On Accuracy and Speed of Geodesic Regression: Do Geometric Priors Improve Learning on Small Datasets?

Adele Myers
UC Santa Barbara
adele@ucsb.edu

Nina Miolane
UC Santa Barbara
ninamiolane@ucsb.edu

Abstract

Image datasets in specialized fields of science, such as biomedicine, are typically smaller than traditional machine learning datasets. As such, they present a problem for training many models. To address this challenge, researchers often attempt to incorporate priors, i.e., external knowledge, to help the learning procedure. Geometric priors, for example, offer to restrict the learning process to the manifold to which the data belong. However, learning on manifolds is sometimes computationally intensive to the point of being prohibitive. Here, we ask a provocative question: is machine learning on manifolds really more accurate than its linear counterpart to the extent that it is worth sacrificing significant speedup in computation? We answer this question through an extensive theoretical and experimental study of one of the most common learning methods for manifold-valued data: geodesic regression.

1. Introduction

Collecting images in scientific domains like biomedicine requires significant investment in time, labor, and financial resources. Thus, such datasets are typically several orders of magnitude smaller than datasets used in traditional machine learning. One way to enhance learning from such small datasets is to utilize the geometry of the data manifold. For example, a growing amount of biomedical research seeks to relate biological shapes to their physiological functions and, in that context, it has been recognized that shapes constitute data points on manifolds [10] — nonlinear generalizations of vector spaces. Performing analyses on these manifolds can ensure that results are invariant to the positioning and orientation of the object, thus enhancing prediction accuracy related to objects' shapes.

However, learning on manifolds is sometimes computationally intensive to the point of being prohibitive. For example, fitting geodesic regression [15] — a generalization of linear regression to manifolds — on surface shapes can take several days even for small datasets of surfaces discretized as meshes with only a few hundred of vertices. This raises the question: is machine learning on manifolds really more accurate than its linear counterpart to the extent that it is worth sacrificing significant speedup in computation?

Contributions We answer this question by considering one of the most common generalizations of linear regression to manifolds: geodesic regression, and its traditional fitting procedure, called geodesic least squares (GLS), which computes the maximum likelihood estimator of a generative model with Gaussian noise constrained to the manifold of interest. We challenge several foundational and widely adopted assumptions. First, we challenge the assumption that noise is indeed Gaussian on the manifold. This is rarely true for a wide range of datasets, as measurement noise from real-world machines (MRIs, rulers, cameras, telescopes, lasers, sensors, etc.) is typically Gaussian in the Euclidean space that contains the data manifold, and *not* subject to the manifold's constraints. Second, we challenge the assumption that GLS provides the most accurate estimates for geodesic regression across several noise models, including the foundational manifold Gaussian noise.

Through theoretical results and experimental validation, we advocate for a paradigm shift in machine learning for manifold data. We recommend a departure from the traditional and computationally intensive fitting approaches such as GLS, and instead champion the adoption of simpler, more efficient, and equally accurate, fitting procedures. Additionally, we strongly advocate that any new method in the field of machine learning on manifolds should compare itself to the baseline of projecting the result of the linear method to the manifold. Our insights open the door to practical, yet reliable, analyses in the specialized fields of science typically characterized by smaller datasets, starting with biomedicine and its shape datasets.

Why is Noise Not Gaussian on the Manifold? We start our discussion by motivating noise models that depart from Gaussian noise on the manifold. To build intuition, let us first consider a toy example on a simple manifold, the

sphere on the left side of Fig. 1, representing the surface of the earth. Consider a practitioner who measures the position Y of an object on the earth. While this object exists in 3D ambient space, physics (gravity) constrains its position to the 2D sphere. The traditional model of geodesic regression describes noise as being added along geodesics of the manifold, represented as blue curves on the sphere in Fig. 1 (top left). This is a realistic model of noise if the practitioner measures the object's position directly on the sphere, for example using a piece of string stretched along the sphere. Only then would the noise experience the same constraints as the data, as (using r, ϕ, θ coordinates) the radius r would be fixed, and the only uncertainty in the measurement would be along ϕ and θ, which are the intrinsic coordinates of the sphere. However, most practitioners use tools that measure the position of the object of interest in the ambient space, i.e., in x, y, z coordinates in our toy example. Such instruments are free to measure position along three degrees of freedom, and therefore measurement uncertainty and noise will exist along all three degrees of freedom (see blue lines departing from the sphere in Fig. 1) (bottom left). Additionally, it is important to note that linear Gaussian noise projected onto the manifold is generally different than Gaussian noise directly generated on the manifold. This motivates the consideration of the Euclidean and projected noise models.

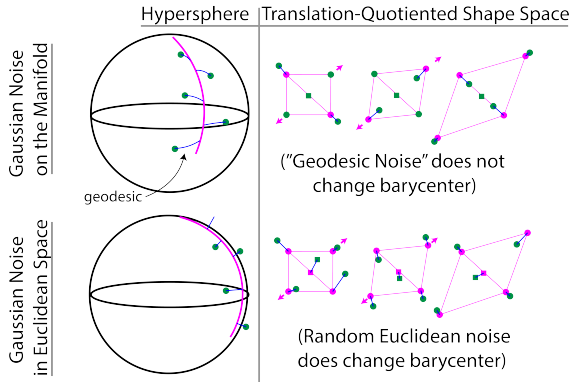

Figure 1. Common noise model (on manifold, top) versus realistic noise model (in ambient Euclidean space, bottom) in machine learning on manifolds, shown along a geodesic on the sphere (left) and on a shape space (right). We challenge the assumption that noise is Gaussian on the manifold (top), since it does not respect any manifold-defining constraints. We argue that it is more realistically modeled as Gaussian in ambient Euclidean space (bottom).

