euc OT UOT
LDA	0.48 0.45 0.45	0.46 0.43 0.51	<u>0.47</u> 0.45 0.42	0.44 0.41 0.44
1NN	0.43 0.36 0.40	0.41 0.38 0.43	0.44 0.36 0.40	0.44 0.36 0.39
3NN	0.45 0.42 0.43	0.45 0.41 0.48	0.44 0.40 0.44	0.44 0.42 0.44
5NN	0.47 0.44 0.44	0.44 0.43 0.50	0.45 0.41 0.44	0.46 0.42 0.47
SVM(L)	0.47 0.46 0.46	0.47 0.44 0.51	0.45 0.45 0.44	0.45 0.44 0.44
SVM(R)	<u>0.47</u> 0.47 0.45	0.46 0.46 0.51	<u>0.46</u> 0.45 0.41	0.47 0.45 0.46
RF	<u>0.48</u> 0.45 0.45	0.46 0.45 0.51	0.49 0.40 0.42	0.47 0.44 0.45
MLR	0.47 0.56 0.45	0.46 0.44 0.50	<u>0.48</u> 0.45 0.44	0.45 0.44 0.44

Table 6: Classification results for RetinaMNIST on Euclidean (left), OT (middle) and UOT (right) embeddings.

UOT outperforms both other metrics 43% of the time, and UOT outperforms OT 81% of the time. Taken together, these experiments indicate that the Hellinger–Kantorovich metrics are finding a more meaningful low-dimensional structure in these datasets, especially compared to standard OT.

7 Conclusion

In this work, we studied how the choice of metric when modeling image data geometry affects task performance in a neighbor embedding framework. In particular, we showed that using the Hellinger–Kantorovich metric from unbalanced optimal transport yields better classification and clustering performance compared to OT-based embeddings a majority of the time, especially on medical imaging benchmarks from the MedMNIST datasets. We also showed that UOT embeddings yield performance improvements over standard Euclidean embeddings some of the time. These results indicate that the constraints of optimal transport, viz. having total mass 1, are too rigid to account for datasets with varying masses and complex structures. We attribute the success of the UOT embeddings to the ability of the HK metrics to create and destroy mass at a distance. It would be interesting to further explore exactly what facets of data geometry are better captured in the HK metrics. Based on our experiments, we suspect that HK metrics are able to account for varied textures within images (evidenced by success on BloodMNIST) as well as variations in overall morphology (as in OrganCMNIST).

Acknowledgments

Both authors were partially supported by a Research Enhancement Program grant from the College of Science at the University of Texas at Arlington and by the Army Research Office and was accomplished under Grant Number W911NF-23-1-0213. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

- [1] Gunnar Carlsson, “Topology and data,” *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.
- [2] Joan Bruna and Stéphane Mallat, “Invariant scattering convolution networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.
- [3] Marina Meilă and Hanyu Zhang, “Manifold learning: What, how, and why,” *Annual Review of Statistics and Its Application*, vol. 11, 2024.
- [4] Mikhail Belkin and Partha Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [5] Ronald R Coifman and Stéphane Lafon, “Diffusion maps,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [6] Kanti V Mardia, John T Kent, and Charles C Taylor, *Multivariate analysis*, John Wiley & Sons, 2024.
- [7] Laurens van der Maaten and Geoffrey Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, 2008.

- [8] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger, “UMAP: Uniform manifold approximation and projection,” *Journal of Open Source Software*, vol. 3, no. 29, 2018.
- [9] Joshua B Tenenbaum, Vin De Silva, and John C Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [10] Mira Bernstein, Vin De Silva, John C Langford, and Joshua B Tenenbaum, “Graph approximations to geodesics on embedded manifolds,” Tech. Rep., Citeseer, 2000.
- [11] Keaton Hamm, Nick Henscheid, and Shujie Kang, “Wassmap: Wasserstein isometric mapping for image manifold learning,” *SIAM Journal on Mathematics of Data Science*, vol. 5, no. 2, pp. 475–501, 2023.
- [12] Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K Rohde, “Optimal mass transport: signal processing and machine-learning applications,” *IEEE signal processing magazine*, vol. 34, no. 4, pp. 43–59, 2017.
- [13] Wei Wang, John A Ozolek, Dejan Slepčev, Ann B Lee, Cheng Chen, and Gustavo K Rohde, “An optimal transportation approach for nuclear structure-based pathology,” *IEEE transactions on medical imaging*, vol. 30, no. 3, pp. 621–631, 2010.
- [14] Gabriel Peyré and Marco Cuturi, “Computational optimal transport: with applications to data science,” *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.
- [15] James C Mathews, Maryam Pouryahya, Caroline Moosmüller, Yannis G Kevrekidis, Joseph O Deasy, and Allen Tannenbaum, “Molecular phenotyping using networks, diffusion, and topology: soft tissue sarcoma,” *Scientific reports*, vol. 9, no. 1, pp. 13982, 2019.
- [16] Matthias Liero, Alexander Mielke, and Giuseppe Savaré, “Optimal transport in competition with reaction: The hellinger–kantorovich distance and geodesic curves,” *SIAM Journal on Mathematical Analysis*, vol. 48, no. 4, pp. 2869–2911, 2016.
- [17] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni, “Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification,” *Scientific Data*, vol. 10, no. 1, pp. 41, 2023.
- [18] Marco Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” *Advances in neural information processing systems*, vol. 26, 2013.
- [19] Mauro Bonafini, Olga Minevich, and Bernhard Schmitzer, “Hellinger–kantorovich barycenter between dirac measures,” *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 29, pp. 19, 2023.
- [20] Samer A. Nene, Shree K. Nayar, and Hiroshi Murase, “Columbia object image library (coil-20),” Tech. Rep. CUCS-005-96, Department of Computer Science, Columbia University, February 1996.
- [21] Rémi Flamary et al., “Pot: Python optimal transport,” *Journal of Machine Learning Research*, vol. 22, no. 78, pp. 1–8, 2021.
- [22] Tianji Cai, Junyi Cheng, Bernhard Schmitzer, and Matthew Thorpe, “The linearized hellinger–kantorovich distance,” *SIAM Journal on Imaging Sciences*, vol. 15, no. 1, pp. 45–83, 2022.

- [23] F Patricia Medina, Linda Ness, Melanie Weber, and Karamatou Yacoubou Djima, “Heuristic framework for multiscale testing of the multi-manifold hypothesis,” *Research in Data Science*, pp. 47–80, 2019.