Probabilistic Nested Homogeneous Spaces for Dimensionality Reduction

Editors: List of editors' names

Abstract

Dimensionality reduction is a key ingredient of many machine learning algorithms and is paramount to their success. For manifold-valued data, the nonlinear equivalent of the well-known principal component analysis (PCA), called, principal geodesic analysis (PGA) is used quite often. An alternative to PGA that is more general and flexible, called "Nested Homogeneous Spaces (NHS)" for dimensionality reduction of manifold-valued data was recently introduced. In this paper, we present a novel probabilistic version of the NHS model (PNHS) for dimensionality reduction of high dimensional manifold-valued data in Riemannian homogeneous spaces. The PNHS model has several advantages over its deterministic counterpart namely, the NHS model. In particular, the ability to, quantify uncertainty in parameter estimates and tackle missing data. We demonstrate these advantages via real and synthetic data examples.

Keywords: Probabilistic model, Homogeneous space, Dimensionality reduction

1. Introduction

Dimensionality reduction is a fundamental technique in data analysis. At the core of this endeavor lies principal component analysis (PCA) (Jolliffe and Cadima, 2016), a fundamental approach in the extraction of representative features from data. Building upon this, probabilistic PCA (PPCA) (Tipping and Bishop, 1999) was developed to overcome PCA's limitation of lacking an associated probabilistic model for observed data. PPCA interprets PCA as the following latent variable model:

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{z} + \boldsymbol{\epsilon}, \tag{1}$$

on $\mathbb{R}^n$ with mean $\boldsymbol{\mu}$, coefficient matrix $\boldsymbol{W}$, latent variables $\boldsymbol{z} \in \mathbb{R}^m \sim N(0, \boldsymbol{I})(m < n)$ and noise $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I})$. This probabilistic approach not only facilitates statistical analysis but also provides practical benefits, such as handling missing data. PCA and PPCA, however, are limited to data in vector spaces and are not applicable to manifold-valued data.

Manifolds naturally arise as a relevant choice for data representation in many scenarios. For example, directional data involving unit vectors in $\mathbb{R}^n$ is best represented by the unit sphere (Mardia et al., 2000). More recently, hyperbolic space has become popular in machine learning for efficiently modeling hierarchical data (Sarkar, 2011; Nickel and Kiela, 2017). Manifolds also represent shapes in shape analysis (Kendall, 1984) and symmetric positive definite (SPD) matrices, which are useful in computer vision tasks as they correspond to covariance matrices (Tuzel et al., 2006, 2008).

In the context of dimensionality reduction, PCA is *not* applicable to data on curved manifolds, as vector space operations in PCA are undefined on manifolds. To address this, principal geodesic analysis (PGA) (Fletcher et al., 2004) was introduced as a generalization of PCA to manifolds. PGA involves projecting data onto principal geodesic submanifolds centered around an intrinsic (Frechét) mean (FM), with the objective of identifying lower-dimensional geodesic submanifolds that minimizes the geodesic distance from the original data to the projected data. Zhang and Fletcher (2013) proposed probabilistic PGA (PPGA) as a latent variable model for PGA, akin to PPCA, thereby introducing the probabilistic framework to PGA. In particular, considering an n-dimensional manifold $\mathcal{M}^n$, the latent space in PPGA is an m-dimensional linear subspace $\mathbb{R}^m$ $(m < n)$ of the tangent space at the FM $\boldsymbol{\mu}$, $T_{\boldsymbol{\mu}}\mathcal{M}^n$. The latent variable $\boldsymbol{z} \in \mathbb{R}^m \sim N(0, \boldsymbol{I})$. A linear transformation of the latent variable $\boldsymbol{z}$ forms a new tangent vector $\boldsymbol{W}\boldsymbol{z} \in T_{\boldsymbol{\mu}}\mathcal{M}^n$. Next, $\boldsymbol{W}\boldsymbol{z}$ is mapped back to $\mathcal{M}^n$ using the exponential map to generate the location parameter of a Riemannian normal distribution, from which the data $\boldsymbol{x}$ is drawn. Note that the latent space in PPGA is a Euclidean space rather than a manifold, and the tangent space approximation of data would lead to inaccuracies when the data are not clustered near the FM.

A distinctive feature of PCA is its ability to produce nested linear subspaces, where reduced-dimensional principal subspaces are hierarchically organized. Exploiting this notion, Jung et al. (2012) proposed principal nested spheres by embedding an $(n-1)$-sphere into an n-sphere. As a result, principal nested spheres surpasses PGA in flexibility by not requiring the learned submanifold to be a geodesic submanifold passing through the FM. Similar methods were applied to SPD matrices in (Harandi et al., 2018) and to hyperbolic space in (Fan et al., 2022), which developed nested hyperbolic spaces. Moreover, Yang and Vemuri (2021) undertook the task of unifying and generalizing the concept of nesting to encompass various Riemannian homogeneous manifolds — an extensive category that includes hyperspheres, hyperbolic spaces, SPD matrices, Grassmannians, Lie groups, etc. These nested constructions excel at dimensionality reduction, consistently outperforming PGA. This superiority of nested homogeneous spaces (NHS) primarily stems from the fact that the learned lower-dimensional subspace is not restricted to be a geodesic submanifold passing through the FM. In addition, Fan et al. (2022) extended the nested framework to design an encoding layer in a hyperbolic neural network for feature extraction, leveraging the nested construction to create a low-dimensional feature space while preserving hyperbolic geometry. Despite this progress, contemporary nested space techniques on Riemannian homogeneous spaces are still deterministic models and lack a probabilistic interpretation.

Our Contributions: Analogous to PPCA and PPGA, this work introduces a latent variable model for NHS, termed Probabilistic Nested Homogeneous Spaces (PNHS). The theoretical novelty lies in the development of a novel manifold-valued model and an associated Expectation Maximization (EM) algorithm tailored for parameter (manifold-valued) inference. The advantage of our PNHS model over the deterministic NHS model is further demonstrated by the inclusion of uncertainty in parameter estimates for dimensionality reduction and the ability to handle missing data. Quantifying the uncertainty in parameter estimates provides a confidence interval which is highly beneficial in capturing the reconstruction error in dimensionality reduction problems. Further this confidence interval is also highly desirable in all predictive applications, e.g., classification, recognition, etc. In contrast to the NHS, note that the PNHS is a latent space model and hence a generative model. This generative feature of our framework will be explored in future work.

