

# Spherelets

Stat 185 Term Paper

*Caleb Ren*

*December 14, 2019*

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Method</b>	<b>3</b>
2.1	Theory . . . . .	3
2.1.1	Covering Number . . . . .	3
2.1.2	Main Theorem . . . . .	3
2.2	Spherical PCA . . . . .	5
2.2.1	Notation . . . . .	5
2.2.2	Assumptions . . . . .	6
2.2.3	Estimation . . . . .	6
2.3	Algorithm . . . . .	7
<b>3</b>	<b>Discussion</b>	<b>8</b>
3.1	Strengths . . . . .	8
3.2	Weaknesses . . . . .	8
<b>4</b>	<b>Examples</b>	<b>9</b>
4.1	Euler Spiral . . . . .	9
4.2	Helix . . . . .	10
4.3	Cylinder . . . . .	10
<b>5</b>	<b>References</b>	<b>12</b>

# 1 Introduction

Data that exists in a high-dimensional ambient space may instead be considered to lie near a lower dimensional manifold (e.g. a circle in  $\mathbb{R}^2$  that lies ambiently in  $\mathbb{R}^3$  space). Many techniques are focused on reducing the ambient space to one closer to the intrinsic space such as clustering (Duda, Stork, and Hart 2000), data compression (Hurtado 2012), and predictive modelling (J. A. Lee and Verleysen 2008). Many of these techniques that attempt to approximate manifolds embedded in a lower dimensional space are locally linear, use multiscale dictionaries, and as such struggle with areas with high Gaussian curvature. Additionally, many techniques that approximate local manifolds do not perform well with out-of-sample error as they do not reveal much information on the traits of the lower dimension manifold.

A simple alternative that is able to handle curvature as well as out-of-sample error is to use pieces of (hyper)spheres to locally approximate the unknown subspace (Didong Li and Dunson 2019). This method was first proposed by Li and Dunson, who termed the technique spherical PCA (SPCA) (Li and Dunson 2017), or *spherelets* for short. Whereas principal component analysis (PCA) is an eigenvalue/eigenvector problem from an inherently *linear* dimension reduction problem, spherelets rely on projection mappings to the surfaces of a hypersphere as an underlying mechanism. As such, in simple cases of SPCA, closed form analytic solutions exist and can be seen as a generalization of PCA that is able to incorporate degrees of curvature. Spherical PCA in general does not have a closed-form solution due to the myriad ways in which a particular dataset can be partitioned and fit with sub-manifolds. Various algorithms exist to best partition the original dataset such as cover trees (Beygelzimer, Kakade, and Langford 2006), METIS (Karypis and Kumar 1998), and iterated PCA (Szlam 2009). The type of partitioning done is context-dependent and multiple types of subsetting can be done to cross-validated to reduce test MSE. As examples, we will examine two types of partitioning: naively—by splitting the dataset into equal-sized clusters—and via partitioning by iterated PCA (Didong Li and Dunson 2019).

Data that is represented in higher dimension  $\mathbb{R}^D$  can be projected down to a manifold in  $\mathbb{R}^d$  consisting of a set of spherelets in  $\mathbb{R}^d$ . Ultimately, the algorithm partitions the dataset into  $k$  subsets, each of which is fit with a submanifold  $M_k \in \mathbb{R}^d$  with a corresponding projection map  $\Psi_k : \mathbb{R}^D \rightarrow \mathbb{R}^d$ . Given a dataset  $\mathbf{X} \in \mathbb{R}^D$ , the spherical PCA algorithm returns manifold  $M$  and the projection map  $\Psi$ , which are the manifold and mapping from the data to the approximate manifold in  $\mathbb{R}^d$ , respectively.

Manifold approximation for model-building is one useful application of spherelets but the technique can be used for other purposes. Spherical PCA can also be used to denoise manifolds and generate visualizations of data in higher dimensions. Spherelets have been shown to outperform competing denoising algorithms like Gaussian Blurring Mean Shift (GBMS), Manifold Blurring Mean Shift (MBMS), and Local Tangent Projection (LTP). Visualization techniques that preserve local geometry like Isomap, Locally Linear Embedding (LLE), and t-distributed Stochastic Neighborhood Embedding (t-SNE) can also be modified to incorporate spherical geometry using the same projection to spherical affine subspace, as in the original spherelets paper (Didong Li and Dunson 2019).

