Abstracts

Diffusion Processes and PCA on Manifolds

Stefan Sommer

Approaching dimensionality reduction on differentiable manifolds with affine connection from a probabilistic viewpoint, we develop a generalization of Principal Component Analysis (PCA) that does not rely on parametric representations of principal submanifolds. The method fits a class of diffusions processes arising as horizontal stochastic flows in the frame bundle to observed data by maximum likelihood. The probabilistic interpretation removes the reliance of previous methods on explicitly constructed submanifolds that are not totally geodesic. In addition, projections to dense geodesics are avoided thus giving a well-defined construction on tori where projections do not exist.

1. Background

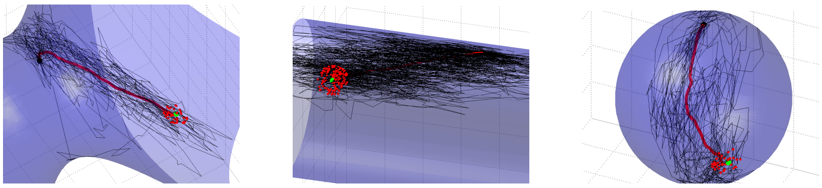
Conventional PCA uses the inner product structure of Euclidean space, a fact that makes generalization of the procedure to differentiable manifolds and stratified spaces difficult. Existing non-Euclidean extensions of PCA include Principal Geodesic Analysis (PGA, [1]) that parametrizes low-dimensional principal components with geodesic sprays from a Frechét mean; Geodesic PCA (GPCA, [2]) that finds principal geodesic curves minimizing residual errors; Principal Nested Spheres (PNS, [3]) that finds low-dimensional spheres; and Horizontal Component Analysis (HCA, [4]) that uses development of curves in the frame bundle of the manifold to construct horizontal subspaces that project to the manifold. These approaches capture different properties of Euclidean PCA but they are all limited by the fact that totally geodesic submanifolds do not exist in general in non-Euclidean spaces. This is in contrast to the Euclidean case where linear subspaces are totally geodesic. In addition, subspaces maximizing captured variance are not equivalent to subspaces minimizing residual errors. Even for distributions with local support, recurring and dense geodesics can make projections minimizing residual errors undefined [5].

2. Diffusion PCA

The principal subspaces that in Euclidean PCA maximize variance of orthogonally projected data samples can be found by eigendecomposing the sample covariance matrix C. The principal axes are given by unit eigenvectors u_j of C, and with $U = (u_1, \dots, u_d)$, the principal components $x_n = U^T(y_n - \mu)$ are projections of the centered data to the span of the principal axes. Conventionally, u_j are ordered according to decreasing eigenvalues.

In [6], a probabilistic formulation of PCA was developed as a maximum likelihood estimate (MLE) of the matrix W in the latent variable model

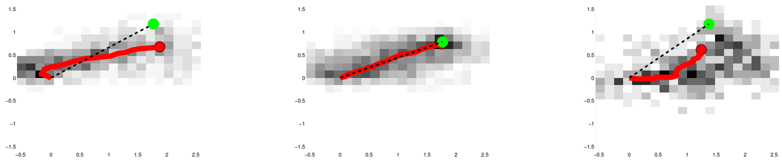
$$y = Wx + \mu + \epsilon \ .$$



(a) sample paths, elliptic hyperboloid, $-2x^2 + y^2 = -z^2$

(b) sample paths, cylinder, $x^2 + 0y^2 = -z^2$

(c) sample paths, sphere, $x^2 + y^2 = -z^2$



(d) $\mathbb{R}^2 \cong T_pM$, mean path (red) and density from (a)

(e) $\mathbb{R}^2 \cong T_pM$, mean path (red) and density from (b)

(f) $\mathbb{R}^2 \cong T_pM$, mean path (red) and density from (c)

Figure 1. Anisotropic diffusion on quadratic hypersurfaces, source $p = (0, 0, 1)$; var. major axis: 4; minor axis: 1.2. (a),(b),(c): sample paths ending near $\mathbf{x} = \mathrm{Exp}_p x$, $x \in T_pM$ (green dot). (d),(e),(f): mean sample path anti-development $\hat{x}_i(t)$ (red) with path densities (background) and shortest path from source $(0, 0)$ to x (dashed line). The paths all reach $\mathbf{x}$ through different anti-developments in $\mathbb{R}^2$. Negative curvature (a),(d): mean path deviates from shortest path by moving along minor diffusion axis before major axis. Zero curvature (b),(e): mean path and shortest path align. Positive curvature (c),(f): mean path aligns with major diffusion axis before moving along minor axis. This case resembles the HCA construction, see text.

