

# HYPERBOLIC DISTANCE BASED ON EMD AND DIFFUSION FOR HYPERSPECTRAL IMAGING

Elad Lavi<sup>\*1</sup>, Amir Bourvine<sup>\*1</sup>, Ya-Wei Eileen Lin<sup>2</sup>, Ronen Talmon<sup>2</sup>

<sup>1</sup>Taub Faculty of Computer Science, Technion

<sup>2</sup>Viterbi Faculty of Electrical and Computer Engineering, Technion

## ABSTRACT

In this paper, we introduce EMD-Based Hyperbolic Diffusion Distance (EMD-HDD), a new method for constructing a meaningful distance metric for hierarchical data with latent hierarchical structure. Our method relies on hyperbolic geometry, diffusion geometry, and the Earth Mover’s Distance (EMD). Specifically, our method embeds data points into a product manifold of hyperbolic spaces, allowing us to recover the hidden hierarchical structure encoded by the mutual relationships between features. We demonstrate the effectiveness of EMD-HDD through experiments on five hyperspectral imaging datasets, showcasing its capability to capture and reveal the intrinsic hierarchical structures inherent in such data.

**Index Terms**— Hyperbolic geometry, Earth Mover’s Distance, Manifold Learning, Hyperspectral Imaging

## 1. INTRODUCTION

Hierarchical high-dimensional data is prevalent in various fields. For example, gene expression profiles exhibit hierarchical structures formed by interactions among genes [1]. Similarly, in natural language processing, words, sentences, or documents form hierarchical structures based on linguistic categories and subcategories [2]. In social network analysis, user interactions create hierarchical relations that influence behavior and information spread [3]. Despite its prevalence, finding a meaningful hierarchical distance for such data presents two key challenges. First, the hierarchical structure of data is often hidden. Second, this intrinsic hierarchical structure usually depends on the mutual relationships between the features, which need to be inferred and incorporated into the development of a meaningful hierarchical distance.

Recently, hyperbolic geometry has emerged as a powerful tool for hierarchical representation learning due to its exponential growth property [4]. It has been effectively utilized across a wide range of fields, e.g., natural language processing [5], social networks [6], and bioinformatics [7–9]. Typically, methods utilizing hyperbolic geometry aim to find hyperbolic embedding and distance of data based on a known hierarchical graph structure [10]. However, the problem we address here differs fundamentally: we focus on finding a distance for hierarchical data based only on the sample and feature mutual relationships without explicitly known structures.

In this paper, we present a new hierarchical distance recovery method for this purpose. Our method builds on diffusion geome-

try [11], a manifold learning technique for capturing the geometric manifold structure underlying high-dimensional data, and Earth Mover’s Distance (EMD), which inherently incorporates the mutual relationships between features to establish a meaningful distance between data [12]. Specifically, we construct a diffusion operator based on the EMD to learn the geometry of the data while incorporating feature mutual relations. Then, by using the EMD-based diffusion operator, we embed the data in a product manifold of hyperbolic spaces [13] based on a multi-scale diffusion process. This embedding recovers the latent hierarchical structure underlying the data, and consequently, it induces a meaningful hierarchical distance [14]. We term this new distance EMD-based Hyperbolic Diffusion Distance (EMD-HDD).

We show application to hyperspectral imaging (HSI) analysis, an area where capturing pixel hierarchy across spectral bands is critically important [15, 16]. Our method is particularly suited to HSI data, with its inherent high dimensionality and the dominant mutual relationships among spectral bands. This type of data is widely used across various fields, including agriculture and healthcare, where effective analysis can lead to significant advancements [17]. Our experiments on several HSI datasets demonstrate that the proposed EMD-HDD significantly improves classification performance in HSI tasks compared to existing methods. In addition, we conduct ablation studies to explore the contributions of each component in our method. We find that using other distance metrics like the Euclidean distance, without accounting for feature (spectral band) relations, results in poor recovery of the hierarchical distances between HSI samples (pixels). Moreover, EMD alone does not effectively capture the hierarchical structure underlying HSI data.

## 2. BACKGROUND

We briefly present the relevant background to our proposed method.

**Diffusion Geometry.** Diffusion maps [11] is a manifold learning method that captures the intrinsic low-dimensional manifold structure of high-dimensional data. It is based on analyzing similarity patterns among data points, leading to an embedding that represents the underlying manifold. The key to this method is a concept known as diffusion geometry, which revolves around the construction of random walks (or “diffusions”) on weighted graphs that are built from the data. Below we present a brief overview of diffusion geometry.