Second, consider a more complex example: shape data, and specifically brain surface shape data represented as surfaces segmented from 3D images acquired with magnetic resonance imaging (MRI). Without considering shape, we can consider each surface to be one point in a Euclidean "object space", which we will call the space of surfaces. However, to make this a "shape space" where distances between surfaces are invariant to the orientation and position of the surfaces, practitioners often equip this space with a Riemannian metric. The difference between Gaussian noise in Euclidean space and Gaussian noise on shape data is shown in Fig. 1 (right). Gaussian noise added in this curved shape space would have to not affect the orientation or position of the barycenter of the surface. However this is almost never the case, as the noise on the brain surface comes from two sources: first, the noise coming from the MRI acquisition, and second, the error coming from the segmentation of the brain surface from the 3D image, both of which are Euclidean and random in nature. The noise on each vertex of the brain surface is added randomly in three (x, y, z) directions, and is typically modelled as multivariate Gaussian in Euclidean space, before being deformed on the manifold equipped with the Riemannian metric. This motivates the consideration of the deformed noise model.

2. Related Works

Table 1. Noise Models in Manifold Regressions. Gaussian noise on manifolds is one of the most popular. Alternative models have been considered, esp. for non-geodesic regressions. However, works do not benchmark their regression approach against different noise models. Here, we compare how learning procedures for geodesic regression compare across noise models.

	Paper	Noise on Manifold		Noise in $\mathbb{R}^d$
		Gaus.	Else	Gaus.
Geodesic	This Paper	✓	✓	✓
	Fletcher [2, 15]	✓		
	Myers [11]			✓
Other Manifold Regressions	Hinkle [6]	✓		
	Hanik [4]	✓		
	Lin [9]	✓		
	Schotz [13]		✓	
	Petersen [12]		✓	
	Kuhnel [8]			✓
	Tsagkrasoulis [17]			✓
	Davis [1]			✓
	Shi [14]	No Discussion of Noise		

We review regression methods on manifolds and characterize existing works in three categories shown in Tab. 1: models with noise 1) Gaussian on the manifold 2) non-Gaussian on the manifold, 3) Gaussian in Euclidean Space $\mathbb{R}^d$. The categories of noise are chosen based on the experiments presented in the corresponding papers. We only review non-Bayesian regression methods where the input space is Euclidean (either one-dimensional $\mathbb{R}$ or higher dimensional $\mathbb{R}^d$) and the output space is a manifold.

Noise on the Manifold. Works that model noise as Gaussian on the manifold (first column in Tab. 1) include geodesic regression [2, 15], polynomial regression on manifolds [7], local extrinsic regression [9], and Bezier splines on manifolds [4]. These are all regression methods with one-dimensional input space and manifold output space.

Works that model noise on the manifold, but not necessarily as Gaussian include the local geodesic regression by Schotz et al. [13], who use the contracted uniform distribution which is obtained from the uniform distribution on the sphere, and the Fréchet regression by Petersen and Muller [12], who add Gaussian noise in the Euclidean space of probability densitys but then deform this noise either through application of the Wasserstein metric or through a Wasserstein transportation.

Noise in Ambient Euclidean Space. Works that model noise in the ambient Euclidean space first include [11], who propose to fit geodesic regression using linear residuals instead of the geodesic residuals in [2, 15]. In the non-geodesic regression class, we find the stochastic development regression [8], random forest [17], and kernel regression [1] that model noise in ambient space.

Impact of Noise is Understudied. However, no work considers the effect of projecting or deforming Euclidean noise to the manifold — two noise models that are, as we argue in introduction, realistic for real-world data. Accordingly, the statistical theory of these noise models, such as maximum likelihood estimation, is never discussed. Our paper fills this first gap. Second, we observe that works do not benchmark their regression approach against different noise models. They usually choose one and perform all of their experiments with it, even if it might not be a realistic noise model for the data. Our paper evaluates the accuracy of geodesic regression fitting approaches across several noise models, thus filling this second gap. Lastly, while manifold regressions are meant to generalize and improve upon regressions in Euclidean space, very few works compare their method against their Euclidean counterpart, obtained by fitting a regression method in Euclidean space and projecting its result to the manifold. Yet, such comparison is crucial to show whether considering the manifold's geometry actually provides a significant advantage. It would show that the added complexity of manifold computations is actually met with added accuracy.

3. Background: Fitting Geodesics

We introduce concepts of differential geometry necessary for geodesic regression. Details can be found at [3].

Manifolds. A *Riemannian metric* G on a smooth manifold $\mathcal{M}$ is a family $(G_p)_{p\in\mathcal{M}}$ of inner products defined on each tangent space $T_p\mathcal{M}$ that varies smoothly with p. A pair $(\mathcal{M}, G_p)$ is called a *Riemannian manifold*.

A *geodesic* is a curve $\gamma : [0,1] \to \mathcal{M}$ which minimizes the energy functional $E(\gamma) = \frac{1}{2}\int_0^1 G(\dot\gamma_t, \dot\gamma_t)_{\gamma_t} dt$, where $\dot\gamma_t \in T_{\gamma(t)}\mathcal{M}$ is the velocity vector at point γ_t. Curves minimizing the energy E also minimize the length of the geodesic, defined as $L_\gamma = \int_0^1 \sqrt{G(\dot\gamma_t, \dot\gamma_t)_{\gamma_t}} dt$. It follows that geodesics are locally distance-minimizing paths on the manifold $\mathcal{M}$, and the Riemannian metric G_p provides a notion of geodesic distance on $\mathcal{M}$. Intuitively, just as a straight line minimizes the distance between two points in Euclidean space, a geodesic minimizes the distance between two points on a manifold.