The latent variables x are assumed normally distributed $\mathcal{N}(0, I)$, and, in contrast to factor analysis, the noise ϵ is isotropic $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. The marginal distribution of the observed data is then again Gaussian $\mathcal{N}(\mu, C_\sigma)$, $C_\sigma = WW^T + \sigma^2 I$. The MLE for W is up to rotation given by $W_{ML} = U(\Lambda - \sigma^2 I)^{1/2}$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ with the usual PCA solution recovered as $\sigma^2 \to 0$. The principal components x_n are defined as the mean of x conditional on the sample y_n and given by $E[x_n|y_n] = (W^T W + \sigma^2 I)^{-1} W_{ML}^T (y_n - \mu)$. This definition again approaches the orthogonal projections used in PCA as $\sigma^2 \to 0$.

Importantly for our purpose, the probabilistic formulation makes no reference to linear subspaces. Thus a generalization to nonlinear manifolds can be obtained without constructing submanifolds that in general cannot be totally geodesic. The main difficulty in generalizing the method instead lies in the latent variable model being additive and using normal distributions, neither which are directly transferable to manifolds.

We now define *diffusion PCA* (DPCA) that rephrases the latent variable model using stochastic paths and diffusions processes. This formulation naturally extends to differentiable manifolds with affine connection. Let W_t be a Wiener process in $\mathbb{R}^d$ and let X_t be given by the $\mathbb{R}^n$ valued stochastic differential equation (SDE) $dX_t = \sigma \circ W_t$ with source $X_0 = \delta_{(0,\ldots,0)}$. Here the $n \times d$ matrix σ is stationary so that X_t is a driftless diffusion with infinitesimal generator $\sigma\sigma^T$. Through the process of stochastic development [7], the process maps to a stochastic process U_t in the frame bundle FM of a manifold M^n with affine connection: If $H_1, \ldots, H_n$ are the horizontal vector fields on FM, U_t satisfies $dU_t = H_i(U_t)\sigma^i_j \circ dW^j_t$, and the source is a point $(p, u) \in FM$, $p \in M$. Trough the bundle projection $\pi : FM \to M$, U_t projects to a manifold valued diffusion $\pi(U_t)$ which is unique given its generator L and initial distribution [7].

Consider the map Diff. : $FM \to \text{Dens}(M)$ that maps (p, u) to $\pi(U_1)$ where the FM diffusion $dU_t = H_i(U_t) \circ dW^i_t$ is started at (p, u) and $W_t \in \mathbb{R}^n$. We let $\Gamma \subset \text{Dens}(M)$ be the image Diff.(FM), i.e. the set of densities resulting from point-sourced diffusions in FM stopped at time $t = 1$. In diffusion PCA, the observed data is assumed to be distributed according to $\mu \in \Gamma$ so that $y \sim \pi(U_1)$ for a diffusion $U_1 \in FM$.

Let μ_0 be a fixed measure on M (e.g. a Riemannian volume form) and for $\mu = p\mu_0 \in \Gamma$ define the log-likelihood

$$\ln \mathcal{L}(\mu) = \sum_{i=1}^{N} \ln p(y_i)$$

for a set of samples $y_1, \ldots, y_N \in M$. Now let $(p, u) \in FM$ be a maximum for $\ln \mathcal{L}(\text{Diff.}(p, u))$. Then (p, u) is an MLE of y_i generalizing the probabilistic PCA formulation to the non-Euclidean case. We denote (p, u) a diffusion PCA. Note that Diff. is not injective as multiple frames at p may lead to the same diffusion process. The MLE is hence not unique though a different formulation of Diff. can correct this (see open questions below).