Consider a set of high-dimensional data points  $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N$ , assumed to lie on a low-dimensional manifold  $\mathcal{M}$  embedded in the ambient space  $\mathbb{R}^m$ . The affinity matrix  $\mathbf{M} \in \mathbb{R}^{N \times N}$  is constructed by  $\mathbf{M}(i, j) = \exp(-d^2(i, j)/\epsilon)$ , where  $d(i, j)$  represents a suitable distance between data points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , and  $\epsilon$  is a scaling parameter, which is typically set as the median distance between data points

<sup>\*</sup> Equal contribution. The authors would like to thank Nimrod Peleg, Yair Moshe, and the Signal and Image Processing Lab (SIPL) staff for their support and helpful suggestions. The work was supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 802735-ERC-DIFFOP.

multiplied by some constant. This affinity matrix  $\mathbf{M}$  can be viewed as the weight matrix of the edges of an undirected weighted graph  $G = (\mathcal{X}, \mathbf{M})$ , whose node set is the dataset  $\mathcal{X}$ . The affinity matrix is then normalized, generating the diffusion operator  $\mathbf{P}$  by

$$\mathbf{P} = \mathbf{M}\mathbf{D}^{-1}, \quad (1)$$

where  $\mathbf{D}$  is a diagonal matrix whose diagonal elements are given by  $\mathbf{D}(i, i) = \sum_j \mathbf{M}(i, j)$ . It is important to note that the diffusion operator  $\mathbf{P}$  is a column-stochastic matrix, and therefore, it can be interpreted as the transition matrix of a Markov chain on the graph. The diffusion operator  $\mathbf{P}$  can be utilized to propagate densities between the nodes on  $G$ . A vector  $\mathbf{p} \in \mathbb{R}^N$  is considered as a density on the graph  $G$  if  $\mathbf{p}(i) \geq 0$  for all  $i \in [N]$  and  $\|\mathbf{p}\|_1 = 1$ . Then, the vector  $\mathbf{p}^t = \mathbf{P}^t \boldsymbol{\delta}_i$  is the propagated diffusion density after  $t > 0$  diffusion time steps, where  $\boldsymbol{\delta}_i \in \mathbb{R}^N$  is the indicator vector of the  $i$ -th node, initially concentrating the entire density at that node.

The diffusion operator was shown to have desirable convergence properties [11]. Specifically, as the number of data points  $N \rightarrow \infty$  and the scaling parameter  $\epsilon \rightarrow 0$ , the operator  $\mathbf{P}^{t/\epsilon}$  converges point-wise to the Neumann heat kernel of the underlying manifold  $\mathcal{M}$ . That is, the diffusion operator serves as a discrete approximation of the continuous heat kernel, which was shown to effectively capture the geometric properties of the manifold [18].

**Hyperbolic Geometry.** Hyperbolic geometry is a non-Euclidean Riemannian geometry characterized by constant negative curvature. In our work, we consider the  $N$ -dimensional Poincaré half-space model in hyperbolic geometry, where the curvature is fixed at  $-1$  [19]. The  $N$ -dimensional Poincaré half-space is defined as  $\mathbb{H}^N = \{\mathbf{x} \in \mathbb{R}^N | \mathbf{x}(N) > 0\}$ , with the Riemannian metric given by  $ds^2 = \frac{dx^2(1) + \dots + dx^2(N)}{\mathbf{x}^2(N)}$ . This metric reflects the exponential growth property inherent to hyperbolic space in the  $N$ -th coordinate. Given  $\mathbf{x}, \mathbf{y} \in \mathbb{H}^N$ , the Riemannian distance is defined by

$$d_{\mathbb{H}^N}(\mathbf{x}, \mathbf{y}) = 2 \sinh^{-1} \left( \frac{\|\mathbf{x} - \mathbf{y}\|_2}{\sqrt{\mathbf{x}(N)\mathbf{y}(N)}}} \right), \quad (2)$$

where  $\|\cdot\|_2$  is the Euclidean norm.

**Earth Mover's Distance.** The earth mover's distance (EMD), also known as the Wasserstein distance of order 1, is a powerful tool for computing a geometrically meaningful distance between discrete probability distributions [20–22]. It has been demonstrated to be effective in various applications, including image retrieval [12], brain imaging [23], and generative models [24].