Geodesics are computed as the solutions of the *geodesic equation*, which is an ordinary differential equation (ODE) written in local coordinates as:

$$\ddot\gamma^k(t) + \Gamma^k_{ij}\dot\gamma^i(t)\dot\gamma^j(t) = 0, \tag{1}$$

for all times $t \in [0,1]$ where Γ^k_{ij} denote the Christoffel symbols, which are computed from the Riemannian metric. Some manifolds, such as hyperspheres and hyperbolic spaces, enjoy analytical expressions for their geodesics, i.e., we know closed-form solutions of Eq. (1). For example, it is well known that the geodesics of the sphere are the great circles. However, this is not the case for all manifolds. For example, we do not know closed-form solutions for the geodesic equation on the manifold of surface shapes, mentioned in introduction and described in more details in the supplementary materials. Therefore, we must calculate (approximate) geodesics between surface shapes by numerically integrating the ODE from Eq. (1).

Operations on Manifolds. Let p be a point on $\mathcal{M}$ and $v \in T_p\mathcal{M}$. The *exponential map* (Exp) is defined as the map $(p, v) \mapsto \gamma_{t=1}$. For manifolds with closed form geodesic solutions, computing the point $\gamma_{t=1}$ is as simple as plugging $t = 1$ into the geodesic solutions. For manifolds with no closed form geodesic solution, we compute $\gamma_{t=1}$ by integrating the geodesic equation over t from $t : 0 \to 1$ with initial conditions $(\gamma_0 = p, \dot\gamma_0 = v)$. The inverse of Exp on its injectivity domain is called the *logarithm map* (Log).

3.1. Geodesic Regression: Geodesic Least Squares

Geodesic regression (GR) [2, 16] models the relationship between an independent variable $X \in \mathbb{R}$ and the dependent variable Y, which is a random variable with values on a manifold $\mathcal{M}$, as:

$$Y = \mathrm{Exp}(\mathrm{Exp}(p, Xv), \epsilon), \tag{2}$$

where Exp is the exponential map defined in the previous section, $\tilde Y = \mathrm{Exp}(p, Xv)$ is a point on the noiseless geodesic, Xv is the tangent vector at $\tilde Y$, and ϵ is isotropic Gaussian noise in the tangent space of $\tilde Y$ with variance σ^2 and mean 0. When the manifold of interest is $\mathcal{M} = \mathbb{R}^d$ (linear), the Exp operator simplifies to addition: $\mathrm{Exp}(p, v) = p + v$. Consequently, this generative model simplifies to

the linear regression generative model, and we can think of p and v as the geodesic regression "intercept" and "slope". We note that the Exp operation appears twice: to model the geodesic itself, and to model the noise ϵ. Fletcher et al. use a Geodesic Least Squares (GLS) estimator, detailed later in Eq. (5), to learn the geodesic that best fits the data, i.e., to estimate the true intercept and slope $(p, v) \in \mathcal{M} \times T_p\mathcal{M}$ by learning $(\hat{p}, \hat{v})$.

Computational Costs. For some manifolds, the geodesic least squares (GLS) problem of Eq. (5) above does not have an analytical solution. For example, the manifold of surface shapes, mentioned in introduction and detailed in the supplementary materials, does not. For such manifolds, we need to compute the estimates of the intercept and slope with (Riemannian) gradient descent [2]. The expression of the Riemannian gradient itself might be unknown in closed form and must be computed numerically. In this case, traditional geodesic regression becomes computationally intensive. Thus, fitting a geodesic via less computationally intensive approaches, as proposed in [11] and detailed in the next subsection, become attractive alternatives.

3.2. Geodesic Regression: Linear Least Squares

Fitting geodesic regression with a Linear Least Squares (LLS) estimator can offer significant speedups from traditional geodesic regression. [11] propose a geodesic generative model with Euclidean noise, given by

$$Y = \text{Exp}(p, Xv) + \epsilon, \tag{3}$$

using the notations of Eq. (2), with the noise ϵ modelled as isotropic Gaussian in ambient Euclidean space. [11] propose to learn $(\hat{p}, \hat{v})$ via Linear Least Squares (LLS) detailed later in Corollary 1.

Computational Costs. For some manifolds, the linear least squares (LLS) problem of Corollary 1 does not have an analytical solution. Thus, LLS requires gradient descent, just as GLS did, yet with a less computationally expensive gradient, as $\epsilon_i = \text{Log}\left(\hat{Y}_i, Y_i\right)$ is replaced by $\epsilon_i = Y_i - \hat{Y}_i$. The replacement of a costly Log computation with a simple subtraction operation offers significant speedup, especially for the space of surface shapes where the closed form Log solution is not known. However, we note that the computational cost of the Exp map defining the geodesic remains.

3.3. Projecting the Results of Linear Regression

Next, we consider the approach which projects the traditional linear regression estimate to the manifold, adapted from [9] but for geodesic regression. We call it projected Linear Least Squares (pLLS). Specifically, the intercept estimate $\hat{p}$ is projected to the manifold, and the slope estimate $\hat{v}$ is projected to the tangent space of the manifold at $\hat{p}$.

Computational Costs. Fitting linear regression with linear least squares enjoys closed form solutions for the intercept and slope estimates, given by the so-called normal equations. Thus, even for manifolds with no closed form geodesic solutions, this method requires no gradient descent operations, as all distances are calculated in Euclidean space. Most of the computational cost of this approach, as an estimation of the geodesic, comes from the projection operations, which happen only once.