2. Background

In this section, we provide a very brief overview of important concepts concerning Riemannian manifolds (Lee and Lee, 2012) and Riemannian homogeneous spaces (Helgason, 1979). We then formulate the problem of dimensionality reduction for manifold-valued data. Thereafter, we review the idea of nested homogeneous spaces as a method for achieving dimensionality reduction on the manifold.

2.1. Riemmanian Manifolds and Riemannian Homogeneous Spaces

Let (M, g) be a n-dimensional Riemannian manifold. The *tangent space* at $p \in M$ is denoted T_pM, which is a n-dimensional vector space. The Riemannian distance $d_{\mathcal{M}}(\boldsymbol{p}, \boldsymbol{q})$ between any two points $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{M}$ is defined as the length of the *geodesic*, the shortest curve connecting them, thereby generalizing the concept of a straight line in Euclidean space.

A Riemannian manifold is called *Riemannian homogeneous space* (Gallier and Quaintance, 2012, §2) if there exists a group G that acts transitively on it, with G being the *isometry* group. In this case, $\mathcal{M}$ can be written as quotient space $\mathcal{M} \cong G/H$ where H is the *isotropy* subgroup of G. Examples include but are not limited to, the n-dimensional spheres $\mathbb{S}^{n-1} \cong \mathbf{SO}(n)/\mathbf{SO}(n-1)$, the n-dimensional Lorentz model of the hyperbolic space $\mathbb{L}^n \cong \mathbf{SO}^+(1, n)/\mathbf{SO}(n)$, and others. Here $\mathbf{SO}(n)$ is the special orthogonal group and $\mathbf{SO}^+(1, n)$ is the positive special Lorentz group (Gallier and Quaintance, 2012, §2.3).

2.2. Notations and Problem Setup

Consider an n-dimensional manifold $\mathcal{M}^n$ and the target lower-dimension manifold $\mathcal{M}^m$, $m < n$. Given the observations $\{\boldsymbol{x}_i\}_{i=1}^N \in \mathcal{M}^n$, the objective of dimensionality reduction is to identify a projection function $\pi : \mathcal{M}^n \to \mathcal{M}^m$ and its corresponding embedding function, $\iota : \mathcal{M}^m \to \mathcal{M}^n$ s.t, the reconstruction error $\sum_{i=1}^N d_{\mathcal{M}}(\boldsymbol{x}_i, \iota(\pi(\boldsymbol{x}_i)))^2$ is minimized. Note that the reconstruction error can be interpreted as the unexplained variance in PCA. For a given observation $\boldsymbol{x}_i$, we denote its lower-dimensional representation on $\mathcal{M}^m$ by $\boldsymbol{z}_i = \pi(\boldsymbol{x}_i)$ and the point $\hat{\boldsymbol{x}}_i = \iota(\boldsymbol{z}_i)$ is the reconstructed point of $\boldsymbol{x}_i$ on the original manifold $\mathcal{M}^n$.

2.3. Nested Homogeneous Spaces

The steps for constructing NHS are illustrated in the commutative diagram Fig 1 (Yang and Vemuri, 2021). Let $\mathcal{M}^m \cong G_m/H_m$ be a m-dimensional Riemannian homogeneous space and its quotient space representation. The main idea is described as follows: define an embedding $\tilde{\iota}$ of the isometry group G in a suitable manner. With this embedding of the isometry group G, the embedding of the homogeneous space G/H follows naturally from the quotient structure. Specifically, since the Riemannian submersion ψ from the isometry group to the quotient space and the identification map f between the quotient space and the manifold is well defined and hence we can

$$
\begin{tikzcd}
G_m \arrow[r, "\tilde{\iota}"] \arrow[d, "\psi"] & G_{m+1} \arrow[d, "\psi"] \\
G_m/H_m \arrow[d, leftrightarrow, "f"] & G_{m+1}/H_{m+1} \arrow[d, leftrightarrow, "f"] \\
\mathcal{M}^m \arrow[r, shift left, "\iota"] & \mathcal{M}^{m+1} \arrow[l, shift left, "\pi"]
\end{tikzcd}
$$

$$
\mathcal{M}^m \underset{\pi_{m+1}}{\overset{\iota_m}{\rightleftarrows}} \mathcal{M}^{m+1} \underset{\pi_{m+2}}{\overset{\iota_{m+1}}{\rightleftarrows}} \cdots \underset{\pi_n}{\overset{\iota_{n-1}}{\rightleftarrows}} \mathcal{M}^n
$$

Figure 1: Commutative diagram of NHS.

follow the path shown in red in the commutative diagram Fig 1. Further, we can induce the embedding ι from the low dimensional manifold to the high dimensional manifold from $\tilde{\iota}$, ψ and f. Finally, the projection $\pi : \mathcal{M}^{m+1} \to \mathcal{M}^m$ can be obtained accordingly.

Let $\iota_d : \mathcal{M}^d \to \mathcal{M}^{d+1}$ and $\pi_{d+1} : \mathcal{M}^{d+1} \to \mathcal{M}^d$. The embedding ι and projection π between any dimensions $\mathcal{M}^m$ and $\mathcal{M}^n$ can be achieved through a sequence of compositions involving embeddings ι_d and projections π_d, as depicted in Fig 1 (below), i.e., $\iota \coloneqq \iota_{n-1} \circ \cdots \circ \iota_{m+1} \circ \iota_m$ and $\pi \coloneqq \pi_{m+1} \circ \pi_{m+2} \circ \cdots \circ \pi_n$. We refer readers to Appendix A for examples of nested constructions for the sphere and hyperbolic space.

3. Probabilistic Nested Homogeneous Spaces

In this section, we first introduce Riemannian normal distribution which is employed as our noise model. Then, we present the probabilistic nested homogeneous spaces (PNHS) model which allows one to interpret the NHS as a latent variable model and the parameters of this latent variable model can be solved by using an expectation maximization (EM) algorithm.

