Probabilistic Riemannian Submanifold Learning with Wrapped Gaussian Process Latent Variable Models: Supplementary Material

1 Details on Manifolds Used

The n-sphere S^n is a Riemannian manifold with exponential and logarithmic maps given by

$$\begin{aligned} \mathrm{Exp}_p(v) &= \cos(\|v\|_2)p + \sin(\|v\|_2)\frac{v}{\|v\|_2}, \\ \mathrm{Log}_p(q) &= \arccos\left(\langle p, q\rangle\right)\frac{q - \langle p, q\rangle p}{\|q - \langle p, q\rangle p\|_2}, \end{aligned} \tag{1}$$

where $\|\cdot\|_2$ is the 2-norm induced by the standard Euclidean innerproduct $\langle\cdot,\cdot\rangle$.

Kendall's shape space forms a quontient manifold of the sphere, so the operations defined for S^n apply, when working with the right quotient representatives. Kendall's shape space has the additional constraint of representing shapes with respect to an optimal translation between a pair of shapes. Let X, Y be the $2 \times N$ data matrices of two shapes, where N is the amount of landmarks, and each column represents the x, y-coordinates after quontienting away scale and translation. Then, the *Procrustean* distance between the shapes X, Y is given by

$$\min_R \|X - RY\|_2, \tag{2}$$

where R is a rotation matrix. The shapes are aligned by choosing a reference point, and aligning the population elements by minimizing the Procrustean distance.

The space $SPD(n)$ of symmetric, positive definite matrices can be given the structure of a Riemannian manifold, by endowing it with the *Log-Euclidean* metric. The tangent space at each point is the space of n-by-n symmetric matrices, and the affine-invariant metric is given by

$$g_P(V, U) = \mathrm{Trace}[V^T U], \tag{3}$$

and the exponential and logarithmic maps are given by

$$\mathrm{Exp}_P(A) = \exp(\log(p) + v), \quad \mathrm{Log}_P(Q) = \log(Q) - \log(P), \tag{4}$$

where exp stands for the matrix exponential and log for the matrix logarithm.

2 Latent Space Visualization

Here we provide the latent space visualizations for the diffusion-tensor and diatom datasets.

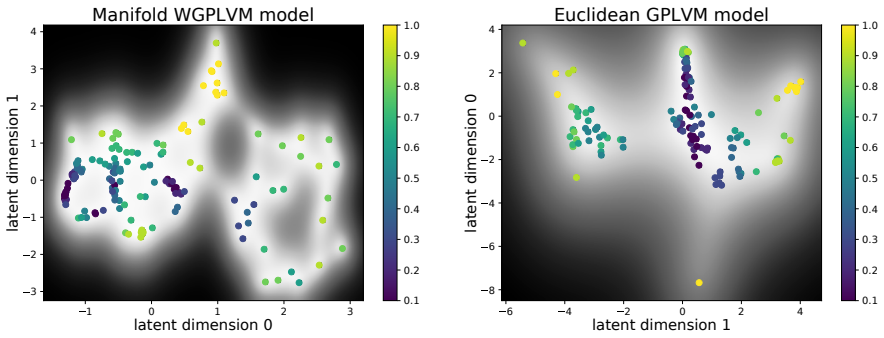

Figure 1: The latent spaces for the diffusion-tensor dataset learned using the WGPLVM and GPLVM models. The colors indicate the FA of the given tensor.

The *fractional anisotropy* (FA) of a 3×3 SPD matrix is a shape descriptor taking values between 0 and 1, where an FA of 0 corresponds to a round tensor, and an FA near 0 corresponds to a very thin one. Given the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ for an SPD matrix, its FA is defined as

$$\sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \hat{\lambda})^2 + (\lambda_2 - \hat{\lambda})^2 + (\lambda_3 - \hat{\lambda})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}},$$

where $\hat{\lambda}$ is the mean of the eigenvalues. In the latent space shown in Fig. 1, the latent variables are colored according to the FA of their associated tensor, and we see that both models provide a smooth transition between different FA values.