In spherical PCA, a point  $\vec{y}_i$  is projected down to a sphere  $S_V(c, r)$ , where  $c$  is the center of the sphere,  $r$  is the radius, and  $V$  specifies the subspace on which the sphere lies. The optimal estimates for the sphere  $S_V(\hat{c}, \hat{r})$  are given by:

$$\begin{aligned}\hat{V} &= (\vec{v}_1, \dots, \vec{v}_{d+1}) \\ \hat{c} &= -\frac{\vec{\eta}^*}{2} \\ \hat{r} &= \frac{1}{N} \sum_{k=1}^N \|\vec{z}_i - \hat{c}\|\end{aligned}$$

Where  $\vec{v}_i$  is the  $i$ th eigenvector of the covariance matrix times  $\Sigma$ ,  $\vec{z}_i$  is the the projected data down to the subspace and  $\vec{\eta}^*$  is a vector based on the entire projected dataset  $\mathbf{Z}$ . After an introduction into notation of spherical PCA, I will introduce relevant theory, including the Main Theorem that undergirds spherical PCA in the following section, as well as assumptions made. I will discuss strengths and weaknesses of the algorithm in Section 3, and then provide numerical examples demonstrating areas in which spherelets perform well and where the algorithm struggles in Section 4.

## 2 Method

### 2.1 Theory

#### 2.1.1 Covering Number

The Main Theorem underlying spherical PCA asserts that the covering number of spherelets is lower than that of an analogous PCA method using hyperplanes. Given a dictionary of basis functions  $\mathcal{B}$ , the Main Theorem shows a that the number of individual “pieces” of spheres to approximate a manifold within error threshold  $\epsilon > 0$  is lower than that of a purely hyperplane method (Didong Li and Dunson 2019). If  $M$  is a  $d$ -dimensional, compact,  $C^3$ -smooth Riemannian manifold, the covering number  $N_{\mathcal{B}}(\epsilon, M)$  is defined as:

$$N_{\mathcal{B}}(\epsilon, M) := \inf_N \{N : \exists \{C_k, \Psi_k, B_k\}_{k=1}^K \text{ s.t. } \|x - \Psi(x)\| \leq \epsilon, \forall x \in M\}$$

Here,  $C_k$  defines the  $k$ th partition of ambient space  $\mathbb{R}^D$ ,  $B_k$  is a local basis within the dictionary of basis functions  $\mathcal{B}$ . We have previously defined  $\Psi$  to be the global projection, so  $\Psi_k$  is the local projection within the  $k$ th partition and  $\Psi = \sum_{k=1}^K 1_{x \in C_k} \Psi_k$  defines the global projection as a combination of local projections.

This definition of covering number holds that there exists some number  $N$  such that we can fit a sphere with error lower than threshold  $\epsilon$ . In the context of spherical PCA, we have two general choices of  $\mathcal{B}$ :  $\mathcal{S}$ —the dictionary of basis functions that use a spherical affine space—and  $\mathcal{H}$ —the dictionary of basis functions that use hyperplanes. Note that if we allow the construction of spheres with infinite radius, we obtain manifolds with  $\kappa = 0$ , which are hyperplanes! Thus,  $\mathcal{H} \subset \mathcal{S}$ . As such, we arrive at the inequality:

$$N_{\mathcal{S}}(\epsilon, M) \leq N_{\mathcal{H}}(\epsilon, M)$$

Spherical PCA asserts that in the worst case with totally linear manifolds, the algorithm will perform as poorly as local PCA that uses hyperplanes instead of sections of spheres as its dictionary of basis functions.

#### 2.1.2 Main Theorem

We can also achieve a tighter upper bound than the above inequality. Given a manifold  $M$  that satisfies the conditions listed above ( $d$ -dimensional with  $d < D$ ,  $C^3$ -smooth, compact, Riemannian), we define additional notation:

- $K(p, \vec{v})$  is a function that assigns a curvature of the geodesic on  $M$  starting at arbitrary point  $p$  with direction  $\vec{v}$ ,

- $K^*$  is the maximum curvature defined above with  $K^* := \sup_{(p, \vec{v}) \in UTM} K(p, \vec{v}) < \infty$ ,
- Similarly, if  $T(p, \vec{v})$  defines the rate of change of curvature, then  $T^* := \sup_{(p, \vec{v}) \in UTM} T(p, \vec{v})$  defines the maximum rate of change in curvature,
- $UTM$  as encountered above defines the unit sphere bundle over  $M$  and  $UT_p M$  defines the unit sphere bundle centered at point  $p \in M$ ,
- $F_\epsilon$  is the set of  $\epsilon$ -spherical points on  $M$  with difference in maximum and minimum curvature under a threshold level parametrized by  $\epsilon$ :

$$F_\epsilon := \left\{ p \in M : \sup_{\vec{v} \in UT_p M} K(p, \vec{v}) - \inf_{\vec{v} \in UT_p M} K(p, \vec{v}) \leq \left( \frac{2\epsilon}{K^*} \right)^{1/2} \right\}$$

- $B(p, \epsilon)$  is a ball with radius  $\epsilon$  and center  $p$ .