Consider two discrete distributions  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}_+^n$  s.t.  $\|\boldsymbol{\mu}_1\|_1 = \|\boldsymbol{\mu}_2\|_1 = 1$ , the transport plan  $\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$  [21] is defined as

$$\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \{\boldsymbol{\gamma} \in \mathbb{R}_+^{n \times n} | \boldsymbol{\gamma} \mathbf{1}_n = \boldsymbol{\mu}_1, \boldsymbol{\gamma}^\top \mathbf{1}_n = \boldsymbol{\mu}_2\}, \quad (3)$$

where  $\mathbf{1}_n = [1, \dots, 1] \in \mathbb{R}^n$ . Given a ground pairwise distance matrix  $\mathbf{C} \in \mathbb{R}^{n \times n}$  between the  $n$  coordinates, the optimal transport (OT) [20] problem seeks to optimize the transport plan  $\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$  that minimizes the cost of transporting  $\boldsymbol{\mu}_1$  to  $\boldsymbol{\mu}_2$  according to the ground distance  $\mathbf{C}$ . The EMD between the two discrete distributions  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$  is then defined as

$$\mathbf{W}_C(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \min_{\boldsymbol{\gamma} \in \Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \langle \mathbf{C}, \boldsymbol{\gamma} \rangle_F, \quad (4)$$

where  $\langle \cdot, \cdot \rangle_F$  is the Frobenius inner product.

### 3. PROPOSED METHOD

Consider a hyperspectral image represented as a three-dimensional tensor  $\mathbf{S} \in \mathbb{R}^{n_x \times n_y \times n_b}$ , where  $n_x$  and  $n_y$  are the spatial dimen-

sions, and  $n_b$  is the number of spectral bands. Let  $\mathbf{s}_i \in \mathbb{R}^{n_b}$  denote the  $i$ -th pixel of the hyperspectral image  $\mathbf{S}$  for  $i = 1, \dots, N = n_x n_y$ , and let  $\mathcal{S} = \{\mathbf{s}_i\}_{i=1}^N$  be the collection of all the pixels in the image. In addition, let  $\mathbf{q}_l \in \mathbb{R}^N$  denote the  $l$ -th feature (spectral band) of  $\mathbf{S}$  for  $l = 1, \dots, n_b$ , such that  $\mathbf{q}_l$  is a column vector of the  $l$ -th frontal slice of the tensor  $\mathbf{S}$ , and let  $\mathcal{Q} = \{\mathbf{q}_l\}_{l=1}^{n_b}$  be the collection of all the features (spectral bands) of the image. We assume that the pixels (i.e., samples)  $\mathcal{S} = \{\mathbf{s}_i\}_{i=1}^N$  lie in a latent hierarchical metric space  $(\mathcal{T}, d_{\mathcal{T}})$ , where both the space  $\mathcal{T}$  and the hierarchical distance  $d_{\mathcal{T}}$  are not explicitly known and depend on the mutual relations between the features  $\mathcal{Q} = \{\mathbf{q}_l\}_{l=1}^{n_b}$ . Our goal is to find embedding of the pixels  $\{\mathbf{s}_i\}_{i=1}^N$  into an intrinsic representation space such that the distance between the embedded pixels recovers the hidden hierarchical distance  $d_{\mathcal{T}}$ .

#### 3.1. Constructing a Diffusion Operator Using EMD

To capture the geometric structure underlying the pixels, we construct a diffusion operator based on EMD. To this end, we first normalize the pixels  $\mathcal{S} = \{\mathbf{s}_i\}_{i=1}^N$  into discrete histograms  $\tilde{\mathbf{s}}_i = \mathbf{s}_i / \|\mathbf{s}_i\|_1 \in \mathbb{R}^{n_b}$  for all  $i \in [N]$ . In the context of the HSI dataset, each pixel  $\mathbf{s}_i$  corresponds to a vector of intensities across spectral bands. Given that these pixel values are non-negative, they can be naturally normalized to histograms, which is particularly useful when the distribution of intensities across the spectral bands is more important than the absolute values of the intensities. Indeed, HSI data is usually normalized as part of a standard preprocessing [25]. With the normalized histograms  $\{\tilde{\mathbf{s}}_i\}_{i=1}^N$  at hand, we compute the EMD between each pair of pixels by

$$\mathbf{W}_C(\tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j) = \min_{\boldsymbol{\gamma} \in \Pi(\tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j)} \langle \mathbf{C}, \boldsymbol{\gamma} \rangle_F, \quad \forall i, j \in [N], \quad (5)$$

