Bayesian Topological Weights in Causal Inference – Toward a Manifold Regression Discontinuity Estimator

PHP2530

Daniel C. Posmik

May 9, 2025

1. 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 learning task[1]. 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 $d \ll D$.

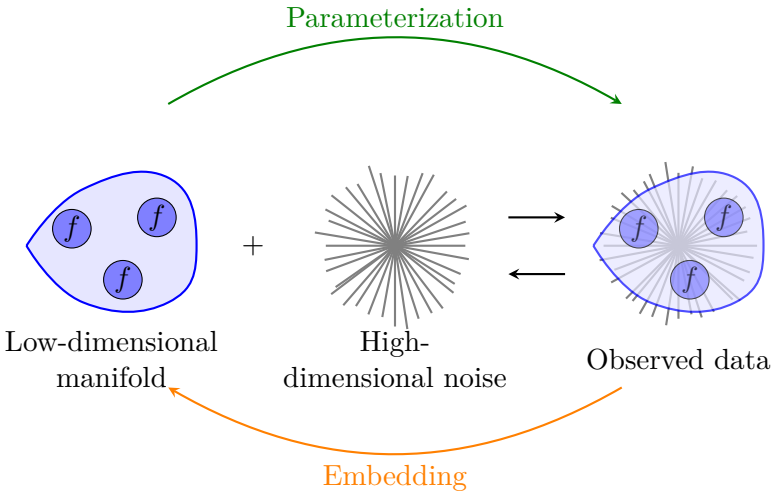

Figure 1: Manifold Learning: Parameterization vs. Embedding

Within the principal manifolds framework (Meng and Eloyan, 2021) – a replicable and flexible framework for manifold learning – we effectively extend linear dimensionality reduction, i.e. the principal components algorithm (PCA), by allowing our proposed manifold to preserve underlying topological structure of our data. In a way, manifold learning reduces

1. For a suitable introduction, see Meilă and Zhang (2023)

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 (Hastie and Stuetzle (1989), Tibshirani (1992)) and has since found a myriad of applications in higher-dimensional settings.

We propose a causal regression discontinuity (RD) method within the manifold framework. To effectively account for the change in underlying topological structure, we propose Bayesian weights that incorporate prior distributional information from the pre-treatment period. Before we dive into the causal estimator, we provide important background on the concept of directional priors in manifolds in the next section. The paper is rounded by a brief presentation, simulation, and discussion on the manifold RD estimator.

2. Manifold Learning and Directional Information

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 $f\left([x_{11} \cdots x_{1D}]^T\right)$. This point minimizes the distance between x_i and $f(x_i)$. We want to stress again that this 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 function space $\mathcal{F}$ (for convenience, assume $\mathcal{F}$ is a Sobolev space). 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 a 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 granted[2]. 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.

2.1 Directional Priors

Given the projected data $[\{x_i\}_{i=1}^n, \{f(x_i)\}_{i=1}^n]$ on the manifold, 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 $[r_i^*; \theta_{i,1}, \cdots, \theta_{i,D-1}]$, i.e. one radius r_i^* and a set of $D-1$ angles suffice to characterize each point x_i's location in space.

2. This 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.

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.

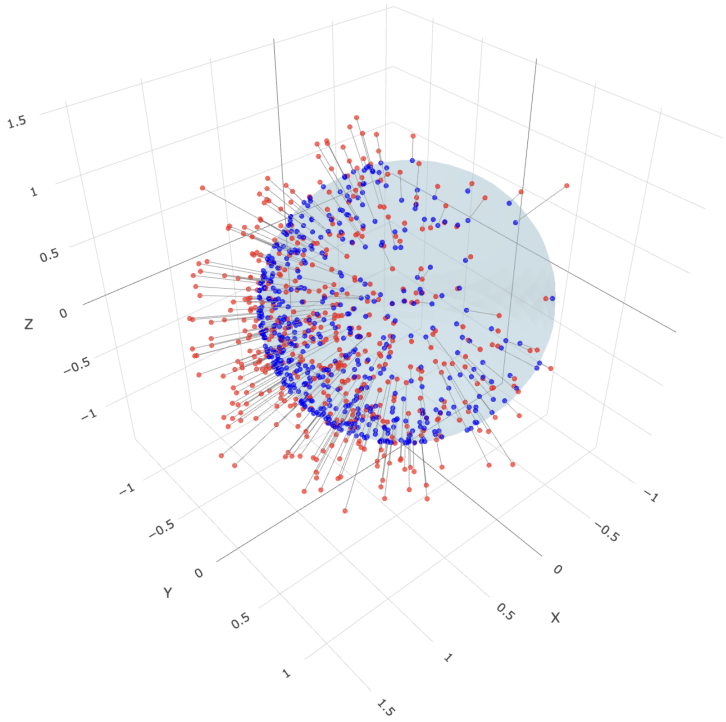

Figure 2: Noisy data projected on $\mathcal{M}_3$; the unit sphere in $\mathbb{R}^3$