Riemannian normal. Following (Zhang and Fletcher, 2013; Pennec, 2006), we adopt the *Riemannian normal distribution*, a generalization of normal distribution for the Riemannian manifolds as our noise model. Let $\boldsymbol{x}$ be a manifold-valued random variable defined on a Riemannian manifold $\mathcal{M}$. Its probability density function (pdf) is given by:

$$p(\boldsymbol{x}|\boldsymbol{\mu}, \tau) = \frac{1}{Z(\tau)} \exp\left(-\frac{\tau}{2} d_{\mathcal{M}}(\boldsymbol{\mu}, \boldsymbol{x})^2\right), \text{ where } Z(\tau) = \int_{\mathcal{M}} \exp\left(-\frac{\tau}{2} d_{\mathcal{M}}(\boldsymbol{\mu}, \boldsymbol{x})^2\right) d\boldsymbol{x} \quad (2)$$

is the normalizing constant. Cheng and Vemuri (2013) were the first to show that the normalizing constant $Z(\tau)$ is independent of $\boldsymbol{\mu}$ in the case of Riemannian normal distribution on the manifold of SPD matrices and this property holds for all Riemannian homogeneous spaces, as shown in (Pennec et al., 2019, §2.5.1.1). We denote it as $\boldsymbol{x} \sim N_{\mathcal{M}}(\boldsymbol{\mu}, \tau^{-1})$. The parameter $\boldsymbol{\mu} \in \mathcal{M}$ is the location parameter on the manifold, while $\tau > 0 \in \mathbb{R}$ is the dispersion parameter, analogous to the precision in a Gaussian distribution. While different extensions of the normal distribution can be employed, e.g., the von Mises distribution and the wrapped normal distribution (Mardia et al., 2000, §3.5.4 & §3.5.7), the inference procedure presented in the subsequent sections can be adapted accordingly.

3.1. The Probabilistic Nested Homogeneous Spaces (PNHS) Model

We are now ready to present our proposed PNHS model, for a manifold-valued random variable $\boldsymbol{x}$ on a n-dimensional Riemannian homogeneous space $\mathcal{M}^n$, that is

$$\boldsymbol{x}|\boldsymbol{z} \sim N_{\mathcal{M}}(\iota(\boldsymbol{z}), \tau^{-1}) \quad (3)$$

where $\boldsymbol{z} \sim N_{\mathcal{M}}(\boldsymbol{\mu}_0, \tau_0^{-1})$ is latent variable in $\mathcal{M}^m$, with $\boldsymbol{\mu}_0$ and τ_0 being predetermined parameters of the latent distribution and τ a scale parameter for the noise (See Fig 2).

In this model, a latent variable $\boldsymbol{z} \in \mathcal{M}^m$, sampled from the latent distribution (which is assumed to be a Riemannian normal distribution with known parameters), is embedded into $\mathcal{M}^n$ using the embedding map ι. This process generates the location parameter $\iota(\boldsymbol{z})$ for a Riemannian normal distribution on $\mathcal{M}^n$, from which the data point $\boldsymbol{x}$ is then drawn. Our method aligns with PPCA in Euclidean space, where the Euclidean space is also a

$$\boldsymbol{z} \sim N_{\mathcal{M}}(\boldsymbol{\mu}_0, \tau_0^{-1}) \qquad \boldsymbol{x}|\boldsymbol{z} \sim N_{\mathcal{M}}(\iota(\boldsymbol{z}), \tau^{-1})$$

$$\mathcal{M}^m \underset{\pi}{\overset{\iota}{\rightleftarrows}} \mathcal{M}^n$$

Figure 2: Illustration of the probabilistic nested homogeneous space.

homogeneous space and thus an embedding $\iota : \mathbb{R}^m \to \mathbb{R}^n, \boldsymbol{z} \mapsto \boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{z}$ can be defined. We would like to emphasize that, in both PPCA and PPGA, the location parameter $\boldsymbol{\mu}$ represents the center (mean/FM) of the data and is derived by maximizing the likelihood. However, the distinctive characteristics of the NHS namely, that the learned subspace does not pass through the FM, result in the exclusion of the mean parameter in PNHS.

3.2. Inference

We now present a maximum likelihood procedure for estimating the parameters of the PNHS model described in Eq 3. These parameters, denoted as $\theta = (\theta_\iota, \tau)$, include the unknown parameter in the embedding ι, denoted as θ_ι and the noise parameter τ. Our approach involves using an expectation maximization (EM) procedure to estimate the parameters. To handle the expectation step over the latent space, we employ a Hamiltonian Monte Carlo (HMC) technique on manifolds proposed in (Brubaker et al., 2012) to draw samples $\boldsymbol{z} \in \mathcal{M}^m$ from the posterior distribution $p(\boldsymbol{z}|\boldsymbol{x}; \theta)$. Given the observations $\{\boldsymbol{x}_i\}_{i=1}^N$ on $\mathcal{M}^n$, the logarithm of the posterior distribution is given by the following expression:

$$\sum_{i=1}^N \log p(\boldsymbol{z}_i|\boldsymbol{x}_i; \theta) \propto -N \log Z(\tau) - \sum_{i=1}^N \frac{\tau}{2} d_{\mathcal{M}}(\iota(\boldsymbol{z}_i), \boldsymbol{x}_i)^2 - \sum_{i=1}^N \frac{\tau_0}{2} d_{\mathcal{M}}(\boldsymbol{\mu}_0, \boldsymbol{z}_i)^2. \tag{4}$$

By integrating this into a Monte Carlo Expectation Maximization (MCEM) framework, we effectively estimate the parameter set θ. The two primary stages, namely the E-step and the M-step, are detailed in the subsequent sections. More details about the EM algorithm (computation time, pseudo code) can be found in Appendix B.