where the ground distance  $\mathbf{C} \in \mathbb{R}^{n_b \times n_b}$  is the Euclidean distance between the features (i.e., spectral bands). Specifically, the Euclidean distance between the features  $\mathbf{q}_l$  and  $\mathbf{q}_g$  is computed by  $\mathbf{C}(l, g) = \|\mathbf{q}_l - \mathbf{q}_g\|_2$ . For brevity, we organize these pairwise EMDs into a matrix  $\mathbf{W}_C \in \mathbb{R}^{N \times N}$  whose  $(i, j)$ -th element is  $\mathbf{W}_C(i, j) := \mathbf{W}_C(\tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j)$ . Next, we construct a pairwise affinity matrix using a Gaussian kernel with the EMD by

$$\mathbf{M}_{\text{EMD}}(i, j) = \exp \left( -\frac{\mathbf{W}_C^2(i, j)}{\epsilon} \right), \quad (6)$$

where  $\epsilon$  is a tunable kernel scale parameter. Following Sec. 2, the diffusion operator  $\mathbf{P}_{\text{EMD}} \in \mathbb{R}^{N \times N}$  is then constructed by

$$\mathbf{P}_{\text{EMD}} = \tilde{\mathbf{D}}^{-1} \mathbf{M}_{\text{EMD}}, \quad (7)$$

where  $\tilde{\mathbf{D}}$  is a diagonal matrix with the diagonal elements  $\tilde{\mathbf{D}}(i, i) = \sum_j \mathbf{M}_{\text{EMD}}(i, j)$ .

We posit that the diffusion operator  $\mathbf{P}_{\text{EMD}}$  encodes the underlying pixel structure by accounting for the spectral band relations, as the EMD matrix  $\mathbf{W}_C$  maps the mutual relations between the features to a geometrically meaningful distance between the pixels [22]. This intuitive argument has the following theoretical support. When the pairwise affinities are based on the Euclidean norm rather than the squared EMD, it was shown that the finite diffusion operator  $\mathbf{P}^{t/\epsilon}$  (see Sec. 2) converges to the Neumann heat kernel of the underlying manifold [11]. Recently, [26] extended this result, demonstrating that the limiting differential operator of a diffusion operator constructed using any norm, including the squared EMD as in our case, also exhibits a similar convergence. Therefore,  $\mathbf{P}_{\text{EMD}}$  indeed represents the geometric manifold structures of the pixels.

### 3.2. Recovering the Hierarchical Structure

To recover the latent hierarchical structure underlying the pixels, we use  $\mathbf{P}_{\text{EMD}}$  to construct an embedding of the pixels into a hyperbolic space. Specifically, we use  $\mathbf{P}_{\text{EMD}}$  to generate a series of propagated densities as follows:  $\mathbf{p}_i^k = \mathbf{P}_{\text{EMD}}^{2^{-k}} \delta_i$ , for each sample  $i \in [N]$  and for diffusion time steps  $t = 2^{-k}$  on a dyadic grid, with  $k \in \mathbb{Z}_0^+$ , following [14]. Subsequently, we compute the element-wise square root of these densities, denoted by  $\psi_i^k = \sqrt{\mathbf{p}_i^k}$ . Although diffusion geometry is typically useful for recovering hidden manifold structures [11], focusing specifically on the propagated densities at exponentially growing diffusion times on a dyadic grid enables us to capture hidden tree-like hierarchical relationships. This approach effectively integrates local information from neighborhoods that grow exponentially around each point, a technique shown to be particularly effective for revealing the underlying tree-like structure [14], as the number of nodes in a tree tends to grow exponentially from the root to the leaves.

Then, we embed every pixel  $i \in [N]$  into a Poincaré half-space  $\mathbb{H}^{N+1}$  according to  $i \mapsto \hat{\mathbf{x}}_i^k = [(\psi_i^k)^\top, 2^{\frac{k}{2}-2}]^\top \in \mathbb{H}^{N+1}$  by concatenating the vector  $\psi_i^k$  with a scalar  $2^{\frac{k}{2}-2}$ . Using Poincaré half-space model for embedding such vectors is natural due to the exponential growth of the metric at the last  $((N+1)$ -th) coordinate.

To account for multiple diffusion time scales simultaneously, we construct a hyperbolic diffusion embedding (HDE) for each pixel  $i \in [N]$ , represented as  $[(\hat{\mathbf{x}}_i^0)^\top, \dots, (\hat{\mathbf{x}}_i^K)^\top]^\top \in \mathcal{H}$ , where  $K \in \mathbb{Z}_0^+$  denotes the maximal scale and  $\mathcal{H} = \mathbb{H}^{N+1} \times \dots \times \mathbb{H}^{N+1}$  represents the product manifold of  $K+1$  Poincaré half-space models.