In contrast to the radii, the $\{D-1\}$-dimensional vector of angles, $\theta_i := [\theta_{i,1}, \cdots, \theta_{i,D-1}]^T$ is random. In simple terms, $\{r_i, \theta_i\}$ is what parameterizes the realized sample $X_i = x_i$ in space. If we draw multiple samples, the random sampling variation in θ_i is what captures the randomness. In biomedical applications, such as cancer medicine, we may have reliable prior information (i.e. from previous trials or expert knowledge) on directional trends of malignant growths. Within a Bayesian framework, using our data to update these prior directional information offers a principled, probabilistic solution to complex inference problems in settings where directionality is a key piece of information.

2.2 Motivating Example: 3D Sphere

For the remainder of this paper, we will stick to the example of a 3D sphere. The noisy data is projected on the sphere of choice, for simplicity we choose the unit sphere. This

yields the data $\{f(x_i)\}_{i=1}^n$. These data are parameterized in a polar coordinate system with $\{r_i, \theta_i\}$, where the (non-random) radius $\vec{r_i}$ is the vector from the origin to $f(x_i)$. Since we are in 3D space, we have two angles. One angle, $\{\theta\}_{i=1}^n$, which parameterizes the angle from the origin on the XY-plane. The angle $\{\phi\}_{i=1}^n$ is the angle between the XY-plane and the point $f(x_i)$.

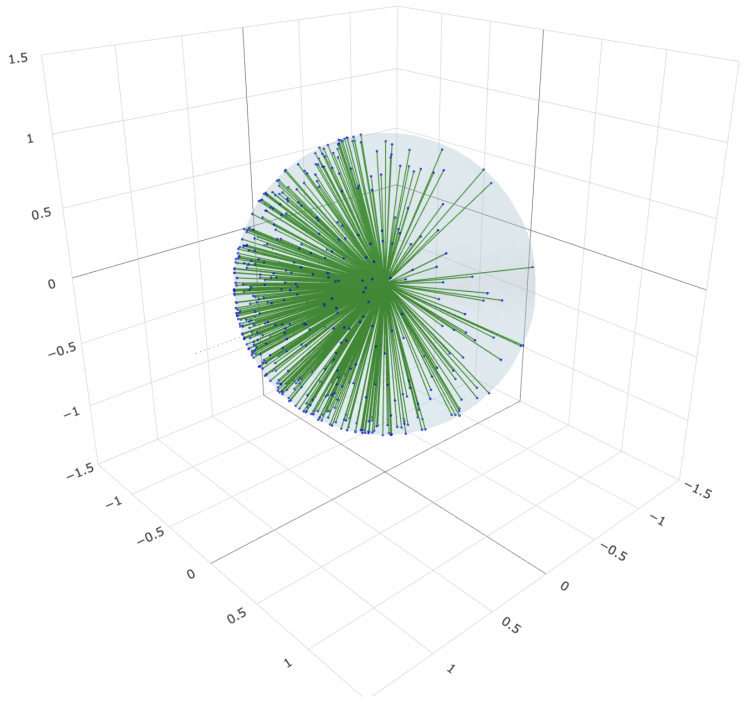

Figure 3: Polar parameterization with $\{r_i, \theta_i, \phi_i\}$ of projected data

2.3 Formulating Directional Priors

We are now ready to formally define directional priors. A directional prior with mean direction $\mu \in [0, \pi]$ allows us to incorporate information about where projected points should concentrate on the manifold. We only posit a prior on the random set $\{\theta_i\}_{i=1}^n$, expressing prior distributional beliefs on point concentrations in the XY-plane. For this example, we consider angles $\{\phi_i\}_{i=1}^n$ fixed, making our example univariate. Extension to the multivariate case is straightforward using multivariate analogs of our likelihood and prior distributions.