The latent space visualization of the diatom dataset is found in Fig. 2; here the latent variables are colored by the species of the corresponding diatom, see Fig. 3 for a visualization of species representatives.

3 Comparing the Geometries

In this section, we compare the geometries in Euclidean and Riemannian cases. The aim is to try and understand, when the performance is improved. We do this by visualizing the distribution of data point distances to the corresponding population means, the distances and means computed according to the corresponding metrics.

As can be seen in Fig. 4, in the femur (2-sphere) and diatom (Kendall's shape space) cases, the distributions look very similar. In fact, in the diatom case, they are essentially the same. The Kendall's shape space forms a quotient

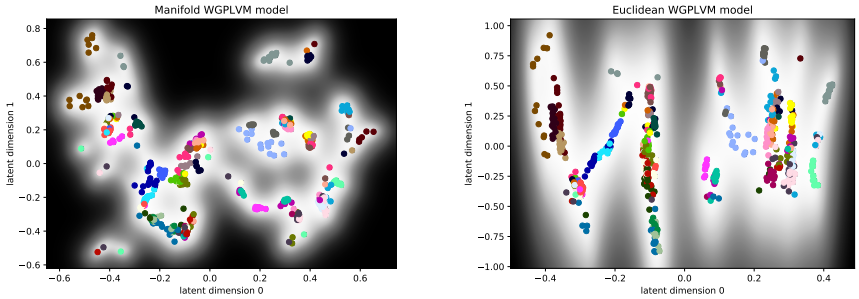

Figure 2: The latent spaces for the diatom dataset learned using the WGPLVM and GPLVM models. The colors indicate the species of the diatom corresponding to the latent variable, see Fig. 3.

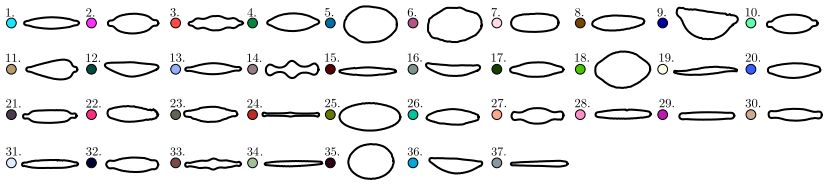

Figure 3: Representatives of each of the 37 diatom classes with corresponding class colors used in Fig. 2. Note that variation inside of each class can be considerable.

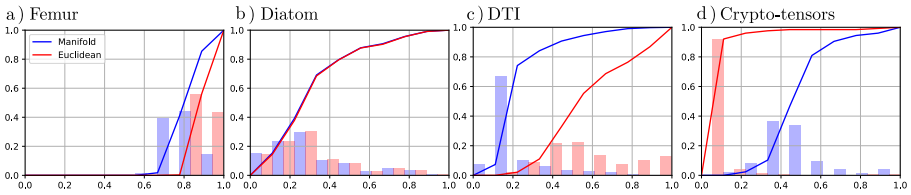

Figure 4: Distributions of distances between the data points and the population means. The bar plots indicate the density of data points that lie x-fraction of the maximum distance away from the mean. The corresponding continuous curves represent the cumulative distributions.

manifold of the sphere, which in this case is high dimensional ($d = 180$). In such high dimension, escaping the manifold becomes increasingly more difficult (most of the volume of the sphere is close to the boundary), and thus both the metrics are essentially the same. This might explain, why the WGPLVM did not improve notably on the GPLVM.

In the crypto-tensor experiment, the distribution implies the presence of

extreme outliers under the Euclidean metric. The Log-Euclidean metric, on the other hand, transforms the metric scale, evening out the distribution. This could very well explain, why we see large improvement with the WGPLVM compared to the GPLVM.

In the DTI experiment, the distribution of Euclidean distances looks more even. This might imply, that in this occasion, the Euclidean distance is better at capturing the trend of the data. However, the improved uncertainty estimates of the WGPLVM could be explained, as the Euclidean models are not confined to $SPD(n)$. Therefore, the distributions do not follow the conic shape of $SPD(n)$.