Finally, we introduce the EMD-based hyperbolic diffusion distance (EMD-HDD), a hierarchical distance between the embedded points in the product manifold  $\mathcal{H}$ . The EMD-HDD is computed using the  $\ell_1$  distance across the product manifold and is defined by

$$\mathbf{H}(i, j) = \sum_{k=0}^K 2 \sinh^{-1} \left( 2^{-\frac{k}{2}+1} \left\| \hat{\mathbf{x}}_i^k - \hat{\mathbf{x}}_j^k \right\|_2 \right), \forall i, j \in [N], \quad (8)$$

where the function  $\sinh^{-1}$  arises from the Riemannian distance in the hyperbolic space  $\mathbb{H}^{N+1}$  in Eq. (2) at each scale. We summarize the proposed EMD-HDD in Alg. 1.

The theoretical foundation of our method stems from the following results. In [14], it was shown that the  $\ell_1$  distance induced by the metric of the product of hyperbolic spaces as in Eq. (8) is equivalent to the hidden distance of the hidden hierarchical space underlying the samples (i.e., pixels in our case). In addition, the diffusion operator  $\mathbf{P}_{\text{EMD}}$  can be viewed as a discrete approximation of the heat kernel on the manifold underlying the pixels (See Sec. 3.1). Therefore, we postulate that the hyperbolic embedding  $\mathcal{H}$  of the pixels constructed based on the diffusion operator  $\mathbf{P}_{\text{EMD}}$  with EMD captures the latent hierarchical structure of the pixels. Specifically, the  $\ell_1$  distance in  $\mathcal{H}$ , induced by the metric of the product of hyperbolic models, approximates the hidden hierarchical distance  $d_{\mathcal{T}}$ . In addition, using the EMD in the construction of the diffusion operator  $\mathbf{P}_{\text{EMD}}$  rather than the Euclidean distance enables us to naturally incorporate the mutual relations between the spectral bands, which are assumed to embody critical information on the structure of the pixels.

## 4. EXPERIMENTAL RESULTS

We demonstrate the advantages of the proposed EMD-HDD in hyperspectral imaging classification tasks. Our method is evaluated on five HSI datasets, including: (i) the Kennedy Space Center (KSC) dataset [27], with image dimensions of  $512 \times 614$  and 176 spectral

---

### Algorithm 1 EMD-based Hyperbolic Diffusion Distance

---

**Input:** The HSI dataset  $\mathcal{S} = \{\mathbf{s}_i\}_{i=1}^N$

**Parameters:** Maximal scale  $K \in [20]$  and diffusion scale  $\epsilon = 2$

**Output:** The EMD-HDD  $\mathbf{H} \in \mathbb{R}^{N \times N}$

**for**  $i \in [N]$  **do**

$\tilde{\mathbf{s}}_i \leftarrow \mathbf{s}_i / \|\mathbf{s}_i\|_1$

**end for**

Compute  $\mathbf{C}$  as the Euclidean distance between the spectral bands

**for**  $i, j \in [N]$  **do**

$\mathbf{W}_{\mathbf{C}}(i, j) \leftarrow \min_{\gamma \in \Pi(\tilde{\mathbf{s}}_i, \tilde{\mathbf{s}}_j)} \langle \mathbf{C}, \gamma \rangle_F$  // Eq. (5)

$\mathbf{M}_{\text{EMD}}(i, j) \leftarrow \exp \left( -\frac{\mathbf{W}_{\mathbf{C}}^2(i, j)}{\epsilon} \right)$  // Eq. (6)

**end for**

$\mathbf{P}_{\text{EMD}} \leftarrow \tilde{\mathbf{D}}^{-1} \mathbf{M}_{\text{EMD}}$  // Eq. (7)

$\mathbf{U} \mathbf{A} \mathbf{V}^\top = \text{eig}(\mathbf{P}_{\text{EMD}})$

**while**  $k \leq K$  **do**

$\hat{\mathbf{X}}_k \leftarrow \left[ \sqrt{(\mathbf{U} \mathbf{A}^{2^{-k}} \mathbf{V}^\top)^\top}, 2^{\frac{k}{2}-2} \mathbf{1}_N \right]^\top$