That is to say, the spherical submanifold  $M_\epsilon$  of  $M$  for a given error  $\epsilon$  is the union of all the balls centered at points in  $F_\epsilon$ , or

$$M_\epsilon := \bigcup_{p \in F_\epsilon} B \left( p, \left( \frac{6\epsilon}{3 + T^*} \right)^{1/3} \right)$$

We have therefore defined a way to define a spherical submanifold  $M_\epsilon$  of base manifold  $M$  given an error level  $\epsilon$ . We define  $V_\epsilon$  as the volume contained within this spherical submanifold  $M_\epsilon$ .

We have now defined the notation necessary for the main theorem, which contrasts covering numbers for hyperplanes and sections of spheres.

**Main Theorem.** (Didong Li and Dunson 2019) For any  $\epsilon > 0$  and compact  $C^3$   $d$ -dimensional Riemannian manifold  $M$ ,

$$N_{\mathcal{H}} \lesssim V \left( \frac{2\epsilon}{K^*} \right)^{-d/2}$$

$$N_{\mathcal{S}} \lesssim V_\epsilon \left( \frac{6\epsilon}{3 + T^*} \right)^{-d/3} + (V - V_\epsilon) \left( \frac{2\epsilon}{K^*} \right)^{-d/2}$$

A key result from the Main Theorem is that the covering number of spherelets has a lower upper bound when the difference between  $V$  and  $V_\epsilon$  is small. In the case when  $V - V_\epsilon \rightarrow 0$ , the second term in the bound for  $N_{\mathcal{S}}$  shrinks to 0 and the first term dominates, which grows at  $O(c^{-\frac{d}{3}})$ , which outperforms the  $O(c^{\frac{d}{2}})$  that  $N_{\mathcal{H}}$  grows at.

In the case when  $V_\epsilon \rightarrow 0$  (when the degree of curvature is low so the spherical submanifold is unable to cover much of the underlying manifold), spherical PCA converges to hyperplane PCA.

According to Li and Dunson, the bounds to the Main Theorem are tight, implying that spherelets often require many fewer pieces than locally linear dictionaries and approximate  $M$  for a given error level  $\epsilon$ . Spherelets provide an exceptionally good approximation of  $M$  when there the curvature across the manifold stays in a relatively narrow range, as these would require many more parameters in locally linear PCA to approximate. As such, spherelets provide a more robust way of fitting datasets that have intrinsic curvature in a lower  $d$ -dimensional manifold compared to hyperplane PCA.

## 2.2 Spherical PCA

### 2.2.1 Notation

Assume we have the  $N \times D$  data matrix  $\mathbf{X}$ , with  $N$  observations and  $D$  variable where  $x_{i,j}$  represents the  $i$ th observation of the  $j$ th variable:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \dots & x_{1,D} \\ \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,D} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,D} \end{bmatrix}$$

As in linear PCA, a more succinct way to represent this matrix is to write:

$$\mathbf{X} = \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_i^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix}$$

We will once again treat  $\vec{x}_1, \dots, \vec{x}_N$  as i.i.d. samples of a random vector in  $\mathbb{R}^D$  from the same underlying distribution  $F$  for which  $E_{x \sim F} \|x\|^2 < \infty$  to assure that the mean and covariance of the data matrix are well-defined. We use:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N \vec{x}_i$$

to denote the sample mean of  $\vec{x}$ . To denote the covariance operator, we use:

$$\Sigma_x = \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \bar{x})(\vec{x}_i - \bar{x})^T = \frac{1}{N} \left( \sum_{i=1}^N \vec{x}_i \vec{x}_i^T \right) - \bar{x} \bar{x}^T = \frac{1}{N} \mathbf{X}^T \mathbf{X} - \bar{x} \bar{x}^T$$

This covariance operator will come into use later on. Additional notation includes:

- $1_N$  is the column vector of all ones with length  $N$  (i.e.  $1_N \in \mathbb{R}^N$ )
- $\|\vec{z}\|$  denotes the Euclidean norm (i.e. for  $\vec{z} \in \mathbb{R}^d$ ,  $\|\vec{z}\| = \left( \sum_{i=1}^d z_i^2 \right)^{1/2} = \sqrt{\vec{z}^T \vec{z}}$ )
- $\Psi : \mathbb{R}^D \rightarrow \mathbb{R}^{d+1}$  is a projection map from space  $\mathbb{R}^D$  to  $\mathbb{R}^{d+1}$ . Note:  $\Psi$  maps from  $\mathbb{R}^D$  to  $\mathbb{R}^{d+1}$  rather than  $\mathbb{R}^d$  because the projected points lie on an affine subspace with one fewer degree of freedom than the manifold. One example is fitting a curve in  $\mathbf{R}^2$  with sections of 2-dimensional circles; while circles are 2D, the projected points only live along the edge of the circle. As such, we include an extra dimension in our projection map to account for the extra degree of freedom.
- $d^2(\cdot, \cdot)$  is the distance operator between two points. Note that  $d^2(x, y)$  is equivalent to  $\sqrt{\|x - y\|}$ .