3.2.1. E-Step

The E-step involves computing the expectation over latent space with respect to the posterior distribution. Specifically, our objective during iteration $k+1$ is to evaluate the following Q function based on current estimate of the parameters θ^k:

$$Q(\theta|\theta^k) = E_{\boldsymbol{z}_i|\boldsymbol{x}_i;\theta^k} \left[\sum_{i=1}^N \log p(\boldsymbol{z}_i|\boldsymbol{x}_i; \theta^k) \right] \tag{5}$$

Given the absence of a closed-form solution of Eq 5 in our case, we approximate the Q function using the following approach. For each $\boldsymbol{z}_i$, we employ the HMC method to draw a set of samples $\{\boldsymbol{z}_{ij}\}_{j=1}^S$, consisting of S samples, from the posterior distribution $p(\boldsymbol{z}_i|\boldsymbol{x}_i; \theta^k)$. Thus the Q function can be approximated as:

$$Q(\theta|\theta^k) \approx \frac{1}{S} \sum_{i=1}^S \sum_{i=1}^N \log p(\boldsymbol{z}_{ij}|\boldsymbol{x}_i; \theta^k). \tag{6}$$

HMC (Duane et al., 1987; Neal, 1996) is a sampling method combining Markov Chain Monte Carlo (MCMC) techniques with Hamiltonian dynamics, using gradient information from the target distribution to propose distant moves to explore complex probability distributions efficiently. In PPGA, the standard HMC method can be applied since one is interested in generating latent samples from Euclidean space, as long as the derivatives (w.r.t. the Euclidean variable) are appropriately calculated. However, in our case, we aim to sample from a posterior distribution on a Riemannian manifold, which introduces additional complexity due to the manifold's constraints. This requires careful attention when defining parameter updates. Note that while (Girolami and Calderhead, 2011) proposed the Riemann manifold Langevin and Hamiltonian Monte Carlo (RMHMC) which employs Riemannian metric on the probability space, RMHMC is not directly applicable for parameters on Riemannian manifolds – the specific focus of the present context. A non-comprehensive list of MCMC methods on manifolds includes (Brubaker et al., 2012; Byrne and Girolami, 2013; Kim et al., 2015; Lelievre et al., 2019), we refer readers to the recent review article (Liu and Zhu, 2022) for more details. For our purposes, we opt to use the constrained HMC (CHMC) presented in (Brubaker et al., 2012), which is applicable to a variety of Riemannian manifolds of interest, including the sphere, hyperbolic space, special orthogonal group and many others. Specifically, CHMC is an extension of HMC that can be applied to manifolds which can be characterized as a subset of Euclidean space defined by constraints. The simulation process involves employing the standard 'leapfrog' method (Neal et al., 2011, §2.3) within the embedded space while adhering to both the imposed constraint on the sample and the induced constraint on the *momentum*. In HMC, an auxiliary 'momentum' variable is introduced, associated with each of the variables of interest, to guide the Markov chain exploration of the target distribution.

3.2.2. M-Step

There are two kinds of parameters of interest in our model, the parameter θ_ι from the nested construction and τ, the noise parameter. Our objective is to maximize the approximated Q function shown in Eq 5. To achieve this, we obtain θ_ι by minimizing the reconstruction error and using the gradient ascent method to update the noise parameter τ.

Estimation of θ_ι. The parameter θ_ι from the embedding ι within the NHS can be obtained by minimizing the following reconstruction error,

$$L(\theta_\iota) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{S} d_{\mathcal{M}}(\iota(\boldsymbol{z}_{ij}), \boldsymbol{x}_i)^2. \tag{7}$$

This typically involves optimizing over manifolds or product spaces, as detailed in the works on nested spheres, hyperbolic spaces and Grassmanians respectively (Jung et al., 2012; Fan et al., 2022; Yang and Vemuri, 2021).

Estimation of τ. The gradient of the Q function in Eq 6 w.r.t. τ involves calculating the derivative of the normalizing constant $Z(\tau)$ in Eq 2, i.e.,

$$\nabla_\tau Q = -N\frac{Z'(\tau)}{Z(\tau)} - \frac{1}{2S}\sum_{i=1}^{N}\sum_{j=1}^{S} d_{\mathcal{M}}(\iota(\boldsymbol{z}_{ij}), \boldsymbol{x}_i)^2 \tag{8}$$

Table 1: Comparison between ground truth and estimates from NHS and PNHS on the sphere and the hyperbolic space respec. $\boldsymbol{I}_3$ is an identity matrix. For matrix parameter $\boldsymbol{R}$ (in Eq. 12 & 13 (sphere) and Eq. 14 & 15 (hyperbolic space)) – in Appendix A, the distance $d(\hat{\boldsymbol{R}}, \boldsymbol{I}_3)$ is reported.

Sphere $\mathbb{S}^2$	$d(\boldsymbol{R}, \hat{\boldsymbol{R}})$	r	τ	Hyperbolic space $\mathbb{L}^2$	$d(\boldsymbol{R}, \hat{\boldsymbol{R}})$	r	τ
Ground Truth	$\boldsymbol{I}_3$	1.00	100	Ground Truth	$\boldsymbol{I}_3$	1.00	100
NHS	0.012	1.01	N/A	NHS	0.029	0.97	N/A
PNHS	0.011	0.99	101.68	PNHS	0.027	0.97	102.53

As discussed in (Pennec et al., 2019, §2.5.1.2), applying a change-of-variables technique to the integration of the normalizing constant transforms it into the following integral.

$$Z(\tau) = A_{m-1} \int_0^R \exp\left(-\frac{\tau}{2} r^2\right) \times \prod_{k=2}^{m} |\kappa_k|^{-1/2} f_k(\sqrt{|\kappa_k|} r) dr \tag{9}$$

where A_{m-1} is the surface area of the $(m-1)$-dimensional sphere $\mathbb{S}^{m-1}$, R is the maximum distance of geodesics originating from $\boldsymbol{\mu}$ in a Riemannian normal distribution $N_{\mathcal{M}}(\boldsymbol{\mu}, \tau^{-1})$,

$$f_k(r) = \begin{cases} \frac{1}{\sqrt{\kappa_k}} \sin(\sqrt{\kappa_k} r) & \text{if } \kappa_k > 0, \\ \frac{1}{\sqrt{-\kappa_k}} \sinh(\sqrt{-\kappa_k} r) & \text{if } \kappa_k < 0, \\ r & \text{if } \kappa_k = 0, \end{cases} \tag{10}$$

here κ_k denotes the sectional curvature of the manifold. Then we can obtain $Z'(\tau)$:

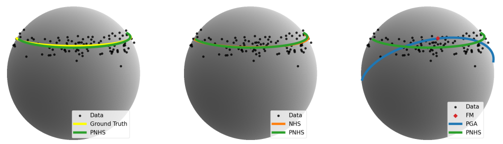
$$Z'(\tau) = A_{m-1} \int_0^R -\frac{r^2}{2} \exp\left(-\frac{\tau}{2} r^2\right) \times \prod_{k=2}^{m} |\kappa_k|^{-1/2} f_k(\sqrt{|\kappa_k|} r) dr \tag{11}$$

4. Experiments

This section evaluates the PNHS model on two common Riemannian homogeneous spaces: the sphere and hyperbolic space. We show parameter estimation results using synthetic data and highlight PNHS's effectiveness as a dimensionality reduction method on real data. We also present reconstruction errors for real data with missing values. We mainly compare our PNHS model with nested sphere (Jung et al., 2012) on the sphere, and with the NHS model (Fan et al., 2022) on the hyperbolic space, We collectively refer to these two methods as NHS (nested homogeneous spaces) throughout this section. We also include PGA (Fletcher et al., 2004) in the comparison. In all the experiments, PNHS stands out for its ability to quantify uncertainty in estimates and manage missing data. More details on the experiments (initialization, dataset,etc.) are provided in Appendix C.

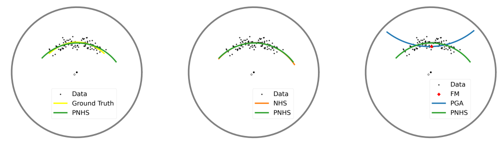
4.1. Parameter Estimation on the Sphere and the Hyperbolic Space

We present parameter estimation results for synthetic data from the sphere and hyperbolic space. Using the latent variable model of our PNHS as shown in Fig 2, we set the latent space



(a) (b) (c)

Figure 3: Ground truth, PNHS, NHS and PGA on sphere.



(a) (b) (c)

Figure 4: Ground truth, PNHS, NHS and PGA on hyperbolic space.

dimension to 1 with a latent distribution $N_{\mathcal{M}}(\boldsymbol{\mu}_0, \tau_0)$, where $\mu_0 = (1, 0)^T$ and $\tau_0 = 0$. We generate 100 random data points on both the 2-dimensional sphere $\mathbb{S}^2$ and the 2-dimensional Lorentz model of hyperbolic space $\mathbb{L}^2$. The ground truth parameters $\theta = (\theta_\iota, \tau)$, $\theta_\iota = (\boldsymbol{R}, r)$ are listed in Table 1. We use our EM algorithm to assess the model's ability to estimate these parameters accurately. Table 1 shows that both NHS and PNHS provide accurate parameter estimates, but PNHS additionally estimates the dispersion within a small error margin of about 2.5%. Fig 3 compares the ground truth, PNHS, NHS, and PGA on the sphere, while Fig 4 shows a similar comparison on hyperbolic space, using the Poincaré disk model for visualization. The close overlap between PNHS results and the ground truth indicates effective parameter recovery by our model. PNHS results are nearly indistinguishable from NHS, as both aim to minimize reconstruction errors. We also compare with PGA/PPGA, which struggles to capture the main trend of the data due to its limitation to geodesic submanifolds through the FM, leading to higher reconstruction errors, see Table C.2 in Appendix C.

4.2. Sphere: CallFish-100

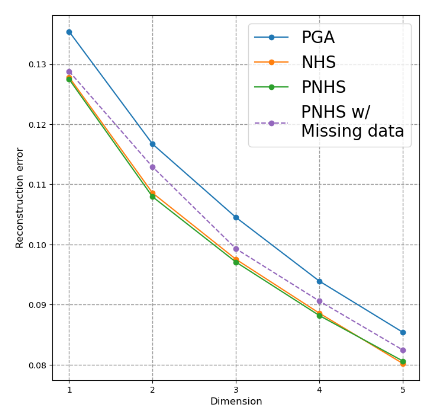


Figure 5: Reconstruction errors.

For data on the sphere, we demonstrate the effectiveness of the PNHS model on a 2D point cloud dataset from the publicly available CallFish-100 dataset (Peter and Rangarajan, 2008), which includes 100 diverse fish shapes extracted from digitized fish drawings. Sample images from this dataset are shown in Fig 6 (Appendix A). Each image is first rescaled to 40×80, then boundary points are extracted and represented using the Schrödinger Distance Transform (SDT) (Gurumoorthy and Rangarajan, 2009; Deng et al., 2014), which maps these point sets onto a hypersphere ($\mathbb{S}^{3199}$). The SDT representation, normalized to have unit $L2$ norm, allows these point sets to be represented as probability densities using square-root density parameterization, resulting in points on a high-dimensional unit sphere in the discrete case, specifically $\mathbb{S}^{3199}$, totaling 100 samples in the dataset. We apply several dimensionality reduction methods to reduce dimensions to a range of 1 to 5. The resulting reconstruction errors for PGA, NHS, and PNHS are compared in Fig 5 . The estimated dispersion $\hat{\tau}$ in PHNS for dim. 1 to 5 are 9.55, 11.03, 12.38, 13.58 and 14.99, respectively. As expected, NHS and PNHS produce nearly identical results due to their shared objective functions in learning the low-dimensional representation space, similar to how the

Table 2: Reconstruction errors (mean and std. dev.) in dim. reduction from $\mathbb{L}^{10}$ to $\mathbb{L}^2$. *Etimated dispersion param. $\hat{\tau}$ in PNHS is included.*

Datasets	balancedtree	phylo tree	diseasome	ca-CSphd
PGA	5.75	121.19	21.53	71.67
HoroPCA	7.80 ± 0.06	108.62 ± 9.20	26.94 ± 0.99	87.99 ± 4.69
NHS	3.35 ± 0.05	24.11 ± 0.68	9.18 ± 0.10	22.68 ± 0.40
PNHS	3.34 ± 0.06	24.14 ± 0.78	9.17 ± 0.09	22.59 ± 0.47
$\hat{\tau}$(**dispersion**)	0.456	0.055	0.173	0.058

principal components in PPCA match those in PCA. Additionally, NHS/PNHS significantly outperform PGA in terms of reconstruction error, which aligns with our expectations since the nested construction offers more flexibility than PGA discussed earlier.