$k \leftarrow k + 1$

**end while**

**for**  $i, j \in [N]$  **do**

$\mathbf{H}(i, j) \leftarrow \sum_{k=0}^K 2 \sinh^{-1} \left( 2^{-\frac{k}{2}+1} \left\| \hat{\mathbf{x}}_i^k - \hat{\mathbf{x}}_j^k \right\|_2 \right)$  // Eq. (8)

**end for**

---

bands, where pixels are categorized into 13 classes, (ii) the Pavia University (PaviaU) dataset [28], featuring images with dimensions of  $610 \times 340$  and 103 spectral bands, with pixels belonging to nine classes, (iii) the Pavia Centre (PaviaC) dataset [28], consisting of images with dimensions of  $1096 \times 715$  and 102 spectral bands, where pixels are classified into nine classes, (iv) the Greeding Village1 (Greeding) dataset [29], comprising images of size  $670 \times 606$  with 126 spectral bands, where pixels are divided into five classes, and (v) the Botswana dataset [30], including images of size  $1476 \times 256$  with 145 spectral bands, where pixels are associated with 14 classes. Two standard pre-processing steps were applied to HSI datasets: (i) the min-max standardization, and (ii) dividing the images into patches of size  $Q \times Q$  centered at each pixel  $\mathbf{s}_i$ . The ground truth for each patch was established based on a majority vote by assigning the most common label among the patch's pixels.

We compare EMD-HDD in Alg. 1 with five hierarchical embedding methods: (i) the tree representation (TR) [31], (ii) the Poincaré embedding (PE) [5], (iii) a PyTorch (PT) implementation of an SGD-based algorithm optimized over a principal geodesic analysis loss function [10], (iv) the hyperbolic multi-dimensional scaling (hMDS), and (v) the hyperbolic embedding obtained by the hyperbolic hierarchical clustering (HHC) [32]. Following the experimental protocol described in [10], we evaluate the PE, hMDS, and PT methods across embedding spaces of dimensions 2, 5, 10, 50, 100, and 200, reporting the best results for each. In addition, following the common practice, we also present the results of the PE into the 2-dimensional Poincaré disk and denote it by PE-2. For consistency across all methods, we also use a distance metric based on the EMD as the input distance for the competing algorithms.

For quantitative evaluation, we employ k-NN classification for HSI data [27] using the distance metric prescribed by each method. Our EMD-HDD uses the proposed distance in Eq. (8). We use cross-validation with ten repetitions, where the dataset is randomly divided into 70% training set and 30% testing set. The classification accu-

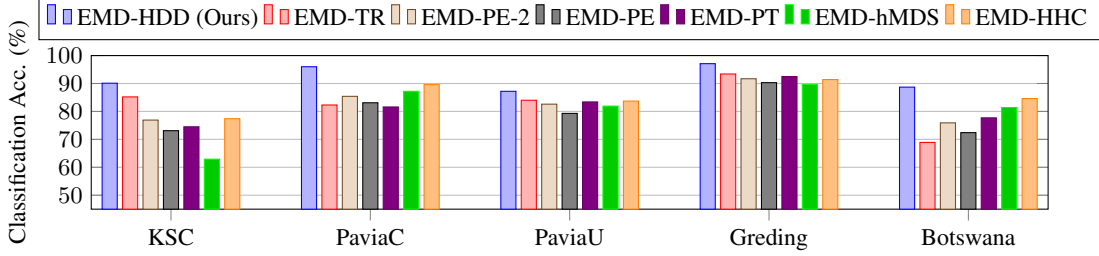


Fig. 1. The HSI classification accuracy based on the hyperbolic representation obtained by the tested methods.

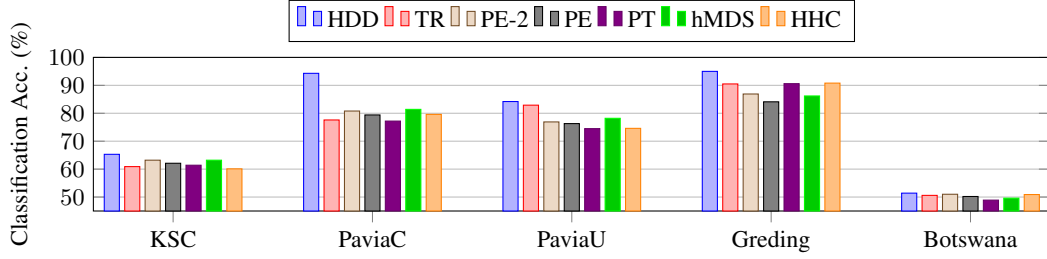


Fig. 2. The HSI classification accuracy based on the hyperbolic representation obtained by the tested methods using Euclidean metric.