### 2.2.2 Assumptions

There are three main assumptions made by Li and Dunson:

- **Distributional Assumption.** The data matrix  $\mathbf{X} \in \mathbb{R}^{N \times D}$  consists of a transformation  $V^* \Lambda^{*\frac{1}{2}} \mathbf{Z}$  where  $\mathbf{Z} \in \mathbb{R}^{N \times D}$  is a data matrix consisting of i.i.d  $z_i$  non-degenerate random variables.
- **Moment Assumption.** For the underlying random variables  $z_i$ ,  $E(z_i) = 0$ ,  $E(z_i^2) = 1$ , and  $E(z_i^6) < \infty$ . We can choose an affine transformation  $V^*$  such that the underlying random variables satisfy the first two moment assumptions. A similar assumption is considered in Lee et al. that uses bounded fourth moments instead to prove convergence in high dimensional PCA (S. Lee, Zou, and Wright 2010).
- **Spike Population Model Assumption.** If  $\lambda_1, \lambda_2, \dots, \lambda_D$  are the ordered eigenvalues of  $\Lambda^*$  as similar to a PCA setting, then there exists an integer  $m > d$  such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \lambda_{m+1} = \dots = \lambda_D = 1$ .

The consequence of these assumptions is that we have a theoretical guarantee that the estimate projection  $\hat{\Psi}$  converges in probability to  $\Psi^*$ . In other words:

$$\lim_{n \rightarrow \infty} P(\sup_{x \in M} \|\hat{\Psi}(x) - \Psi^*(x)\| > \epsilon) = 0, \forall \epsilon < 0$$

### 2.2.3 Estimation

Given a set of data  $\vec{x}_1, \dots, \vec{x}_N \in \mathbb{R}^D$  organized into data matrix  $\mathbf{X}$ , spherical PCA estimates the best approximating sphere  $S_V(c, r)$ , where  $c$  is the center,  $r$  is the radius, and  $V \in \mathbb{R}^{(d+1) \times (d+1)}$  is the  $(d+1)$ th dimensional affine subspace the sphere lives on.

For observation  $\vec{x}_i$ , the closest point  $\vec{y}_i$  lying on the sphere  $S_V(c, r)$  is the point that minimizes Euclidean distance  $d^2(\vec{x}_i, \vec{y}_i)$  between  $\vec{x}$  and  $\vec{y}$ .

Continuing the convention that  $\Psi$  is an orthogonal projection from  $\mathbb{R}^D$  to the affine subspace given by  $c + V$ , then  $\Psi$  is given by:

$$\hat{\Psi}(\vec{x}_i) = \hat{c} + \frac{\hat{r}}{\|\hat{V}\hat{V}^T(\vec{x}_i - \hat{c})\|} \hat{V}\hat{V}^T(\vec{x}_i - \hat{c})$$

Note that  $\vec{x} - \Psi_{V,c}(\vec{x}) \perp \Psi_{V,c}(\vec{x}) - \vec{y}$  since  $\Psi$  is an orthogonal projection.

The optimal subspace  $V$  is given by  $\hat{V} = (\vec{v}_1, \dots, \vec{v}_{d+1})$ , where  $\vec{v}_i, i \in \{1, \dots, d+1\}$  is the  $i$ th eigenvector ranked in descending order of  $(\mathbf{X} - 1_N \bar{\mathbf{X}})^T (\mathbf{X} - 1_N \bar{\mathbf{X}})$ . This is a result taken from linear PCA, where we eigendecomposed the covariance matrix  $\Sigma$  to obtain the decreasing eigenvalues and eigenvectors to obtain the principal component scores and loadings. In fact, the matrix given by  $(\mathbf{X} - 1_N \bar{\mathbf{X}})^T (\mathbf{X} - 1_N \bar{\mathbf{X}})$  is equivalent to:

$$\begin{aligned} (\mathbf{X} - 1_N \bar{\mathbf{X}})^T (\mathbf{X} - 1_N \bar{\mathbf{X}}) &= \mathbf{X}^T \mathbf{X} - \bar{\mathbf{X}}^T 1_N^T \mathbf{X} - \mathbf{X}^T 1_N \bar{\mathbf{X}} + \bar{\mathbf{X}}^T 1_N^T 1_N \bar{\mathbf{X}} \\ &= \mathbf{X}^T \mathbf{X} - N \bar{\mathbf{x}} \bar{\mathbf{x}}^T \\ &= N \cdot \Sigma_X \end{aligned}$$

In other words, the first step of spherical PCA is to eigendecompose the scaled covariance matrix  $\Sigma_X$ ! This is where the ‘‘PCA’’ aspect of spherical PCA arises.