4. Impact of Noise Models on Geodesic Fits

Here, we address the gap in the literature by studying how different geodesic regression approaches compare under different noise models. This section specifically provides theoretical results to answer the question: is GLS really more accurate than faster alternatives (i.e. [9, 11]) to the extent that it is worth sacrificing significant speedup in computation? To this aim, we leverage four different noise models for manifold-valued data, including two new noise models that — as we argue in introduction — are more realistic than the manifold Gaussian noise model. This yields four geodesic regression generative models, shown in the four columns of Fig. 2. Fig. 2 also defines and illustrates notations Y, $\tilde{Y}$, and $\hat{Y}$ used in this section. We will discuss how each estimator (or fitting procedure), shown in the rows of Fig. 2, performs under each generative model.

4.1. Manifold Gaussian Noise

The generative model proposed by [2, 15] assumes that the noise is Gaussian on the manifold. We have recalled this model in Eq. (2), and we illustrate it in Fig. 2 (first column).

Probability Density. This model yields the following probability density of the data [2, 15]:

$$p(Y|X;\, p, v) = \frac{1}{C_{\mathcal{M}}(\sigma)} \exp\left(-\frac{d\left(Y, \tilde{Y}\right)^2}{2\sigma^2}\right), \tag{4}$$

where $C_{\mathcal{M}}(\sigma)$ is the normalization constant, and σ^2 is the variance of the Gaussian noise.

Maximum Likelihood Estimator. The maximum likelihood (ML) estimator is the geodesic least-squares (GLS) estimator [15]:

$$(\hat{p}, \hat{v})^{\text{ML}} = \operatorname*{amin}_{\hat{p},\hat{v}} \sum_{i=1}^{n} d\left(Y_i, \hat{Y}_i\right)^2 \tag{5}$$

$$= \operatorname*{amin}_{\hat{p},\hat{v}} \sum_{i=1}^{n} \|\text{Log}(\hat{Y}_i, Y_i)\|^2_{\hat{Y}_i}, \tag{6}$$

where d denotes the geodesic distance on the manifold, which can be equivalently expressed as the norm (computed with the Riemannian metric) of the Log of the data points Y at the predicted points $\hat{Y}$.

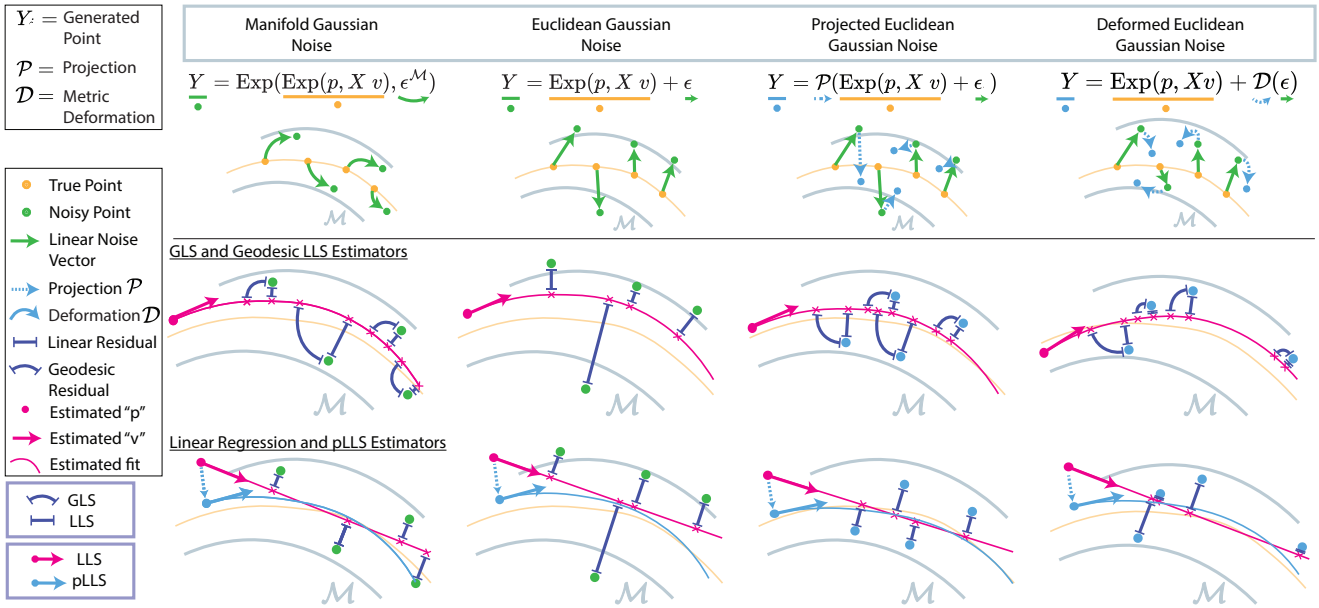

Figure 2. Theoretical Analyses. The four columns represent the models of noise. The four rows show the estimators: GLS, LLS for Geodesic Regression, LLS for Linear Regression, and LLS for Linear Regression projected to the manifold. **Notations:** we denote $\tilde{Y} = \mathrm{Exp}(p, Xv)$ the point on the geodesic before adding noise (orange), Y the final observed data point after Gaussian noise has been added on the manifold or in Euclidean space (green arrows and points) and (if applicable) projection $\mathcal{P}$ or deformation $\mathcal{D}$ have been applied (blue arrows and points). Finally, $\hat{Y} = \mathrm{Exp}(\hat{p}, X\hat{v})$ is the predicted point on the geodesic (pink).

4.2. Euclidean Gaussian Noise

As discussed in introduction, noise added by real-world measurements is often Euclidean in nature. The generative model proposed by [11] assumes that the noise is Gaussian in ambient Euclidean space. We have recalled this model in Eq. (3), and we illustrate it in Fig. 2 (second column). In this model, the data point Y is observed in the ambient Euclidean space, and not on the manifold. We note that [11] do not provide the theory associated with this model, which we contribute here.

