

Directional Priors in Manifolds

A Causal Interpretation

Daniel Posmik

2025-04-03

1 Note

Prof. Oganisian – as discussed, I have attached the abstract and the introduction of my paper. It already exists in its rough form, but I have not yet written the causal section. Please let me know what suggestions and thoughts you have, I would be eager to incorporate them.

2 Abstract

This paper introduces an approach for incorporating directional prior information into manifold learning, specifically within the principal manifolds framework. We formulate a Bayesian method that leverages directional statistics to update prior knowledge about the orientation of data points projected onto manifolds. With a spherical example in \mathbb{R}^3 , we demonstrate how directional uncertainty can be modeled using the von Mises distributions. We exploit established conjugacy results, letting the concentration parameter encode the (fitted) local Gaussian curvature of the manifold at a projection point. Within this (empirical) Bayesian paradigm, we also explore the Manifold analogue of a causal regression discontinuity estimator where pre- and post treatment manifolds are fitted. Treating the curvature of pre-treatment manifolds as prior variance information, we explore the construction of a causal estimator as the posterior projection angles change in the post-treatment period. This paper outlines the foundation for incorporating prior directional knowledge into the Principal Manifolds framework, suggesting applications in fields like cancer medicine, and exploring the viability of causal effect identification within a Bayesian paradigm.

3 Introduction

Manifold learning is dimensionality reduction technique that has proven useful in settings where data are high-dimensional and non-linear. Often, manifold learning algorithms are used

when the topological structure of the data are to be preserved in a statistical learning¹. Within the manifold learning framework, data are assumed to live on a lower dimensional manifold and are corrupted by high-dimensional noise. We say that D -dimensional data can be embedded in d -dimensional manifold where $d \leq D$ but generally $d \ll D$.

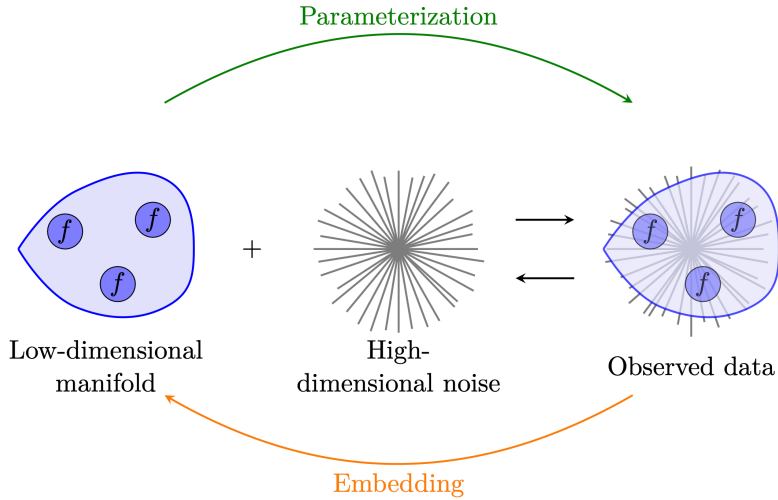


Figure 1: Manifold Learning

Within the principal manifolds framework (Meng2021) – a replicable and flexible framework for manifold learning – the process of fitting a manifold to our data contains multiple steps. The key idea of the fitting step is that we fit a d -dimensional manifold to our D -dimensional data by minimizing the sum of squares between our data and the proposed manifold. An important extension to linear dimensionality reduction, i.e. the principal components algorithm (PCA), is that we allow our proposed manifold to preserve underlying topological structure of our data. In a way, manifold learning reduces the dimensionality of data with an explicit focus on the topology of it. We note that – although certainly intuitive – this topological structure is not only limited to spatial abstraction, but may be extended to arbitrary dimensions of interest. This framework was pioneered as an extension to the PCA algorithm with curves (HastieStuetzle1989, Tibshirani1992) and has since found a myriad of applications in higher-dimensional settings.