Missing Data Experiment: Our EM algorithm for estimating PNHS model parameters can handle missing data, similar to the approach in PPCA This involves starting with an initial guess for the missing values, which are then iteratively updated in each E-step to minimize reconstruction error. We simulated missing data by randomly dropping 10% of values from the CallFish-100 dataset. We then applied our PNHS model to reduce dimensionality to between 1 and 5. The reconstruction errors, shown in Fig 5, reveal that performance decreases with missing data but still exceeds that of PGA. Notably, neither PGA nor NHS can handle missing data. These results highlight PNHS's effectiveness in managing missing data while maintaining performance comparable to complete datasets.

4.3. Hyperbolic Space: Embeddings of Trees

In hyperbolic space, we focus on reducing the dimensionality of trees embedded in this space. We validate our method on four datasets from (Chami et al., 2021). We use (Gu et al., 2018) to embed these tree datasets into a 10-dimensional Poincaré ball, then apply various dimensionality reduction methods to reduce the dimension to 2. In addition to PGA and NHS, we compare with HoroPCA (Chami et al., 2021), which uses horospherical projection to map points onto a geodesic submanifold. Reconstruction errors for each method across datasets are reported in Table 2. PNHS and NHS perform similarly and better than PGA and HoroPCA. This highlights the effectiveness of PNHS in hyperbolic space. Additionally, we report the estimated dispersion parameter $\hat{\tau}$ for PNHS, which quantifies variance/uncertainty and is not available in the NHS model.

5. Conclusion

In this paper, we presented PNHS, a probabilistic nested homogeneous spaces model for dimensionality reduction. NHS, a SOTA dimensionality reduction technique for high-dimensional Riemannian homogeneous spaces, is adapted into PNHS by treating it as a latent variable model. We estimate the model parameters with a novel adaptation of the EM algorithm to manifold-valued data. Empirical tests show PNHS's effectiveness using sphere and hyperbolic space data, highlighting its advantage over other models and its ability to provide uncertainty estimates, unlike deterministic models. We also address missing data with a slight adjustment to the EM algorithm. Future work will explore PNHS's use as a neural network decoding layer and its extension to multi-class generative modeling.

References

Marcus Brubaker, Mathieu Salzmann, and Raquel Urtasun. A family of mcmc methods on implicitly defined manifolds. In *Artificial intelligence and statistics*, pages 161–172. PMLR, 2012.

Simon Byrne and Mark Girolami. Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, 2013.

James W Cannon, William J Floyd, Richard Kenyon, and Walter R Parry. Hyperbolic geometry. *Flavors of geometry*, 31:59–115, 1997.

Ines Chami, Albert Gu, Dat P Nguyen, and Christopher Ré. Horopca: Hyperbolic dimensionality reduction via horospherical projections. In *International Conference on Machine Learning*, pages 1419–1429. PMLR, 2021.

Guang Cheng and Baba C Vemuri. A novel dynamic system in the space of spd matrices with applications to appearance tracking. *SIAM journal on imaging sciences*, 6(1):592–615, 2013.

Yan Deng, Anand Rangarajan, Stephan Eisenschenk, and Baba C Vemuri. A riemannian framework for matching point clouds represented by the schrodinger distance transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3756–3761, 2014.

Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

Xiran Fan, Chun-Hao Yang, and Baba C Vemuri. Nested hyperbolic spaces for dimensionality reduction and hyperbolic nn design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 356–365, 2022.

P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

Jean Gallier and Jocelyn Quaintance. Notes on differential geometry and lie groups. *University of Pennsylvannia*, 4:3–1, 2012.

Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(2):123–214, 2011.

Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2018.

Karthik S Gurumoorthy and Anand Rangarajan. A schrödinger equation for the fast computation of approximate euclidean distance functions. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 100–111. Springer, 2009.

Mehrtash Harandi, Mathieu Salzmann, and Richard Hartley. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):48–62, 2018.

Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces.* Academic Press, 1979.

Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

Sungkyu Jung, Ian L Dryden, and James Stephen Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, 2012.

David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.

Hyunwoo Kim, Jia Xu, Baba Vemuri, and Vikas Singh. Manifold-valued dirichlet processes. In *International Conference on Machine Learning*, pages 1199–1208. PMLR, 2015.

Niklas Koep and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17: 1–5, 2016.

John M Lee and John M Lee. *Smooth manifolds.* Springer, 2012.

Tony Lelievre, Mathias Rousset, and Gabriel Stoltz. Hybrid monte carlo methods for sampling probability measures on submanifolds. *Numerische Mathematik*, 143:379–421, 2019.

Chang Liu and Jun Zhu. Geometry in sampling methods: A review on manifold MCMC and particle-based variational inference methods. In Arni S.R. Srinivasa Rao, G. Alastair Young, and C.R. Rao, editors, *Advancements in Bayesian Methods and Implementation*, volume 47 of *Handbook of Statistics*, pages 239–293. Elsevier, 2022. doi: https://doi.org/10.1016/bs.host.2022.07.004. URL https://www.sciencedirect.com/science/article/pii/S0169716122000426.

Kanti V Mardia, Peter E Jupp, and KV Mardia. *Directional statistics*, volume 2. Wiley Online Library, 2000.

RM Neal. Bayesian learning for neural networks. *Lecture Notes in Statistics*, 1996.

RM Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30:6338–6347, 2017.

Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127–154, 2006.

Xavier Pennec, Stefan Sommer, and Tom Fletcher. *Riemannian geometric statistics in medical image analysis*. Academic Press, 2019.