Probability Density. We provide the probability density associated with this model in the lemma below.

Lemma 1. *The probability density associated with the generative model with Euclidean Gaussian noise is:*

$$p(Y \mid X;\, p, v) = \frac{1}{C(\sigma)} \exp\left(-\frac{\|Y - \tilde{Y}\|^2}{2\sigma^2}\right),$$

where $C(\sigma) = \sqrt{(2\pi)^d \sigma^{2d}}$ *is the normalization constant.*

The proof is immediate by definition of the Gaussian distribution. This density differs from that of the classical linear regression, since $\tilde{Y} = \mathrm{Exp}(p, Xv)$ is on a geodesic.

Maximum Likelihood Estimator. We provide the maximum likelihood estimator associated with this model in the lemma below. The proof is immediate given Lemma 1.

Corollary 1. *The maximum likelihood estimator associated with the generative model with Euclidean Gaussian noise is the linear least-squares estimator (LLS):*

$$(\hat{p}, \hat{v})^{ML} = \underset{\hat{p},\hat{v}}{\mathrm{amin}} \sum_{i=1}^{n} \|Y_i - \hat{Y}_i\|^2 \text{ for } \hat{Y}_i = \mathrm{Exp}(\hat{p}, X_i\hat{v}),$$

where the Euclidean distance $\|.\|$ *compares* $\hat{Y}$ *and* Y.

4.3. Projected Euclidean Gaussian Noise

Now, we consider a new generative model that models noise on the manifold as the result of orthogonally projecting Gaussian noise in ambient Euclidean space to the manifold. The typical examples for this case are data on the sphere S^2, that is isometrically embedded in the Euclidean space $\mathbb{R}^3$, and the hyperboloid H^2 that is (non-isometrically) embedded in the vector space $\mathbb{R}^3$. Despite being a realistic noise model, as argued in introduction, the theory behind this framework has not been investigated, as we propose to do here.

Generative Model. We propose the following generative model, shown in Fig. 2 (third column):

$$Y = \mathcal{P}(\tilde{Y} + \epsilon) \quad \text{with } \tilde{Y} = \mathrm{Exp}(p, Xv), \tag{7}$$

using notations of Eq. (2), with ϵ isotropic Gaussian noise in ambient Euclidean space $\mathbb{R}^d$, of distribution $\mathcal{N}(0, \sigma^2)$, and

$\mathcal{P}$ an orthogonal projection to the manifold. This model explicitly acknowledges that, even though the noise ϵ pushes the data outside of the manifold, processing steps usually project noisy data onto the manifold. For example, if a practitioner studies positions of cities on the earth (modelled as a sphere of radius 1), and realizes that noisy measurements give a position x that is slightly outside the sphere, their first step would be to normalize $\frac{x}{\|x\|}$, hence projecting the data on the manifold. Often, most theory ignores the impact of this processing step. Here, we explicitly include it through the orthogonal projection $\mathcal{P}$.

Probability Density. We provide the probability density for this model. The proof is in supplementary materials.

Proposition 1. *The probability density associated with the model with projected Euclidean noise of Eq. (8) is:*

$$p(Y \mid X;\ p, v) = \frac{1}{\sqrt{2\pi\sigma^2}^m} \exp\left(-\frac{\|P_Y^{\perp}(\tilde{Y}) - \tilde{Y}\|^2}{2\sigma^2}\right),$$

where we introduce the projection $P_Y^{\perp}$, different from $\mathcal{P}$, that projects the noiseless data point $\tilde{Y}$ onto the subspace $T_Y\mathcal{M}^{\perp}$, as shown in Fig. 3, and m is the intrinsic dimension of $\mathcal{M}$.

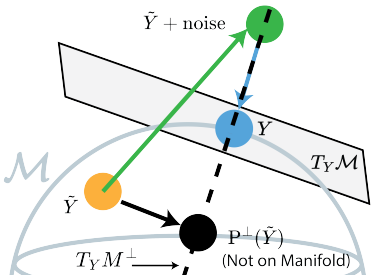

Figure 3. Notations for the Projected Euclidean Noise. $\tilde{Y} = \mathrm{Exp}(p, Xv)$: noise-less data point (orange), Y: noisy data point (blue) after projection via $\mathcal{P}$ (blue arrow) on $\mathcal{M}$. We decompose the ambient space $\mathbb{R}^d$ into $T_Y\mathcal{M}$ (gray plane) and its orthogonal $T_Y\mathcal{M}^{\perp}$ (dashed line). The orthogonal projection to $T_Y\mathcal{M}^{\perp}$ is denoted $P_Y^{\perp}$ (black arrow).

Maximum Likelihood Estimator. We provide the maximum likelihood estimator for this model in the corollary below. The proof is immediate given Proposition 1.

Corollary 2. *The maximum likelihood estimator for the generative model with projected Euclidean Gaussian noise is:*

$$(\hat{p}, \hat{v})^{ML} = \operatorname*{amin}_{\hat{p},\hat{v}} \sum_{i=1}^{n} \|P_{Y_i}^{\perp}(\hat{Y}_i) - \hat{Y}_i\|^2,$$

for $\hat{Y}_i = \mathrm{Exp}(\hat{p}, X_i\hat{v})$, where $P_{Y_i}^{\perp}$ is the projection onto the subspace $T_{Y_i}\mathcal{M}^{\perp}$, as shown in Fig. 3. The estimator depends on the data point Y through this projection. We call this estimator projected least squares (PLS).