Our posterior distribution on the random vector θ can be expressed as follows:

$$f_{\theta|\mu}(\theta|\mu) \propto \left\{ \prod_{i=1}^{n} f_{\theta_i|\mu}(\theta_i|\mu) \right\} \cdot f_{\mu}(\mu)$$

where μ is our mean vector with prior $\mu \sim f_\mu(\mu)$. For our likelihood, we initially consider the wrapped normal distribution with density:

$$f(\theta|\mu, \sigma^2) = \frac{1}{2\pi\sigma^2} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(\theta - \mu + 2\pi k)^2}{2\sigma^2}\right)$$

However, this infinite series presents computational challenges. Instead, we use the von Mises distribution, which provides a more tractable circular analog to the normal distribution:

$$f(\theta|\mu, \kappa) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}$$

where θ is the angle, μ is the mean direction, $\kappa \geq 0$ is the concentration parameter (analogous to precision, or inverse variance), and $I_0(\kappa)$ is the modified Bessel function of first kind of order 0. As κ increases, the distribution concentrates more tightly around μ. When $\kappa = 0$, the distribution becomes uniform on the circle. The von Mises distribution scales naturally for hyperspheres in $\mathbb{R}^p$, making it ideal for spherical settings in manifold learning applications.

2.4 Variance Encodes Curvature

In smooth manifolds, choosing the inverse Gaussian curvature as our variance term yields a principled approach to encode topological structure into the variance paramater. Intuitively, in regions with little topological variation, we might update our projection more liberally. Conversely, if a small change in projection angle moves our data point into a topologically distinct region, we would be more conservative with updates. This approach has an elegant connection to the score test in likelihood-based inference, which uses the slope at restricted MLE estimates as a similarity measure. Considering κ as random nuisance parameter also provides a bridge to the mixed models literature.

The Gaussian curvature of a surface can be defined as the product of the principal curvatures:

$$K = k_1 \cdot k_2$$

where k_1 and k_2 are the principal curvatures at a point on the surface. For a sphere of radius R, due to perfect symmetry, both principal curvatures equal $\frac{1}{R}$, giving a Gaussian curvature of $K = \frac{1}{R^2}$. For a unit sphere, the Gaussian curvature is $K = 1$, an intrinsic property related to the total curvature integrated over the entire sphere equaling 4π by the Gauss-Bonnet theorem. This curvature-based approach provides a meaningful way to weight observations based on their location in the manifold's topological structure. Areas with high curvature (significant bending of the manifold) will contribute less to updating our beliefs than areas with lower curvature (flatter regions), which aligns with our geometric intuition about confidence in projections.

2.5 Conjugacy Results

A key advantage of the von Mises distribution is conjugacy, which allows for straightforward Bayesian updates of our directional priors. Following the work of Mardia and El-Atoum (1976), a von Mises prior with mean direction μ and concentration parameter κ is conjugate to the likelihood. After observing angles $\theta_1, \ldots, \theta_n$, the posterior distribution takes the form:

$$f(\mu_i, \mu^* | \{\theta_i\}_{i=1}^n) \propto \exp(\kappa \cdot \sum_{i=1}^{n} \cos(\theta_i - \mu_i) + \kappa^* \cdot \sum_{i=1}^{n} \cos(\mu_i - \mu^*))$$

where angles $\{\theta_i\}_{i=1}^n \sim \mathcal{VM}(\mu_i, \kappa)$ and the prior is $\mu_i \sim \mathcal{VM}(\mu^*, \kappa^*)$. In our case, the data were generated with $\mu_i = \mu = \text{circular}(0)$.

The posterior form reveals a shrinkage estimator structure. The component involving observed angles $\{\theta_i\}_{i=1}^n$ is weighted by κ, which in our framework represents the curvature at the projected point. The difference between the data mean and prior mean is weighted by the prior concentration parameter κ^*. This structure extends shrinkage estimation to our topological interpretation: when the data concentration parameter κ is low, more weight is attributed to our prior information.