If  $\vec{z}_i = \bar{\mathbf{X}} + \hat{V}\hat{V}^T(\vec{x}_i - \bar{\mathbf{X}})$  are a change of basis to affine subspace  $V$ , then it can be shown that the minimizing pair  $(\vec{\eta}^*, \vec{\xi}^*)$  of loss function  $g(\vec{\eta}, \vec{\xi}) = \sum_{k=1}^N (\vec{z}_i^T \vec{z}_i + \vec{\eta}^T \vec{x}_i + \vec{\xi})^2$  is:

$$\begin{aligned}\vec{\eta} &= -H^{-1}\omega \\ \vec{\xi} &= -\frac{1}{N} \sum_{k=1}^N (\vec{z}_i^T \vec{z}_i + \vec{\eta}^T \vec{z}_i)\end{aligned}$$

where  $H$  and  $\omega$  are defined as:

$$\begin{aligned}H &= \sum_{i=1}^N (\vec{z}_i - \bar{\vec{z}})(\vec{z}_i - \bar{\vec{z}})^T \\ &= N\Sigma_z \\ \vec{\omega} &= \sum_{i=1}^N \left( \|\vec{z}_i^T \vec{z}_i\| - \frac{1}{N} \sum_{j=1}^N \|\vec{z}_j^T \vec{z}_j\| \right) (\vec{z}_i - \bar{\vec{z}})\end{aligned}$$

The  $H$  matrix can be seen as a centered matrix that is the covariance matrix of new coordinates  $\mathbf{Z}$  multiplied by number of observations  $N$ .  $\vec{\omega}$  is a vector where  $\omega_i$  is the centered  $i$ th  $z$ -coordinate scaled by a weight, where the weight is a centered magnitude of the  $i$ th  $z$ -coordinate.

The optimal parametrization  $(\hat{V}, \hat{c}, \hat{r})$  of the projection of  $\mathbf{X} \in \mathbb{R}^{N \times D}$  onto the sphere  $S_V(c, r)$  is:

$$\begin{aligned}\hat{V} &= (\vec{v}_1, \dots, \vec{v}_{d+1}) \\ \hat{c} &= -\frac{\vec{\eta}^*}{2} \\ \hat{r} &= \frac{1}{N} \sum_{k=1}^N \|\vec{z}_i - \hat{c}\|\end{aligned}$$

The projection map  $\hat{\Psi}$  of data matrix  $\mathbf{X}$  onto sphere  $S_{\hat{V}}(\hat{c}, \hat{r})$  is the projection map onto affine subspace  $\hat{c} + \hat{V}$ , given by:

$$\hat{\Psi}(\vec{x}_i) = \hat{c} + \frac{\hat{r}}{\|\hat{V}\hat{V}^T(\vec{x}_i - \hat{c})\|} \hat{V}\hat{V}^T(\vec{x}_i - \hat{c})$$

## 2.3 Algorithm

We have now defined spherical PCA (SPCA) to project the data  $\mathbf{X}$  down to single sphere  $S_V$ . In general, this single sphere will typically not be a sufficient approximation for the inherent manifold  $M$ . Instead, we partition the space  $\mathbb{R}^D$  into  $k$  disjoint subsets  $C_1, \dots, C_k$ .

For the  $k$ th disjoint subset, we can define a data matrix  $\mathbf{X}_k = \{X_i : X_i \in C_k\}$  that is a partition of the original data that lies within  $C_k$ . After applying SPCA to  $\mathbf{X}_k$ , we obtain spherical volume, center, and radius  $(\hat{V}_k, \hat{c}_k, \hat{r}_k)$  alongside projection map  $\Phi_k$  as a map from  $x \in C_k$  to  $y \in S_{\hat{V}_k}(\hat{c}_k, \hat{r}_k)$ . A spherelets estimation  $\hat{M}$  of the manifold  $M$  can be obtained by setting  $\hat{M} = \bigcup_{k=1}^K \hat{M}_k$ , where  $\hat{M}_k$  is the local SPCA in the  $k$ th region and  $\hat{M}_k = S_{\hat{V}_k}(\hat{c}_k, \hat{r}_k) \cap C_k$ .