4.4. Deformed Euclidean Gaussian Noise

Next, we consider a new generative model that models noise on the manifold by "deforming" a Gaussian noise, in the sense described below. In this setting, we assume that the Riemannian manifold $\mathcal{M}$ is a subset of a Euclidean space of same dimension. Therefore, there are two Riemannian metrics that we can consider on $\mathcal{M}$: the Euclidean metric or its own Riemannian metric. The typical examples for this case are the Poincaré disk model, as a subset of the plane $\mathbb{R}^2$, and the manifold of surface shapes with elastic metric, as a subset of Euclidean space of immersions [5]. We model the case where the noise is Gaussian for the Euclidean metric, but not Gaussian (instead, "deformed") for the Riemannian metric. Despite being a realistic noise model, as argued in introduction, the theory behind this framework has not been studied, as we propose to do here.

Generative Model. We propose the following generative model, shown in Fig. 2 (third column):

$$Y = \tilde{Y} + \mathcal{D}(\epsilon) \quad \text{with } \tilde{Y} = \mathrm{Exp}(p, Xv), \tag{8}$$

using notations of Eq. (2), with ϵ isotropic Gaussian noise of variance σ^2 for the Euclidean metric, and $\mathcal{D}$ indicates that it will be deformed since we consider the Riemannian metric on $\mathcal{M}$.

Probability Density. We provide the probability density for this model in the proposition below.

Proposition 2. *The probability density for the generative model with deformed Euclidean Gaussian noise is:*

$$p(Y \mid X;\ p, v) = \frac{1}{C(\sigma)\sqrt{\det G(Y)}} \exp\left(-\frac{\|Y - \tilde{Y}\|^2}{2\sigma^2}\right),$$

with $C(\sigma) = \sqrt{(2\pi)^D\sigma^{2D}}$ the normalization constant, and G the matrix of the Riemannian metric of $\mathcal{M}$ at Y.

This probability density differs from that of the classical linear regression, through the term $\sqrt{\det G(Y)}$. The proof is given in the supplementary materials.

Maximum Likelihood Estimator. We provide the maximum likelihood estimator associated with this model.

Corollary 3. *The maximum likelihood estimator for the model with deformed Euclidean Gaussian noise is:*

$$(\hat{p}, \hat{v})^{ML} = \operatorname*{amin}_{\hat{p},\hat{v}} \sum_{i=1}^{n} \|Y_i - \hat{Y}_i\|^2 \ \textit{for}\ \hat{Y}_i = \mathrm{Exp}(\hat{p}, X_i\hat{v}).$$

The proof is immediate by noting that the term $\sqrt{\det G(Y)}$ is independent of the parameters p, v. Notably, the maximum likelihood estimator for deformed Gaussian noise is the same as for (undeformed) Gaussian noise.

4.5. Comparing estimators

We discuss how the maximum likelihood (ML) estimators for the different models differ, i.e., how GLS, LLS and PLS from this section compare. This amounts to comparing the terms that appear in their respective definitions $d^2(Y, \hat{Y})$, $\|Y - \hat{Y}\|^2$ and $\|P_{\hat{Y}}^{\perp}(\hat{Y}) - \hat{Y}\|^2$.

First, LLS can be used as an approximation to GLS, since the Euclidean distance is an approximation of the geodesic distance d, especially for nearby points. Here, Y and $\hat{Y}$ are indeed close, if the noise variance σ^2 is small at the scale of the curvature of the manifold. Quantitatively:

$$d(\hat{Y}, Y)^2 = \|\hat{Y} - Y\|^2 + \delta_{\mathcal{M}}, \tag{9}$$

where $\delta_{\mathcal{M}}$ denotes the difference between the squared Riemannian distance on the manifold and the squared Euclidean distance in $\mathbb{R}^d$. For a manifold $\mathcal{M}$ isometrically immersed in $\mathbb{R}^d$, it depends on the external curvature of $\mathcal{M}$.

Second, PLS can approximate LLS (and thus, GLS), if the noise variance σ^2 is small, and if the manifold has low curvature in ambient Euclidean space. Indeed, the point $P_{\hat{Y}}^{\perp}(\hat{Y})$ is close to Y when the manifold has small curvature, has shown in Fig. 3. Quantitatively, we have:

$$\|P^{\perp}(\hat{Y}) - \hat{Y}\|^2 = \|\hat{Y} - Y\|^2 - \|P^{\perp}(\hat{Y}) - Y\|^2, \tag{10}$$

by Pythagorean theorem in the ambient Euclidean space $\mathbb{R}^d$. The difference between PLS and LLS only comes from the second term in Eq. (10), typically small.

5. Experiments

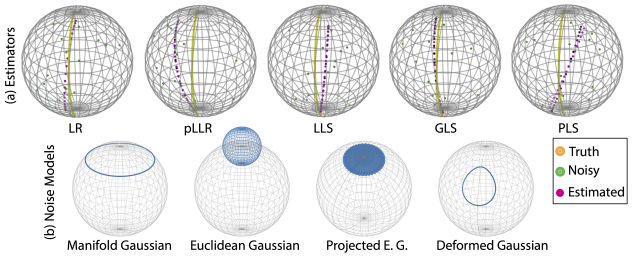

Figure 4. Comparison of (a) estimators and (b) noise models.

Here, we complement our theoretical analyses by studying how different geodesic regression approaches compare under different noise models through extensive experiments. This section specifically provides practical, numerical results to answer the question: is GLS really more accurate than faster alternatives to the extent that it is worth sacrificing significant speedup in computation?

We analyze how the estimators GLS, LLS, PLS, and pLLS behave under different, realistic noise models (see Fig. 4). Such an analysis is rarely performed in the literature, where authors often focus on one noise model, usually Gaussian. This will give us unique insights on whether practitioners should use GLS, LLS, PLS, or pLLS, depending on which noise model is believed to best approximate their actual data. We initialize all regression computations with the result of linear regression in ambient space $\mathbb{R}^d$.