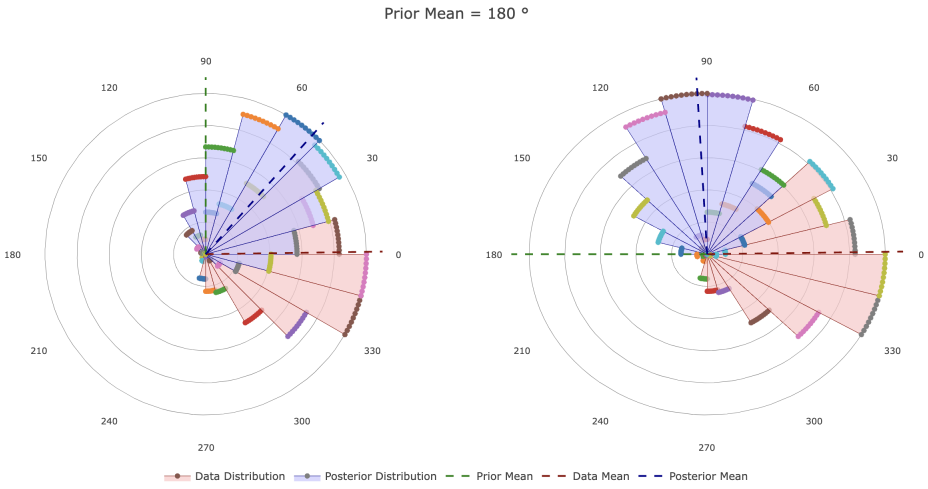

Figure 4: Posterior distribution under prior θ_0: Left $\theta_0 = \pi/2$; Right: $\theta_0 = \pi$

Figure 4 demonstrates that the posterior mean effectively balances the prior and data means. This occurs because our manifold is a sphere with equivalent curvature at every point. For manifolds with heterogeneous curvature, we would observe less reprojection with respect to the prior mean in regions where curvature is more extreme, thus preserving the topological characteristics of the data.

The conjugacy relationship for von Mises distributions has an elegant geometric interpretation. For a von Mises likelihood $f(x|\mu) \propto \exp(\kappa \cos(x - \mu))$ and a von Mises prior on the mean direction $\pi(\mu) \propto \exp(\kappa_0 \cos(\mu - \mu_0))$, the posterior distribution is also von Mises: $\pi(\mu|x) \propto \exp(\kappa_n \cos(\mu - \mu_n))$. The posterior parameters can be expressed in terms of resultant vectors:

$$\kappa_n \cos(\mu_n) = \kappa_0 \cos(\mu_0) + \kappa \cos(x)$$
$$\kappa_n \sin(\mu_n) = \kappa_0 \sin(\mu_0) + \kappa \sin(x)$$

These equations can be rewritten in vector form as:

$$\kappa_n = \sqrt{(\kappa_0 \cos(\mu_0) + \kappa \cos(x))^2 + (\kappa_0 \sin(\mu_0) + \kappa \sin(x))^2}$$

and

$$\mu_n = \text{atan2}(\kappa_0 \sin(\mu_0) + \kappa \sin(x), \kappa_0 \cos(\mu_0) + \kappa \cos(x))$$

This representation yields a powerful geometric interpretation: the posterior mean direction and concentration are determined by the vector sum of the prior and likelihood information vectors in $\mathbb{R}^2$. Each vector has direction given by unit vectors pointing at angles μ_0 and x respectively, with magnitude determined by concentration parameters κ_0 and κ. The resultant vector's length determines the posterior concentration parameter κ_n, while its direction determines the posterior mean direction μ_n.

This insight demonstrates that when prior and likelihood point in similar directions, the posterior becomes more concentrated (higher κ_n), reflecting increased certainty. Conversely, conflicting directions lead to a less concentrated posterior (lower κ_n), reflecting uncertainty.

The posterior mean naturally represents a weighted compromise between prior and data, with weights proportional to their respective concentration parameters. For multiple observations, the vectors from each observation simply add to the prior vector, providing an efficient sequential updating mechanism. This vector-based interpretation elegantly connects Bayesian inference on the circle with vector addition in the plane, making it particularly suitable for our manifold learning framework where directional data naturally arises from projections onto curved surfaces.

3. A Causal Extension – Toward a Manifold RD Estimator

Our goal is to propose an estimator that exploits topological information from pre- and post-treatment manifolds. Before we begin, recall the definition of a regression discontinuity (RD) estimator when treatment is administered at time t^*.

$$\tau(t^*) = \lim_{\delta \to 0} \mathbb{E}[\theta | Z = 1, t = t^* + \delta] - \lim_{\delta \to 0} \mathbb{E}[\theta | Z = 0, t = t^* - \delta]$$

As it relates to the principal manifolds framework, after treatment happens at $t = t^*$, we can fit the post-treatment manifold $\mathcal{M}_d^{t \geq t^*}$. In the spirit of RD-based causal estimation, we want to emphasize that $\mathcal{M}_d^{t \geq t^*}$ is *not* fit iteratively for each post-treatment unit, i.e. we do not obtain the post-treatment unit by combining the post-treatment units with pre-treatment units. Instead, we fit two manifolds with disjoint pre- and post-treatment data.[3]

3. That being said, it may be interesting to let the choice of manifold (e.g., the shape, highest order, etc.) from the pre-treatment case inform the post-treatment case, effectively borrowing prior topological information on a more global level. This is a really interesting direction in the Bayesian paradigm and touches on the notion of persistent homology. Although this discussion goes beyond the scope of this paper, it highlights how nuanced various levels/ sources of topological information can be.