Now, consider a setting where we have fit a manifold \mathcal{M}_d to our D -dimensional data by means of minimizing the orthogonal distance between the data and the manifold. We consider this manifold fixed and will not touch on the fitting procedure itself. Given \mathcal{M}_d , for each data point, i.e. the row vector $[x_{11} \cdots x_{1D}]^T$, we can now define the point on \mathcal{M}_d , say $f([x_{11} \cdots x_{1D}]^T)$. This point minimizes the distance between x_i and $f(x_i)$. We want to stress again that this

¹For an introduction, see (Meila2023)

procedure does not mean we are fitting the manifold to the data, we are simply retrieving the distance-minimizing projection point. We write

$$\arg \min_{f \in \mathcal{F}} \|x^* - f(x^*)\|_2$$

where we consider each projection function f to be a member of an arbitrary Sobolev space \mathcal{F} . We define the distance metric as the L^2 distance.

If there exists only one projection point $f(x_i)$ for every x_i , every $f \in \mathcal{F}$ is one-to-one and onto (“bijective”) mapping. We find it interesting to highlight that the projection functions in the PCA algorithm are inherently bijective, and for inferential purposes, this is a property that is often taken for granted². In a manifold learning framework, this is no longer the case. Albeit highly interesting, due to the limited scope of this paper, we shall treat this scenario as an edge case, reserving rigorous treatment for the blissful times that follow the author’s qualifying exam.

Now, given the data $[\{x_i\}_{i=1}^n, \{f(x_i)\}_{i=1}^n]$, we can reparameterize our space into polar coordinates to obtain a vector representation of the collection $f \in \mathcal{F}$. Converting a Cartesian parameterization in space with D dimensions into polar coordinates yields the d -dimensional vector $[r_i^*, \theta_{i,1}, \dots, \theta_{i,D-1}]$, i.e. one radius r_i^* and a set of $D - 1$ angles suffice to characterize each point x_i ’s location in space.

Recognize that the parameter r^* is not random. This is because it is simply the result from our previous projection distance-minimizing procedure. Usually, polar parameterizations assume that all angles and radii are centered at the origin. Luckily, simple vector addition and subtraction readily generalizes our parameterizations in space. For instance, to obtain the vector from the point x_i and $f(x_i)$, we simply subtract $f(x_i) - x_i$. It is important that the issue of defining the origin is explicitly clarified when dealing with directional information. For simplicity, we will henceforth consider data centered at the origin.

We are now ready to introduce the Bayesian framework with directional priors.

²This is because a principal axes is a straight line, i.e. neither convex or concave. Although a point’s distance to its projection may be co-minimal across ≤ 2 dimensions, it only has one $f(x_i)$ in one principal axis.

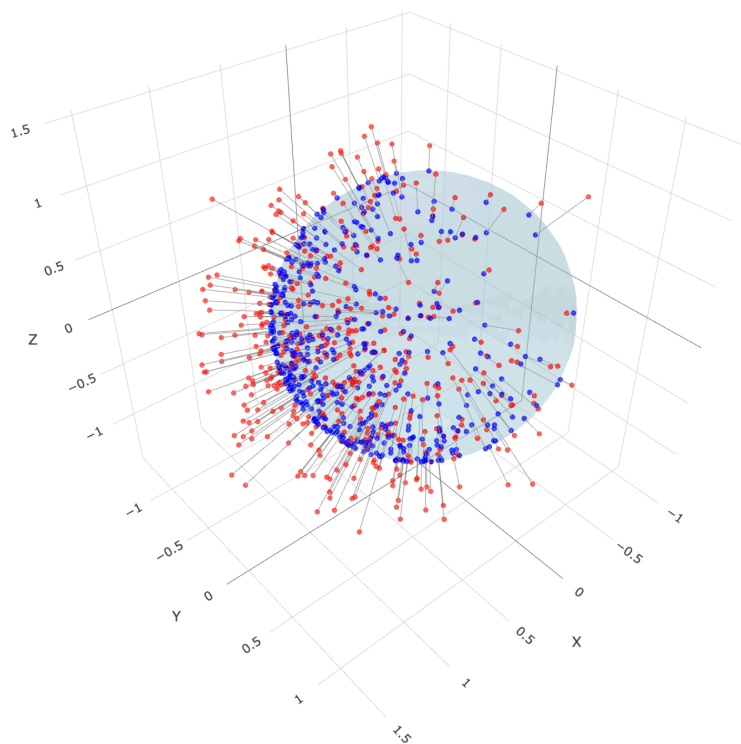


Figure 2: Noisy data projected on \mathcal{M}_3 ; the unit sphere in \mathbb{R}^3