Abstracts

On the Geometry of Latent Variable Models

Søren Hauberg

Latent variable models (LVMs) describe the distribution of data $\mathbf{y} \in \mathcal{Y} = \mathbb{R}^D$ through a low-dimensional random variable $\mathbf{x} \in \mathcal{X} = \mathbb{R}^d, (d \ll D)$ and a (generally nonlinear) stochastic mapping $f : \mathcal{X} \to \mathcal{Y}$. Here we discuss the random Riemannian geometry induced by this stochastic mapping. The presented results was first stated in [1, 2].

To make the discussion explicit, we consider a Gaussian Process (GP) LVM [3] where f has component-wise conditionally independent Gaussian process entries,

$$f_i(\mathbf{x}) \sim \mathcal{GP}(m_i(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \qquad \forall i = 1, \ldots, D. \tag{1}$$

Here m_i and k are the mean and covariance functions of the i^{th} GP. Note that we assume the same covariance function across all dimensions as this simplify future calculations. The key presented results holds regardless of this simplification.

Assuming k is sufficiently smooth covariance then the image of a sample from f is a smooth d-dimensional immersed manifold. Note that this manifold is only locally diffeomorphic to d-dimensional Euclidean space, and it may globally self-intersect. It is then natural to consider the pull-back metric $\mathbf{M} = \mathbf{J}^\top \mathbf{J}$ over $\mathcal{X}$, where $\mathbf{J} \in \mathbb{R}^{D \times d}$ is the Jacobian of f. This defines a Riemannian metric over $\mathcal{X}$. Since f is stochastic, $\mathbf{M}$ is a stochastic object as well.

Since Gaussian variables are closed under differentiation, then $\mathbf{J}$ follows a GP,

$$\mathbf{J} \sim \prod_{j=1}^{D} \mathcal{N}(\mu(j,:), \Sigma) = \prod_{j=1}^{D} \mathcal{N}(\partial \mathbf{K}_{\mathbf{x},*}^\top \tilde{\mathbf{K}}_{\mathbf{x},\mathbf{x}}^{-1} \mathbf{Y}_{:,j}, \partial^2 \mathbf{K}_{*,*} - \partial \mathbf{K}_{*,\mathbf{x}}^\top \mathbf{K}_{\mathbf{x},\mathbf{x}}^{-1} \partial \mathbf{K}_{*,\mathbf{x}}), \tag{2}$$

where we use standard notation for GPs [4]. It then follows that $\mathbf{M}$ at a given point is governed by a non-central Wishart distribution [5]

$$\mathbf{M} \sim \mathcal{W}_d(D, \Sigma, \mathbb{E}[\mathbf{J}]^\top \mathbb{E}[\mathbf{J}]). \tag{3}$$

The entire metric by definition follows a generalized Wishart process [6].

Since the metric is a stochastic variable, we cannot apply standard Riemannian geometry to understand the space $\mathcal{X}$ (e.g. curvature is stochastic, geodesics are solutions to a stochastic differential equation, etc.). We can, however, inspect the leading moments of the metric

$$\mathbb{E}[\mathbf{M}] = \mathbb{E}[\mathbf{J}^\top \mathbf{J}] = \mathbb{E}[\mathbf{J}]^\top \mathbb{E}[\mathbf{J}] + D\,\Sigma \qquad = \mathcal{O}(D) \tag{4}$$

$$\operatorname{var}[M_{ij}] = D(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) + \mu_j^\top \Sigma \mu_j + \mu_i^\top \Sigma \mu_i \qquad = \mathcal{O}(D) \tag{5}$$

which we see both grow linearly with the dimension of $\mathcal{Y}$. This motivate the question as to how the pull-back metric behaves in high dimensions, $D \to \infty$. To ensure that the inner product of $\mathcal{Y}$ converges to the usual L^2 inner product in the limit $D \to \infty$ we let