Namely, we observe the outcomes of m additional units in the immediate post-treatment period, and fit $\mathcal{M}_d^{t\geq t^*}$ with data $\{\theta_t\}_{i=n+1}^{m}|_{t\geq t^*}$. The mean projection angle of $\mathcal{M}_d^{t\geq t^*}$, $\bar{\theta}^{t\geq t^*}$, is the post-treatment mean projection angle. It reflects the directionality under treatment. The following figure presents this admittedly complex idea visually.

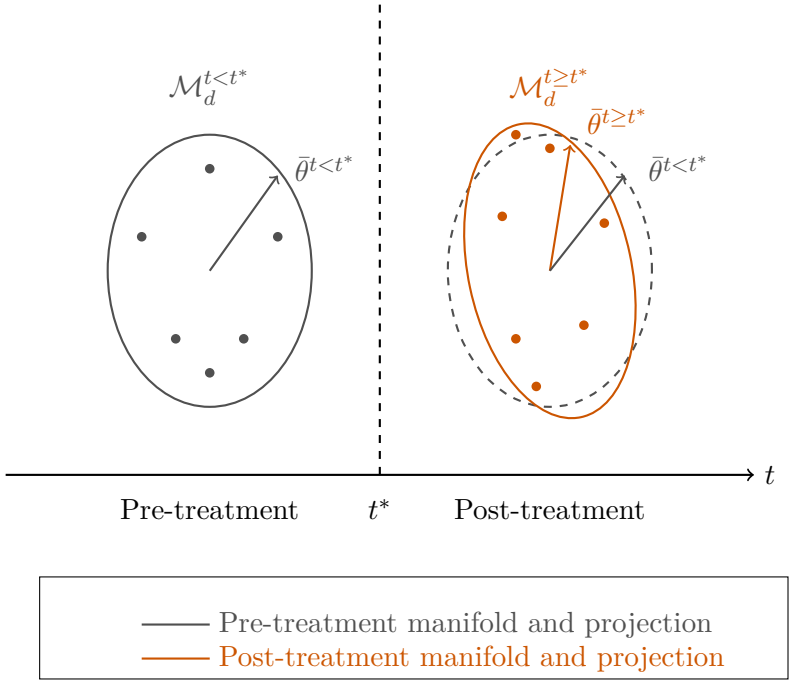

Figure 5: The Manifold RD Estimator

3.1 The Manifold RD Estimator

It is common to estimate the causal effect, write $\hat{\tau}(t^*)$, using a weighting approach. Units that are located closer to t^* are weighted more, while units that are located further away are weighted less. This makes sense in the context of the identification assumption underlying the RD estimator, i.e. units close to the cutoff only differ in treatment assignment. Thus, treatment is considered quasi-random.

Our crux is to translate this weighting scheme into a topological setting. Intuitively, we want to achieve the following: For each additional unit we observe after the treatment has been administered, evaluate whether this observation is *topologically similar* to our pre-treatment mean. If it is, we want to assign a larger weight to it. If it is not, we want to assign less weight to it. Remember that we formalized the notion of topological similarity using the curvature $\kappa|\theta_i$, i.e. for unit i, the curvature at the projection angle encodes its topological neighborhood. Thus, if some unit is embedded close to what we would expect from our pre-treatment data, we want to assign a lot of weight to it since it is topologically similar.

This thought process yields the following weighting approach. Let $\hat{\tau}(t^*)$ be the weighted treatment effect estimator,

$$\hat{\tau}(t^*) = \sum_{j=n+1}^{m} \frac{w_j \cdot \theta_j^{t \geq t^*}}{m-n} - \sum_{i=1}^{n} \frac{w_i \cdot \theta_i^{t<t^*}}{n}$$

For all pretreatment observations, $i \in \{1, \cdots, n\}$, the weight w_i simply quantifies the difference of unit i's projection angle θ_i from the mean pre-treatment projection angle $\bar{\theta}^{t<t^*}$. This is analogous to the canonical weighting scheme in literature. When we enter the post-treatment period, with observations $j \in \{n+1, \cdots, m\}$, the weight w_j is drawn from a posterior with prior around the pre-treatment projection angle. Its likelihood is updated for each additional point in the post-treatment period. Thus, we formally define our weight function as follows:

$$w_i(\kappa|\theta) = \frac{1}{\|\kappa_i^{t<t^*}|\{\theta_i\}_{i=1}^{t} - \kappa^{t<t^*}|\bar{\theta}^{t<t^*}\|}$$