Datasets. For each manifold $\mathcal{M}$ considered (see below), we generate a synthetic geodesic between two points on $\mathcal{M}$. Next, we extract n data points along this geodesic, testing $n \in \{5, 10, 20, 30\}$. Then, we add noise to each data point, according to the noise model of interest. The Gaussian noises have standard deviations $\sigma \in \{0, 0.01, 0.1, 0.2, 0.4, 0.6\}$.

We test the generative model with manifold Gaussian noise, Euclidean Gaussian noise, and projected Euclidean Gaussian noise on the hypersphere and hyperboloid for dimensions $m \in \{2, 3, 5, 10\}$. The hypersphere and the hyperboloid are canonical examples of manifolds with curvatures 1 and -1 respectively. On these manifolds, geodesics have well-known closed forms. Testing on these classical manifolds thus allows us a chance to investigate whether LLS and pLLS have advantages over GLS for manifolds that do not require numerical approximations of geodesics.

We test the generative model with deformed Euclidean Gaussian noise on surface shapes. For the geodesic, we choose two simple surface shapes as our start and end points: an (undeformed) cube and a cube that has been twisted and stretched. We discretize the surfaces with 8 vertices. We compare LLS, pLLS and GLS for this noise.

Evaluation Metrics. We measure the *accuracy* of each estimator by computing the root mean square deviation (RMSD) between data points and predictions. For data on the hyperboloid and hypersphere, we compute RMSD as $\text{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d(Y, \hat{Y})^2}$, where d is the geodesic distance. For surface shapes, we compute a normalized RMSD (NRMSD), where the RMSD given above is normalized by the diameter of the mesh and the number of mesh vertices. Each prediction from the linear regression is projected to the manifold so that the geodesic distance d can be used. We measure the *speed* of each estimator as the duration of the fitting procedure, after initialization with the result of the projected linear regression.

Results. Remarkably, regardless of noise model, our experiments show that the RMSD does not vary significantly between the estimators, while the projected linear regression result (pLLS) and linear regression itself (with predictions projected to the manifold) take far less time.

Comparison: Manifold and Euclidean Gaussian Noises. As shown in Fig. 5, even for the manifold Gaussian noise, all four estimators perform similarly to GLS. Interestingly, we do observe a slight increase in accuracy for GLS and LLS in the case of the hyperboloid only. Yet, linear regression and its projection pLLS are faster, especially for higher dimensional hyperboloids. For example, for the 10 dimensional hyperboloid, LLS and GLS take 44 and 23

seconds, respectively, and LR and pLLS take 11 seconds.

Though we cannot test GLS on the Euclidean Gaussian noise model, we see in Fig. 6 that all estimators achieve similar RMSDs in this case, while linear regression does so in much less time than the geodesic regression model. Compared to evaluation under manifold Gaussian noise, estimators exhibit lower accuracy, esp. for the hyperboloid.

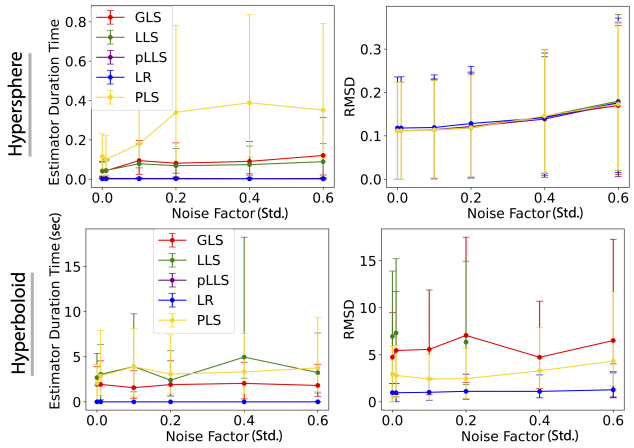

Figure 5. Comparing estimators for manifold Gaussian noise on the hypersphere (top), and hyperboloid (bottom) according to speed (left) and accuracy (right).

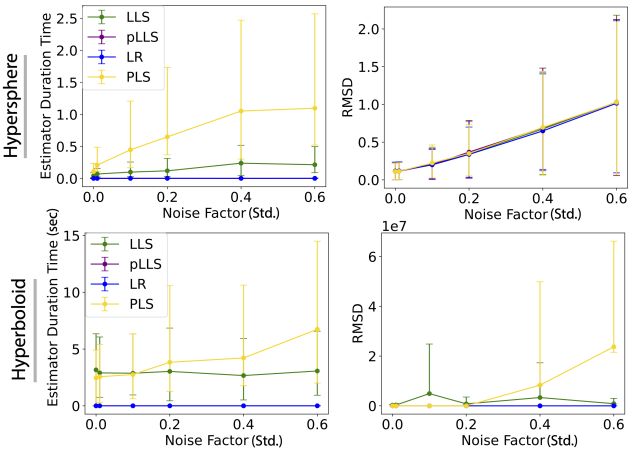

Figure 6. Comparing estimators for Euclidean Gaussian noise on the hypersphere (top), and hyperboloid (bottom) according to speed (left) and accuracy (right). Notice that here, we are forced to measure linear RMSD, as Y does not fall on the manifold.

Comparison: Projected Euclidean Gaussian Noise. As seen in Fig. 7, all estimators perform similarly well, with a better accuracy compared to the evaluations under the Euclidean Gaussian noise model. Again, unsurprisingly, the linear regression estimators take far less time.