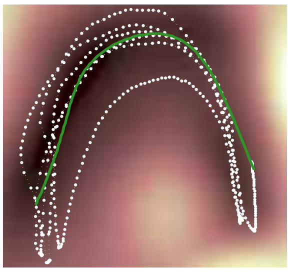
$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{Y}} = \frac{1}{D} \sum_{i=1}^{D} a_i b_i \xrightarrow[D\to\infty]{} \int a_t b_t \mathrm{d}t. \tag{6}$$

Then the natural pull-back becomes $\tilde{\mathbf{M}} = \frac{1}{D}\mathbf{J}^\top \mathbf{J}$, which has moments

$$\mathbb{E}[\tilde{\mathbf{M}}] = \mathbb{E}\left[\frac{1}{D}\mathbf{J}^\top \mathbf{J}\right] = \frac{1}{D}\mathbb{E}[\mathbf{J}]^\top \mathbb{E}[\mathbf{J}] + \Sigma \qquad = \mathcal{O}(1) \tag{7}$$

$$\mathrm{var}[\tilde{M}_{ij}] = \frac{1}{D}(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) + \frac{1}{D^2}\mu_j^\top \Sigma \mu_j + \frac{1}{D^2}\mu_i^\top \Sigma \mu_i \qquad = \mathcal{O}\left(\frac{1}{D}\right) \tag{8}$$

In the limit $D \to \infty$ we, thus, see that the variance vanishes and the metric becomes fully deterministic even if the underlying manifold is a stochastic object.



Implications and Extensions. This simple-to-prove result is rather surprising: even if we only have stochastic information about the underlying data manifold, its metric is deterministic. Furthermore, from Eq. 7 we see that this deterministic metric correspond to the (usual) pull-back metric of the mean f plus an additional term capturing the uncertainty of the manifold. This imply that the metric is large in regions of low data density (where the manifold is uncertain), and consequently, that geodesics will tend to avoid such regions. One such example is shown in the figure. Here human motion capture data $\mathbf{y}$ is used to estimate a two-dimensional manifold [1]. In the figure white points correspond to low-dimensional representations of the data, the green curve is an example geodesic computed under the expected metric, and the background color is proportional to the volume measure induced by the expected metric. We see that the metric is "larger" in regions of low data density and that geodesics consequently follow the structure of the data. The latter is a useful property when analyzing real data as distance-based data distribution will adapt well to the data [2].

From a practical point of view, geodesics can be computed in $\mathcal{X}$ by numerically solving the usual system of ordinary differential equations under the expected metric. The solution will be a curve in $\mathcal{X}$, which correspond to a GP in $\mathcal{Y}$. As such, geodesics remain stochastic objects, but they can be determined by solving a set of deterministic equations.

The presented derivations rely on the dimensions of $f(\mathcal{X})$ being conditionally independent, which is a common assumption. It can be eased upon: if the dimensions are (imperfectly) correlated, then the variance will still decrease, albeit at a slower rate than D^{-1}. Consequently, as a general rule of thumb, *the stochastic pull-back metric of an uncertain manifold immersed in a high-dimensional space is well approximated by the (deterministic) expected metric.*

Acknowledgments. SH was supported by a research grant (15334) from VILLUM FONDEN. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n^{o} 757360).

References

[1] A. Tosi, S. Hauberg, A. Vellido, and N. Lawrence, *Metrics for Probabilistic Geometries*, In The Conference on Uncertainty in Artificial Intelligence (2014).
[2] G. Arvanitidis, LK. Hansen, and S. Hauberg, *Latent Space Oddity: on the Curvature of Deep Generative Models*, In the International Conference on Learning Representations (2018).
[3] N. Lawrence, *Probabilistic non-linear principal component analysis with Gaussian process latent variable models*, Journal of machine learning research 6. Nov (2005): 1783-1816.
[4] CE. Rasmussen, and CKI. Williams, *Gaussian Processes for Machine Learning*, University Press Group Limited (2006).
[5] RJ. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons (2005).
[6] AG. Wilson, and Z. Ghahramani, *Generalised Wishart Processes*, In The Conference on Uncertainty in Artificial Intelligence (2011).