Here, $\kappa^{t<t^*}|\theta_i$ denotes the Gaussian curvature on the pre-treatment manifold, evaluated at θ_i. Meanwhile, $\kappa^{t<t^*}|\bar{\theta}^{t<t^*}$ denotes the Gaussian curvature on the pre-treatment manifold, evaluated at the mean pre-treatment projection angle. In the post-treatment period, this quantity is always non-random.

Before we proceed, we want to emphasize the difference in interpretation of $\{\theta_i\}_{i=1}^{t}$ from the pre-treatment to the post-treatment case. In the pre-treatment case, $\{\theta_i\}_{i=1}^{t}$ represents the prior distribution using pre-treatment observations. In the post-treatment case, $\{\theta_i\}_{i=1}^{t}$ becomes a posterior distribution, where the prior distribution is updated with each post-treatment observation in $j \in \{n+1, \cdots, t\}$. This implies that we can express our data vector as a disjoint union of sets:

$$\{\theta_i\}_{i=1}^{t} = \begin{cases} \{\theta_i\}_{i=1}^{t}, & \text{if } t < t^* \\ \{\theta_i\}_{i=1}^{n} \cup \{\theta_j\}_{j=n+1}^{t}, & \text{if } t \geq t^* \end{cases}$$

In the pre-treatment case, we have a canonical von Mises distribution. In the post-treatment case, we make use of our conjugacy result from Section 2.1. Naturally, in a causal setting, we will focus on the case $t \geq t^*$, i.e. when post-treatment data is available. This yields our hierarchical manifold RD estimator:

$$\begin{aligned} &\tau|w_i \\ &\quad w_i = w_i(\kappa^{t<t^*}|\theta_i) \\ &\quad \theta_i \sim \mathcal{VM}(\mu, \sigma^2) \end{aligned}$$

where

$$\mu = \frac{1/\kappa^{t<t^*}}{1/\kappa^{t<t^*} + 1/\kappa^{t \geq t^*}} \cdot \bar{\theta}^{t<t^*} + \frac{1/\kappa^{t \geq t^*}}{1/\kappa^{t<t^*} + 1/\kappa^{t \geq t^*}} \cdot \bar{\theta}^{t \geq t^*}$$

$$\sigma^2 = \sqrt{\left[\kappa^{t<t^*} \cos(\bar{\theta}^{t<t^*}) + \kappa^{t \geq t^*} \cos(\bar{\theta}^{t \geq t^*})\right]^2 + \left[\kappa^{t<t^*} \sin(\bar{\theta}^{t<t^*}) + \kappa^{t \geq t^*} \sin(\bar{\theta}^{t \geq t^*})\right]^2}$$

by the conjugacy results of Section 2.5.

We want to emphasize that the Bayesian inference procedure is not the causal estimator $\hat{\tau}$ directly. Rather the weights as a function of κ (evaluated at θ). Posterior inference through the above mechanism is possible as soon as we observe at least one post-treatment unit. An immediate advantage of this estimation procedure is that we can update our estimates readily for each new observation.

The manifold RD estimator is similar to weighted RD estimators in literature. A common and intuitive estimation procedure for RD estimators is the Local Linear Regression (LLR) framework, i.e. measure the LATE as the difference in post- and pre-treatment means, given that they are sufficiently close to the treatment cutoff t^*. Branson et al. (2019) considers a Bayesian RD setting where a Gaussian process prior is placed on the pre- and post-treatment mean response functions. We believe this work, although not directly related to the use of topological information in causal inference, encapsulates the desire for more flexible, data-driven, and principled causal inference through Bayesian paradigms.

3.2 Estimator Properties and Simulation

An analytical derivation of our estimator's properties is simplified greatly by our conjugacy results. Naturally, outside of the von Mises parameterization, estimator properties can be determined by posterior sampling and numerical methods. To provide further insight into the estimator, we will analytically derive its first two moments and provide a brief simulation.