Comparison: Deformed Euclidean Gaussian Noise. Here, on surface shape data, we see in Fig. 8 that linear

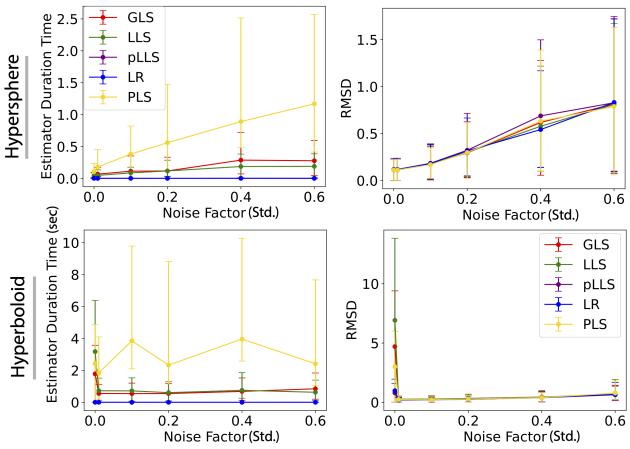

Figure 7. Comparing estimators for projected Euclidean Gaussian Noise on the hypersphere (top), and hyperboloid (bottom) according to speed (left) and accuracy (right).

techniques shine even brighter than for the hypersphere and hyperboloid. Indeed, this manifold does not have a closed form solution for geodesics. We see again that RMSD is not affected by choice of estimator, but GLS is thousands of times slower than its linear counterparts.

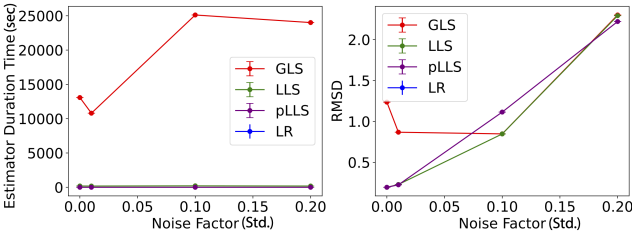

Figure 8. Comparing estimators for Deformed Euclidean Gaussian noise on surface shapes w.r.t. speed (left) and accuracy (right).

6. Conclusion

We asked the provocative question: do manifold regressions really provide estimates that are more accurate than their linear counterparts to the extent that it is worth sacrificing significant speedup in computation? We provided an analysis of the most common manifold regression: geodesic regression. We showed that the computational cost of geodesic regression does not actually make the result more accurate. In fact, its linear counterparts are as accurate and significantly faster. We hope that this analysis can spark a discussion on when geometric priors help, compared to cases where simpler alternatives bring faster results without sacrificing accuracy.

Acknowledgements A.M. and N.M. were funded by the NSF GRFP fellowship and NSF CAREER 2240158.

References

[1] Brad C Davis, P Thomas Fletcher, Elizabeth Bullitt, and Sarang Joshi. Population shape regression from random design data. *International journal of computer vision*, 90:255–266, 2010. 2, 3

[2] Thomas Fletcher. Geodesic regression on riemannian manifolds. In *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pages 75–86, 2011. 2, 3, 4

[3] Nicolas Guigui, Nina Miolane, Xavier Pennec, et al. Introduction to riemannian geometry and geometric statistics: from basic theory to implementation with geomstats. *Foundations and Trends® in Machine Learning*, 16(3):329–493, 2023. 3

[4] Martin Hanik, Hans-Christian Hege, Anja Hennemuth, and Christoph von Tycowicz. Nonlinear regression on manifolds for shape analysis using intrinsic bézier splines. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 617–626. Springer, 2020. 2, 3

[5] Emmanuel Hartman, Yashil Sukurdeep, Eric Klassen, Nicolas Charon, and Martin Bauer. Elastic shape analysis of surfaces with second-order sobolev metrics: a comprehensive numerical framework. *International Journal of Computer Vision*, 131(5):1183–1209, 2023. 6

[6] Jacob Hinkle, P Thomas Fletcher, and Sarang Joshi. Intrinsic polynomials for regression on riemannian manifolds. *Journal of Mathematical Imaging and Vision*, 50(1):32–52, 2014. 2

[7] Jacob Hinkle, Prasanna Muralidharan, P Thomas Fletcher, and Sarang Joshi. Polynomial regression on riemannian manifolds. In *European conference on computer vision*, pages 1–14. Springer, 2012. 3

[8] Line Kühnel and Stefan Sommer. Stochastic development regression on non-linear manifolds. In *Information Processing in Medical Imaging*, 2017. 2, 3

[9] Lizhen Lin, Brian St. Thomas, Hongtu Zhu, and David B. Dunson. Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association*, 112:1261 – 1273, 2015. 2, 3, 4

[10] K V Mardia. Directional statistics and shape analysis. *Journal of Applied Statistics*, 26(8):949–957, 1999. 1

[11] Adele Myers, Caitlin Taylor, Emily Jacobs, and Nina Miolane. Geodesic regression characterizes 3d shape changes in the female brain during menstruation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2542–2551, 2023. 2, 3, 4, 5

[12] Alexander Petersen and Hans-Georg Muller. Fréchet regression for random objects with euclidean predictors. *The Annals of Statistics*, 2016. 2, 3

[13] Christof Schötz. Nonparametric regression in nonstandard spaces. *Electronic Journal of Statistics*, 16(2):4679–4741, 2022. 2, 3

[14] Xiaoyan Shi, Martin Styner, Jeffrey Lieberman, Joseph G Ibrahim, Weili Lin, and Hongtu Zhu. Intrinsic regression models for manifold-valued data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 192–199. Springer, 2009. 2

[15] P Thomas Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International journal of computer vision*, 105:171–185, 2013. 1, 2, 3, 4

[16] P Thomas Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International journal of computer vision*, 105:171–185, 2013. 3

[17] Dimosthenis Tsagkrasoulis and Giovanni Montana. Random forest regression for manifold-valued responses. *Pattern Recognition Letters*, 101:6–13, Jan. 2018. 2, 3