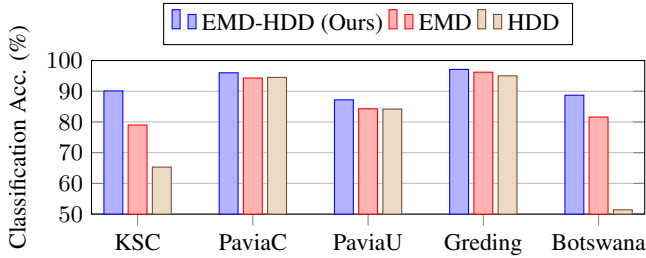


Fig. 3. The HSI classification accuracy using our EMD-HDD compared to EMD alone (without hierarchical distance recovery) and HDD alone (based solely on Euclidean distance without considering feature relations).

accuracy obtained from these experiments is presented in Fig. 1. The results demonstrate that EMD-HDD consistently outperforms all baseline methods by a significant margin across all tested datasets. This performance advantage indicates that EMD-HDD effectively captures the underlying tree-like hierarchical structure underlying HSI data, leading to superior classification accuracy.

To investigate the role of the EMD between the pixels, which incorporates information on the mutual relations between the spectral bands, we repeat the previous HSI classification but use the Euclidean distance between the pixels (without taking into account the spectral bands mutual relationships) instead of the EMD. This Euclidean distance serves as the input metric for the competing algorithms. The classification tasks are performed following the same procedure as in Fig. 1. The classification accuracy results using the Euclidean distance are presented in Fig. 2. The results indicate that while HDD [14] outperforms existing hierarchical distance recovery methods when using Euclidean distance, our EMD-HDD in Alg. 1, as shown in Fig. 1, achieves significantly better performance than all competing methods evaluated in Fig. 2. This result indicates the crit-

icality of the mutual relationships between the spectral bands for the accurate recovery of the hierarchical structures underlying HSI data.

We further conduct an ablation study to assess the effectiveness of our method’s components. In addition to exploring the impact of EMD, we examine the role of the hierarchical representation method in EMD-HDD. Specifically, we perform the HSI classification task directly using the EMD without utilizing it for hierarchical distance recovery. The HSI classification is performed in the same way as in Fig. 1 and Fig. 2. The classification accuracy results obtained by using solely the EMD distance are presented in Fig. 3. We observe that EMD-HDD outperforms both the sole use of EMD and HDD, indicating that, indeed, the use of the EMD accounting for the feature relationship and the hyperbolic representation based on diffusion operator with the EMD has a critical contribution to the extraction of the hierarchical structure.

## 5. CONCLUSIONS

In this paper, we introduced EMD-HDD, a new method that integrates feature relationships to recover hierarchical distances in data where the structure is not explicitly known. We build a diffusion operator based on EMD, which integrates the feature information into an informative distance between samples. This diffusion operator provably encodes the underlying sample structure. We then exploit this diffusion operator to construct a product manifold of hyperbolic spaces, which is specifically designed to reveal the latent hierarchical structure. In addition, we showed that the obtained embedding provably recovers hidden tree-like structures. We apply our method to HSI classification tasks. The empirical results demonstrate the effectiveness of our method, where it consistently outperforms existing baselines by a large margin. Our findings support the existence of a hierarchical structure underlying the pixels in HSI that critically depends on the mutual relationships between the spectral bands. In future work, we plan to extend the method by adding regularization to the EMD, e.g., entropy [33], which has improved efficiency and robustness for noisy data.