We thus arrive at the spherical PCA algorithm. The basic strokes of the spherical PCA process are:

1. Subdivide the space (and dataset) into  $k$  partitions:  $\{C_k\}_{k=1}^K$  and  $\{M_k\}_{k=1}^K$ ,
2. Calculate a submanifold  $\hat{M}_k$  and projection map  $\hat{\Psi}_k$  for each submanifold,

3. Find the union of  $\{M_k\}_{k=1}^K$  and  $\Psi_k$  to obtain estimates  $\hat{M}$  and  $\Psi$ .

The specific algorithm is as follows:

---

**Algorithm 1** Spherelets

---

**Input:** Data matrix  $\mathbf{X}$ ; intrinsic dimension  $d$ ; partition  $\{C_k\}_{k=1}^K$

**Output:** Local estimated manifolds  $\hat{M}_k$  and projection map  $\hat{\Psi}_k, k \in \{1, \dots, K\}$ ; global estimated manifold  $\hat{M}$  of intrinsic manifold  $M$  and projection map  $\Psi$ .

- 1: **for** ( $k = 1 : K$ ) **do**
  - 2:   Define  $\mathbf{X}_{[k]} = \mathbf{X} \cap C_k$
  - 3:   Calculate  $\hat{V}_k, \hat{c}_k, \hat{r}_k$
  - 4:   Calculate  $\hat{\Psi}_k(x) = \hat{c}_k + \frac{\hat{r}_k}{\|\hat{V}_k \hat{V}_k^T (x - \hat{c}_k)\|} (x - \hat{c}_k)$
  - 5:   Calculate  $\hat{M}_k = S_{\hat{V}_k}(\hat{c}_k, \hat{r}_k) \cap C_k$
  - 6: **end for**
  - 7: Calculate  $\Psi(x) = \sum_{k=1}^K \mathbf{1}_{\{x \in C_k\}} \hat{\Psi}_k(x)$ , and  $\hat{M} = \bigcup_{k=1}^K \hat{M}_k$ .
- 

### 3 Discussion

Previously within this paper, I have incorporated parts of the geometric intuition and connections to linear PCA in other sections but will rehash it briefly in this section as well. Within the Section 2.1, I also discuss the Main Theorem and its theoretical guarantees for the relative performance of spherelets in terms of covering number.

Within this section, I will discuss the various strengths and weaknesses that spherelets provide relative to local linear PCA using hyperplanes. I will also incorporate extensions and applications of spherelets to other aspects of dimension reduction and data visualization.

#### 3.1 Strengths

The two main draws for spherical PCA are its flexibility and its ability to offer out of sample testing without reinitialization of the entire algorithm.

- **High curvature.** As mentioned prior, the motivating desire behind spherical PCA is to approximate manifolds in lower dimensions that other fully linear methods are not able to capture. Spherical PCA projects data down to
- Performs well in areas with high curvature that local PCA can't approximate
- **Out of sample testing.** Can perform OOS assessments and returns the underlying manifold

#### 3.2 Weaknesses

While spherical PCA is capable of fitting manifolds that have higher degrees of curvature than linear PCA to a certain extent, spherelets still struggle with areas, sometimes in similar regards to linear PCA. Main disadvantages of spherelets include sensitivity to non-uniform dimensions and curvature along the manifold and computational inefficiency when paired with a poor choice of partitioning algorithm and choice of intrinsic manifold dimensionality  $d$ .

- Struggles with areas of non-uniform curvature
- Struggles with non-uniform dimensions



- Must specify inherent dimension  $d$
- Computationally expensive
- Dependent on choice of manifold subsetting