Adrian M Peter and Anand Rangarajan. Maximum likelihood wavelet density estimation with applications to image and shape matching. *IEEE Transactions on Image Processing*, 17(4):458–468, 2008.

Rik Sarkar. Low distortion delaunay embedding of trees in hyperbolic plane. In *International Symposium on Graph Drawing*, pages 355–366. Springer, 2011.

Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622, 1999.

Oncel Tuzel, Fatih Porikli, and Peter Meer. Region covariance: A fast descriptor for detection and classification. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9*, pages 589–600. Springer, 2006.

Oncel Tuzel, Fatih Porikli, and Peter Meer. Pedestrian detection via classification on riemannian manifolds. *IEEE transactions on pattern analysis and machine intelligence*, 30 (10):1713–1727, 2008.

Chun-Hao Yang and Baba C Vemuri. Nested grassmanns for dimensionality reduction with applications to shape analysis. In *International Conference on Information Processing in Medical Imaging*, pages 136–149. Springer, 2021.

Miaomiao Zhang and Tom Fletcher. Probabilistic principal geodesic analysis. *Advances in neural information processing systems*, 26, 2013.

Appendix A. Examples of Nested Homogeneous Spaces

We illustrate the explicit nested structures of two specific Riemannian homogeneous spaces: sphere $\mathbb{S}$ and the Lorentz model of the hyperbolic space $\mathbb{L}$. This is accomplished by detailing the embedding and projection employed within these spaces.

Nested spheres (Jung et al., 2012; Yang and Vemuri, 2021). An m-dimensional sphere $\mathbb{S}^m$ encompasses all points in $\mathbb{R}^{m+1}$ with unit length, i.e., $\mathbb{S}^m = \{\boldsymbol{x} = (x_0, x_1, \dots, x_m)^T \in \mathbb{R}^{m+1} : \sum_{i=0}^{m} x_i^2 = 1\}$. Sphere can be written as quotient space $\mathbb{S}^m \cong \mathbf{SO}(m+1)/\mathbf{SO}(m)$, where $\mathbf{SO}(m)$ is is the group of $m \times m$ orthogonal matrices with determinant 1. The induced embedding $\iota_m : \mathbb{S}^m \to \mathbb{S}^{m+1}$ is

$$\iota_m(\boldsymbol{z}) = \boldsymbol{R} \begin{bmatrix} \sin(r)\boldsymbol{z} \\ \cos(r) \end{bmatrix} = \sin(r)\tilde{\boldsymbol{R}}\boldsymbol{z} + \cos(r)\boldsymbol{v} \tag{12}$$

where $\boldsymbol{z} \in \mathbb{S}^m$, $r \in \mathbb{R}$, $\boldsymbol{R} = [\tilde{\boldsymbol{R}} \quad \boldsymbol{v}] \in \mathbf{SO}(m+2)$, $\tilde{\boldsymbol{R}}$ is an $(m+2, m+1)$ matrix consisting of the first $m+1$ columns of $\boldsymbol{R}$, $\boldsymbol{v}$ is the last column of $\boldsymbol{R}$. Corresponding to ι_m, we have projection $\pi_{m+1} : \mathbb{S}^{m+1} \to \mathbb{S}^m$:

$$\pi_{m+1}(\boldsymbol{x}) = \frac{1}{\sin(r)}\tilde{\boldsymbol{R}}^T\boldsymbol{x} = \frac{\tilde{\boldsymbol{R}}^T\boldsymbol{x}}{\|\tilde{\boldsymbol{R}}^T\boldsymbol{x}\|} \quad \boldsymbol{x} \in \mathbb{S}^{m+1} \tag{13}$$

The geodesic distance between $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^m$ is $d_{\mathbb{S}}(\boldsymbol{x}, \boldsymbol{y}) = \cos^{-1}(\boldsymbol{x}^T\boldsymbol{y})$.

Nested hyperbolic spaces (Fan et al., 2022). The Lorentz (hyperboloid) model is one of the five isometric models (Cannon et al., 1997) of hyperbolic space which can be regarded as a homogeneous space. An m-dimensional Lorentz model of hyperbolic space $\mathbb{L}^m$ is defined as $\mathbb{L}^m = \{\boldsymbol{x} = (x_0, x_1, \dots, x_m)^T \in \mathbb{R}^{m+1} : -x_0^2 + \sum_{i=1}^{m} x_i^2 = -1, x_0 > 0\}$ which can be written as quotient space $\mathbb{L}^m \cong \mathbf{SO}^+(1,m)/\mathbf{SO}(m)$, where $\mathbf{SO}^+(1,m)$ is the positive specical Lorentz group (see (Gallier and Quaintance, 2012; Fan et al., 2022) for more details). The induced embedding $\iota_m : \mathbb{L}^m \to \mathbb{L}^{m+1}$ is

$$\iota_m(\boldsymbol{z}) = \boldsymbol{\Lambda} \begin{bmatrix} \cosh(r)\boldsymbol{z} \\ \sinh(r) \end{bmatrix} = \cosh(r)\tilde{\boldsymbol{\Lambda}}\boldsymbol{z} + \sinh(r)\boldsymbol{v} \tag{14}$$

where $\boldsymbol{z} \in \mathbb{L}^m$, $r \in \mathbb{R}$, $\boldsymbol{\Lambda} = [\tilde{\boldsymbol{\Lambda}} \quad \boldsymbol{v}] \in \mathbf{SO}^+(1, m+1)$, $\tilde{\boldsymbol{\Lambda}}$ is an $(m+2, m+1)$ matrix consisting of the first $m+1$ columns of $\boldsymbol{\Lambda}$, $\boldsymbol{v}$ is the last column of $\boldsymbol{\Lambda}$. Corresponding to ι_m, we have projection $\pi_{m+1} : \mathbb{L}^{m+1} \to \mathbb{L}^m$:

$$\pi_{m+1}(\boldsymbol{x}) = \frac{1}{\cosh(r)} J_m\tilde{\boldsymbol{\Lambda}}^T J_{m+1}\boldsymbol{x} = \frac{J_m\tilde{\boldsymbol{\Lambda}}^T J_{m+1}\boldsymbol{x}}{\|J_m\tilde{\boldsymbol{\Lambda}}^T J_{m+1}\boldsymbol{x}\|_L} \quad \boldsymbol{x} \in \mathbb{L}^{m+1} \tag{15}$$

The geodesic distance between $\boldsymbol{x} = (x_0, x_1, \dots, x_m)^T, \boldsymbol{y} = (y_0, y_1, \dots, y_m)^T \in \mathbb{L}^m$ is $d_{\mathbb{L}}(\boldsymbol{x}, \boldsymbol{y}) = \cosh^{-1}(-(-x_0y_0 + \sum_{i=1}^{m} x_iy_i))$.