## 6. REFERENCES

- [1] A. Tanay and A. Regev, “Scaling single-cell genomics from phenomenology to mechanism,” *Nature*, vol. 541, no. 7637, pp. 331–338, 2017.
- [2] T. Mikolov, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [3] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [4] R. Sarkar, “Low distortion delaunay embedding of trees in hyperbolic plane,” in *International symposium on graph drawing*, pp. 355–366, Springer, 2011.
- [5] M. Nickel and D. Kiela, “Poincaré embeddings for learning hierarchical representations,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [6] K. Verbeek and S. Suri, “Metric embedding, hyperbolic space, and social networks,” in *Proceedings of the thirtieth annual symposium on Computational geometry*, pp. 501–510, 2014.
- [7] J. Ding and A. Regev, “Deep generative model embedding of single-cell RNA-Seq profiles on hyperspheres and hyperbolic spaces,” *Nature communications*, vol. 12, no. 1, p. 2554, 2021.
- [8] Y.-W. E. Lin, Y. Kluger, and R. Talmon, “Hyperbolic Procrustes analysis using Riemannian geometry,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 5959–5971, 2021.
- [9] Y.-W. E. Lin, Y. Kluger, and R. Talmon, “Hyperbolic diffusion procrustes analysis for intrinsic representation of hierarchical data sets,” in *ICASSP 2024*, pp. 6325–6329, IEEE, 2024.
- [10] F. Sala, C. De Sa, A. Gu, and C. Ré, “Representation trade-offs for hyperbolic embeddings,” in *International conference on machine learning*, pp. 4460–4469, PMLR, 2018.
- [11] R. R. Coifman and S. Lafon, “Diffusion maps,” *Applied and computational harmonic analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [12] Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval,” *International journal of computer vision*, vol. 40, pp. 99–121, 2000.
- [13] B. H. Bowditch, “A course on geometric group theory,” *mathematical society of japan, tokyo, 2006. MR2243589 (2007e:20085)*, vol. 16, 2007.
- [14] Y.-W. E. Lin, R. R. Coifman, G. Mishne, and R. Talmon, “Hyperbolic diffusion embedding and distance for hierarchical representation learning,” in *International Conference on Machine Learning*, pp. 21003–21025, PMLR, 2023.
- [15] P. Guccione, L. Mascolo, and A. Appice, “Iterative hyperspectral image classification using spectral–spatial relational features,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3615–3627, 2015.
- [16] P. Ghamsi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, “Advanced spectral classifiers for hyperspectral images: A review,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 8–32, 2017.
- [17] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, “Recent advances of hyperspectral imaging technology and applications in agriculture,” *Remote Sensing*, vol. 12, no. 16, p. 2659, 2020.
- [18] P. W. Jones, M. Maggioni, and R. Schul, “Manifold parametrizations by eigenfunctions of the laplacian and heat kernels,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 1803–1808, 2008.
- [19] A. F. Beardon, *The geometry of discrete groups*, vol. 91. Springer Science & Business Media, 2012.
- [20] G. Monge, “Mémoire sur la théorie des déblais et des remblais,” *Mem. Math. Phys. Acad. Royale Sci.*, pp. 666–704, 1781.
- [21] L. V. Kantorovich, “On the translocation of masses,” in *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, pp. 199–201, 1942.
- [22] C. Villani, *Optimal transport: old and new*, vol. 338. Springer, 2009.
- [23] H. Janati, M. Cuturi, and A. Gramfort, “Spatio-temporal alignments: Optimal transport through space and time,” in *International conference on artificial intelligence and statistics*, pp. 1695–1704, PMLR, 2020.
- [24] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*, pp. 214–223, PMLR, 2017.
- [25] S. Prasad and J. Chanussot, *Hyperspectral image analysis: advances in machine learning and signal processing*. Springer Nature, 2020.
- [26] J. Kileel, A. Moscovich, N. Zelesko, and A. Singer, “Manifold learning with arbitrary norms,” *Journal of Fourier Analysis and Applications*, vol. 27, no. 5, p. 82, 2021.
- [27] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, “Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning,” *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 55–66, 2014.
- [28] X. Huang and L. Zhang, “A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy,” *International Journal of Remote Sensing*, vol. 30, no. 12, p. 3205–3221, 2009.
- [29] W. Gross, D. Tuia, U. Soergel, and W. Middelmann, “Non-linear feature normalization for hyperspectral domain adaptation and mitigation of nonlinear effects,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5975–5990, 2019.
- [30] Y. Li, J. Wang, T. Gao, Q. Sun, L. Zhang, and M. Tang, “Adoption of machine learning in intelligent terrain classification of hyperspectral remote sensing images,” *Computational Intelligence and Neuroscience*, vol. 2020, no. 1, p. 8886932, 2020.
- [31] R. Sonthalia and A. Gilbert, “Tree! I am no Tree! I am a low dimensional hyperbolic embedding,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 845–856, 2020.
- [32] I. Chami, A. Gu, V. Chatziafratis, and C. Ré, “From trees to continuous embeddings and back: Hyperbolic hierarchical clustering,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 15065–15076, 2020.
- [33] M. Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” *Advances in neural information processing systems*, vol. 26, 2013.