## 4 Examples

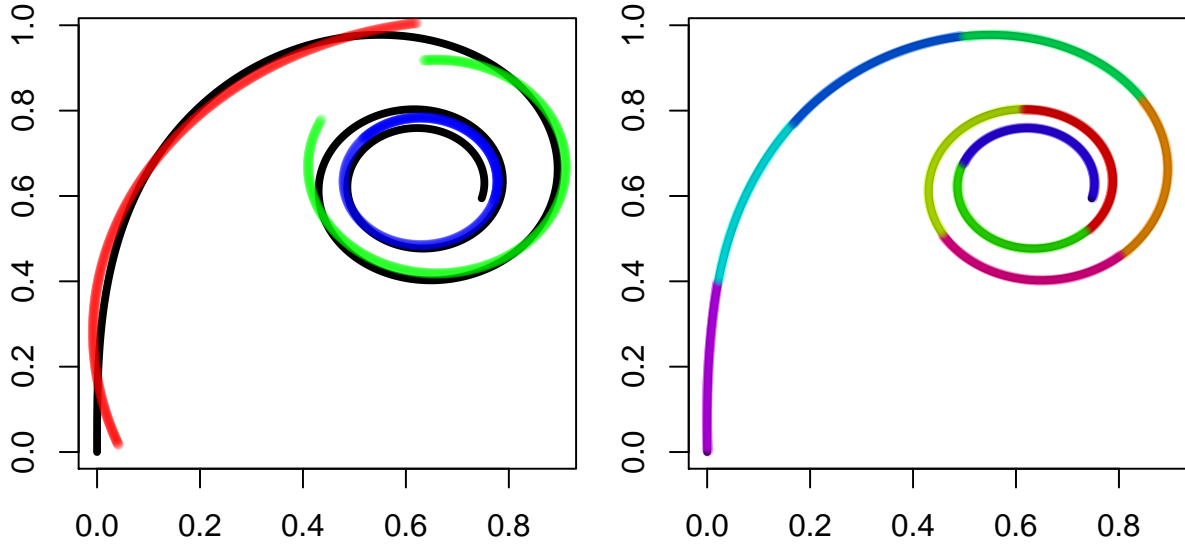
To generate numerical examples, I used the `SPCA` and `SS_calc` functions written by co-author Minerva Mukhopadhyay (mmukhopadhyay 2019). The `SPCA` function takes in a matrix of  $N$  observations  $\vec{x}_i \in \mathbb{R}^D, i \in 1, \dots, N$  and returns the error given by spherical and local PCA (`SS` and `SS_new`), as well as the projected values `Y_D`.

### 4.1 Euler Spiral

The Euler spiral is a curve in  $\mathbb{R}^1$  that has a curvature that changes linearly with arc length. In other words,  $\kappa(s) = s$ . The Euler spiral can be parametrized as follows:

$$\begin{bmatrix} x(s) \\ y(s) \end{bmatrix} = \begin{bmatrix} \int_0^s \cos(t^2) dt \\ \int_0^s \sin(t^2) dt \end{bmatrix} \quad s \in [0, 4]$$

We use the Euler spiral as a demonstration to see how spherical PCA is able to handle regions with curvature.



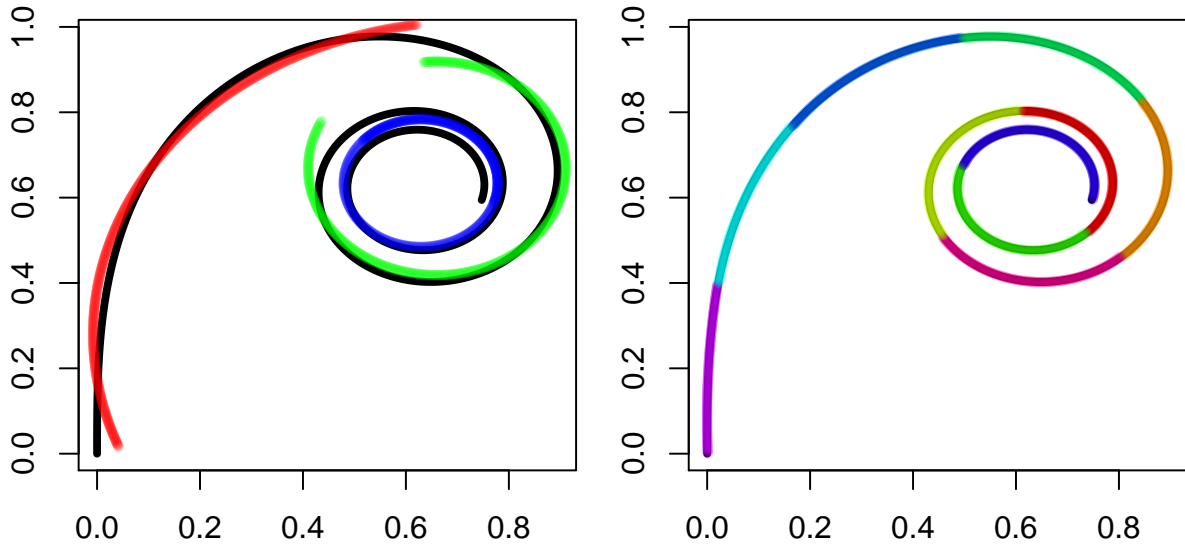


Figure 1: Spherical PCA performed on an Euler spiral with  $k = 3, 8$ .