Dimensionality Reduction. With the outlined embedding ι and projection π, as well as the geodesic distance measure, the task of dimensionality reduction on the manifold is carried out by determining the unknown parameters in ι and π. This involves minimizing the reconstruction error for the given observations. The detailed optimization procedure is presented in detail in the original contributions (Jung et al., 2012; Fan et al., 2022). In the case of nested spheres, for a set of samples $\{\boldsymbol{x}_i\}_{i=1}^N \in \mathbb{S}^n$, the unknown parameters are identified as $(\boldsymbol{R}, r) \in \mathbf{SO}(n+1) \times \mathbb{R}$, as per Eq. 12 and 13. The loss function is then formulated as $L(\boldsymbol{R}, r) = \frac{1}{N}\sum_{i=1}^N d_{\mathbb{S}}(\hat{\boldsymbol{x}}_i, \boldsymbol{x}_i)^2$, where $\hat{\boldsymbol{x}}_i = \iota(\pi(\boldsymbol{x}_i))$ is the reconstructed point of $\boldsymbol{x}_i$ in $\mathbb{S}^n$. The minimization of this loss function is an optimization process over the product space $(\boldsymbol{R}, r) \in \mathbf{SO}(n+1) \times \mathbb{R}$. This can be solved using a Riemannian gradient-based optimization technique, e.g., Pymanopt (Koep and Weichwald, 2016).

Appendix B. Supplementary Details about EM Algorithm

B.1. Computation Time

The computation time of the iterative process within the EM algorithm is dependent upon the distinct characteristics of both the E-step and the M-step, which vary across different manifolds. The E-step involves a sampling procedure that can be expedited through parallel computation of multiple Markov chains. Meanwhile, the compute time for the M-step mainly depends on the updation of θ_ι. We refer the readers to (Jung et al., 2012; Yang and Vemuri, 2021; Fan et al., 2022) for details of the optimization methods which are also applicable in our nested homogeneous spaces model.

B.2. Pseudo Code for PNHS

Algorithm 1: EM Algorithm for PNHS

Input: Data $\boldsymbol{x}$ on manifold $\mathcal{M}^n$, reduced dimension m
Output: Optimized parameters $\theta = (\theta_\iota, \tau)$
Initialize $\theta = (\theta_\iota, \tau)$;
repeat
  E-step: Sample according to Eq. 6 in Sec 3.2.1 ;
  M-step: Update parameters $\theta = (\theta_\iota, \tau)$ according to Eq. 7 and Eq. 8 in Sec 3.2.2;
until *convergence*;

Appendix C. Supplementary Details about Experiments

C.1. Initialization of Parameters

For each experiment, we first initialize the parameters, specifically, $\theta = (\theta_\iota, \tau)$, using the same approach as in solving PPCA with the EM algorithm. In our PNHS, the initialization of θ_ι, denoted as $\theta_\iota^{(0)}$, is obtained by applying the corresponding NHS model, for instance, as Nested Spheres. The initialization of τ is derived from its estimates $\hat{\tau}^{(0)} = \frac{1}{2}\left(\sum_{i=1}^N \frac{1}{N} d_{\mathcal{M}}(\iota_{\theta_\iota^{(0)}}(\boldsymbol{z}_i), \boldsymbol{x}_i)^2\right)^{-1}$.

C.2. Comparison with PPGA on Synthtic Data

In the context of comparing state-of-the-art (SOTA) methods for dimensionality reduction on manifold-valued data, the PPGA is the primary method relevant to our study. A direct

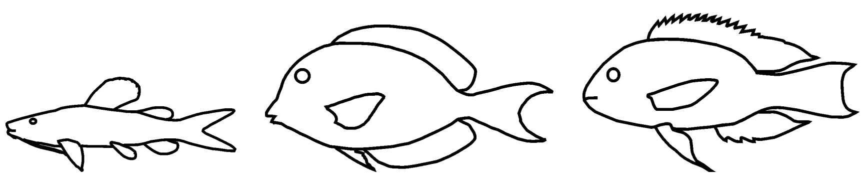
Table 3: Comparison of dispersion estimates and reconstruction errors between NHS/PNHS and PGA/PPGA on synthetic data.

Sphere	$\hat{\tau}$	Reconstruction Error	**Hyperbolic Space**	$\hat{\tau}$	Reconstruction Error
ground truth	100		ground truth	100	
NHS	N/A	0.016	NHS	N/A	0.014
PNHS	101.68	0.016	PNHS	102.53	0.014
PGA	N/A	0.038	PGA	N/A	0.036
PPGA	42.63	0.038	PPGA	39.86	0.036

comparison between PPGA and our proposed PNHS method is challenging due to the differing latent space models, which prevent straightforward parameter comparisons. However, a practical approach to evaluating the methods is through an assessment of dispersion in reconstruction accuracy.

To this end, we present a comparison of the dispersion estimates for PPGA and PNHS using synthetic data on both spherical and hyperbolic manifolds in Sec 4.1. This evaluation is restricted to the synthetic data setting, as it requires the availability of ground truth to accurately measure reconstruction error. The results shown in Table C.2 demonstrate that PNHS provides more accurate estimates of dispersion compared to PPGA, aligning more closely with the ground truth values.

C.3. Samples from CallFish-100



(a) (b) (c)

Figure 6: Digitized fish shapes from CallFish-100.