The expectation of $\hat{\tau}(t^*)$, where we let $w_{i,j} = w_{i,j}(\kappa|\mu)$ denote the respective weights, is the following:

$$\mathbb{E}(\hat{\tau}(t^*)) = \sum_{j=n+1}^{m} \frac{w_j \cdot \theta_j^{t \geq t^*}}{w_j} - \sum_{i=1}^{n} \frac{w_i \cdot \theta_i^{t<t^*}}{w_i}$$

Intuitively, we see that our manifold RD estimator is a topologically weighted average of post-treatment observations, that accounts for the pre-treatment "baseline". For our variance,

$$\begin{aligned}
\mathbb{V}(\hat{\tau}(t^*)) &= \mathbb{V}\left(\sum_{j=n+1}^{m} \frac{w_j \cdot \theta_j^{t \geq t^*}}{w_j} - \sum_{i=1}^{n} \frac{w_i \cdot \theta_i^{t<t^*}}{w_i} \right) \\
&= \mathbb{V}\left(\sum_{j=n+1}^{m} \frac{w_j \cdot \theta_j^{t \geq t^*}}{w_j} \right) + \mathbb{V}\left(\sum_{i=1}^{n} \frac{w_i \cdot \theta_i^{t<t^*}}{w_i} \right) \\
&\underbrace{-2\,\mathrm{Cov}\left(\sum_{j=n+1}^{m} \frac{w_j \cdot \theta_j^{t \geq t^*}}{w_j}, \sum_{i=1}^{n} \frac{w_i \cdot \theta_i^{t<t^*}}{w_i} \right)}_{=0}
\end{aligned}$$

Since the pre-treatment and post-treatment observations are assumed to be independent, the covariance term equals zero. The evaluation of the variance term depends on the underlying curvature of the manifold, and is thus highly dependent on the choice of manifold.

Returning to our spherical example, our moment expressions simplify greatly. This is because in a sphere, our curvature is constant everywhere, thus yielding no variation in the weights. Therefore, our manifold RD estimator reduces to our canonical RD estimator, i.e. the difference of post- and pre-treatment angles.

$$\mathbb{E}(\hat{\tau}(t^*)) = \mu^{t \geq t^*} - \mu^{t < t^*}$$
$$\mathbb{V}(\hat{\tau}(t^*)) = \frac{(\kappa^{t \geq t^*})^{-1}}{n-m} + \frac{(\kappa^{t < t^*})^{-1}}{n}$$

Although the spherical example is thus somewhat trivial, we want to present a simulation study of the manifold RD estimator when both manifolds are spheres. We recycle the toy example from Section 2.5. Suppose we have a post-treatment mean projection angle, $\bar{\theta}^{t \geq t^*} = \pi$ and a pre-treatment mean projection angle, $\bar{\theta}^{t < t^*} = \pi/2$. We consider both of these manifolds to be circles, but the post-treatment manifold has radius 2. Thus, the manifold we have fit grows from a unit circle (radius = 1) to a post-treatment circle with radius 2. This is relevant to biomedical settings where we believe our object of interest, e.g., a tumor growth, grows equally in size after treatment. This means that while the pre-treatment Gaussian curvature stays 1, i.e. $\kappa^{t < t^*} = 1$, our post-treatment curvature is now $\kappa^{t \geq t^*} = \frac{1}{R^2} = \frac{1}{4}$. We simulate 500 draws and observe the following results:

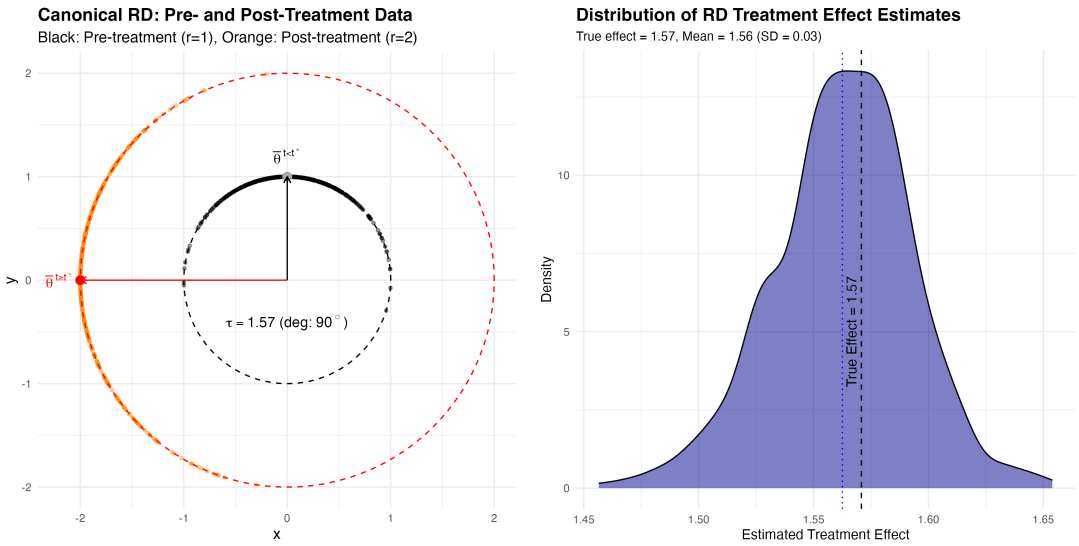

Figure 6: Distribution of Pre- and Post-Treatment Data, Treatment Effect, and Estimated Treatment Effect

We can see that our estimator captures the treatment effect well. We want to reiterate that in the spherical setting, the manifold RD estimator simplifies to the canonical estimator. That being said, it is reassuring to see well-behaved simulation-based results in this simplest of settings.

4. Discussion and Future Direction

When compared to well-established LLR methods, the mean projection angles are simply a spherical reparameterization of the outcome variable of interest. This formulation naturally preserves identification assumptions required for causal interpretation while accommodating the topological abstractions within the principal manifold framework. The magnitude of the posterior variance directly reflects how drastically treatment has altered the local geometry of our outcome space, e.g., if we have a large difference, we want to encode that topological uncertainty into our LATE.

An important point of emphasis is what this procedure is not. Namely, we are not simply updating our fitted manifold with the new data, and letting the posterior angle update iteratively with each new post-treatment observation. This is certainly an interesting angle in the context of statistical learning procedures (e.g., prediction tasks), but would complicate causal identification. Rather, we are using the pre-treatment manifold to borrow topological information for our post-treatment mean vector. In that sense, after having fit $\mathcal{M}_d^{t \geq t^*}$ to the post-treatment data only, we assign the prior mean (i.e. $\bar{\theta}^{t<t^*}$) a weight of zero, only using the post-treatment observations.

We round out this section with an appealing information-theoretic point of view. This manifold RD estimate is fundamentally analogous to the existing methods that penalize observations for being further away from treatment onset in time. The main difference is that this procedure accounts for the topological information of our embedding space. Existing methods are quick to discard this information and assume it away. Through our Bayesian topological weights, we now have a procedure for iteratively updating our beliefs in the underlying topological structure of our data. Each additional observation in the post-treatment period allows us to strengthen our beliefs in local topological changes.

Future work could address the use of priors for the construction of the mean projection vectors in the pre- and post-treatment periods, respectively. That is, within the existing manifold RD framework, posit hyperpriors within each treatment period. This connects neatly to the idea of encoding global manifold prior information. A promising avenue within the broader topological data analysis is mmanifold learning drawing from persistent homology. To the author's knowledge, there currently exists no method to obtain causal estimates in a manifold RD setup. With that in mind, this idea is rudimentary and can benefit from more rigorous analysis and robust theoretical guarantees.

5. Code and Data Availability Statement

Code necessary to reproduce the simulation study in this paper is available at the GitHub repository: https://github.com/posmikdc/directional-priors

References

Zach Branson, Maxime Rischard, Luke Bornn, and Luke W. Miratrix. A nonparametric bayesian methodology for regression discontinuity designs. *Journal of Statistical Planning and Inference*, 202:14–30, September 2019. ISSN 0378-3758. doi: 10.1016/j.jspi.2019.01.003. URL http://dx.doi.org/10.1016/j.jspi.2019.01.003.

Trevor Hastie and Werner Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406):502–516, 1989. ISSN 01621459, 1537274X. URL http://www.jstor.org/stable/2289936.

K. V. Mardia and S. A. M. El-Atoum. Bayesian inference for the von mises-fisher distribution. *Biometrika*, 63(1):203–206, 1976. ISSN 00063444. URL http://www.jstor.org/stable/2335106.

Marina Meilă and Hanyu Zhang. Manifold learning: what, how, and why, 2023. URL https://arxiv.org/abs/2311.03757.

Kun Meng and Ani Eloyan. Principal manifold estimation via model complexity selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(2):369–394, March 2021. ISSN 1467-9868. doi: 10.1111/rssb.12416. URL http://dx.doi.org/10.1111/rssb.12416.

Robert Tibshirani. Principal curves revisited. *Statistics and Computing*, 2(4):183–190, 1992. doi: 10.1007/BF01889678. URL https://doi.org/10.1007/BF01889678.