#### 4.2 Helix

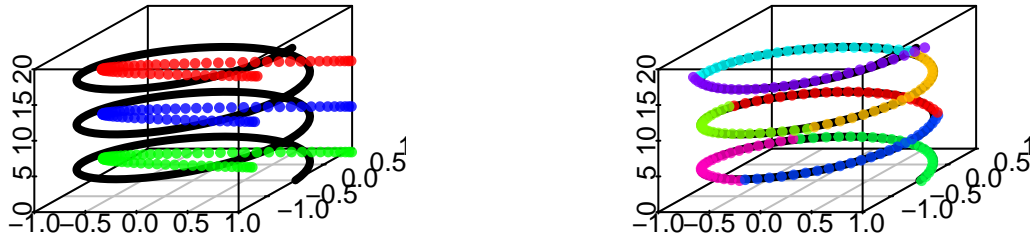


Figure 2: Spherical PCA performed on a helix with  $k = 3, 8$ .

#### 4.3 Cylinder

## [1] 0.00269493

## [1] 7.353559e-05

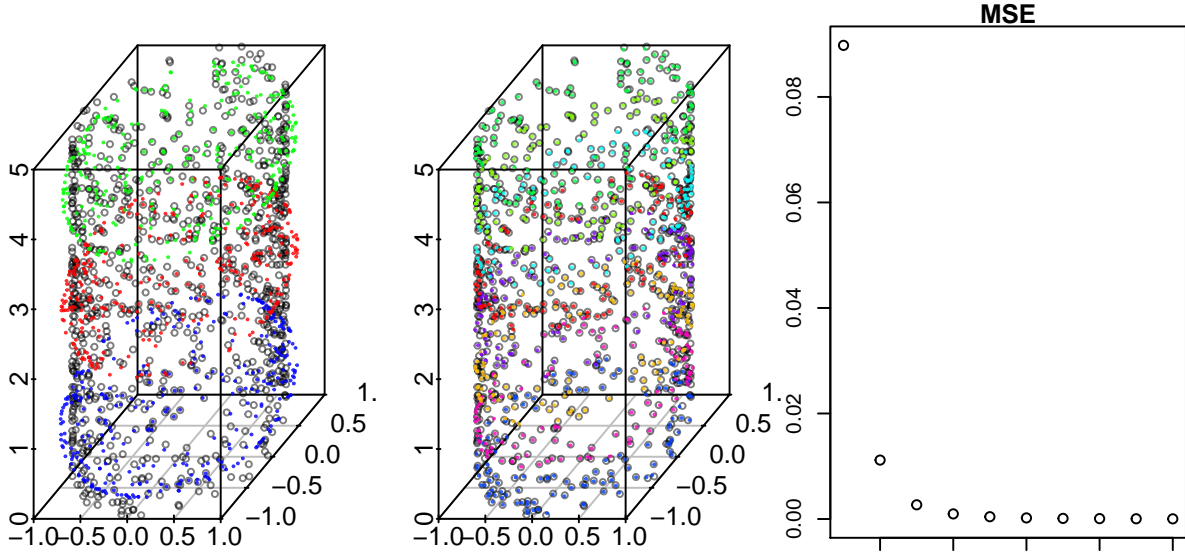


Figure 3: Spherical PCA performed on a cylinder with  $k = 3, 8$ .

We see that SPCA is not fully capable of handling a cylinder.

## 5 References

- Beygelzimer, Alina, Sham Kakade, and John Langford. 2006. “Cover Trees for Nearest Neighbor.” ACM.
- Didong Li, Minerva Mukhopadhyay, and David B Dunson. 2019. “Efficient Manifold Approximation with Spherelets.” *arXiv*, February.
- Duda, Richard O, David G Stork, and Peter E Hart. 2000. *Pattern Classification*. 2nd ed. Wiley.
- Hurtado, Jorge E. 2012. *Structural Reliability*. Vol. 17. Springer-Verlag.
- Karypis, George, and Vipin Kumar. 1998. “A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs.” *SIAM Journal on Scientific Computing* 20 (1): 359–92. doi:10.1137/s1064827595287997.
- Lee, John A, and Michel Verleysen. 2008. *Nonlinear Dimensionality Reduction*. Springer Science.
- Lee, Seunggeun, Fei Zou, and Fred A Wright. 2010. “Convergence and Prediction of Principal Component Scores in High-Dimensional Settings.” *The Annals of Statistics* 38 (6): 3605–29. doi:10.1214/10-aos821.
- Li, Didong, and David B Dunson. 2017. “Efficient Manifold and Subspace Approximations with Spherelets.” *arXiv*, June. ResearchGate.
- mmukhopadhyay. 2019. “Efficient Manifold Learning Using Spherelets.” Github. April 9.
- Szlam, Arthur. 2009. “Asymptotic regularity of subdivisions of Euclidean domains by iterated PCA and iterated 2-means.” *Applied and Computational Harmonic Analysis*, March. Academic Press. <https://www.sciencedirect.com/science/article/pii/S1063520309000190>.