3. The Principal Components

In probabilistic PCA, the mean of the latent variables conditional on the observed data $E[x_n|y_n]$ converges to the principal components as $\sigma^2 \to 0$. With non-zero curvature, single vectors cannot summarize the observations in this way because of the path dependences of the diffusion, see Figure 1. Instead, the mean sample paths reaching y_i $\hat{x}_i(t) = E[x(t)|x(1) = y_i]$ take the role of the latent variables in probabilistic PCA. Note that given the source $(p, u) \in FM$, the sample paths can be equivalently viewed as paths on M or as paths in $\mathbb{R}^n$ trough (anti-)development. Examples of mean paths are illustrated in Figure 1. In $\mathbb{R}^n$, the data can be further summarized by integrating out the time dependence from $\hat{x}_i$ giving $\tilde{x}_i = \int_0^1 \hat{x}_i(t)dt$.

If M is Riemannian, let $u_0 \in OM$ be an orthonormal frame at p such that the matrix of uu^T in the u_0 basis is diagonal with decreasing diagonal. Anti-developing $\hat{x}_i$ with base (p, u_0) gives $\mathbb{R}^n$ valued paths with the major variation residing in the low coordinates. The vectors $\tilde{x}_i$ here provide a Euclideanization of the data similar to those provided by PGA, GPCA, and HCA. In general, the linearization will differ from the linearizations provided by the existing methods.

In effect, the complicated geometric problems arising when defining parametric subspaces of non-linear manifolds and projecting data are removed with the probabilistic approach. In particular, recurring and dense geodesics on tori prevent a well-defined notion of projection as closest point on a geodesic as required by existing methods. With diffusion PCA, the mean path $\hat{x}_i$ is defined without projecting to a submanifold. Note in Figure 1 (f) that $\hat{x}_i$ resembles the axis-aligned curves of HCA [4] indicating a new characterization of HCA as approximating diffusion processes in positively curved spaces.

4. The Geometry of Γ

The set $\mathrm{Dens}(M) = \{\mu \in \Omega^n(M) : \int_M \mu = 1, \mu > 0\}$ is an infinite dimensional manifold in the Fréchet topology of smooth functions [8], and it can be equipped with the Fisher-Rao metric $G_\mu^{FR}(\alpha, \beta) = \int_M \frac{\alpha}{\mu}\frac{\beta}{\mu}\mu$ [9]. In information geometry, finite dimensional submanifolds of $\mathrm{Dens}(M)$ are called statistical manifolds. In coordinates, the matrix form of the metric is the Fisher Information Matrix. If Γ locally has the structure of a statistical manifold, the MLE in diffusion PCA can be found by a gradient flow with respect to the gradient inherited from the Fisher-Rao metric, i.e.

$$\dot{\theta}_s = -\nabla_\theta \left(\sum_{i=1}^N \ln p_\theta(y_i) \right) \tag{1}$$

where θ is a local chart for Γ and p_θ is the distribution for a given value of θ.

5. Open Questions

At the workshop, we identified a number of questions related to the diffusion PCA construction that are open for further research. These include finding the precise topological and geometric structure of Γ as a subset of $\mathrm{Dens}(M)$. We conjecture that Γ in the Riemannian case can be parametrized by the bundle of symmetric positive-definite covariant tensors of order 2. Conditions for the convergence of the gradient flow (1) to the MLE estimate needs to be identified. In addition, a scaled Brownian motion can be added to the diffusion PCA model similarly to the isotropic error ϵ in probabilistic PCA. The exact form of this construction and the convergence as $\sigma^2 \to 0$ remains to be explored.

References

[1] P.T. Fletcher, Conglin Lu, S.M. Pizer, and Sarang Joshi, *Principal geodesic analysis for the study of nonlinear statistics of shape*, Medical Imaging, IEEE Transactions on (2004).

[2] Stephan Huckemann, Thomas Hotz, and Axel Munk, *Intrinsic shape analysis: Geodesic PCA for riemannian manifolds modulo isometric lie group actions*, Statistica Sinica **20** (2010), no. 1, 1–100.
[3] Sungkyu Jung, Ian L. Dryden, and J. S. Marron, *Analysis of principal nested spheres*, Biometrika **99** (2012), no. 3, 551–568 (en).
[4] Stefan Sommer, *Horizontal dimensionality reduction and iterated frame bundle development*, Geometric Science of Information (Frank Nielsen and Frederic Barbaresco, eds.), Lecture Notes in Computer Science, Springer Berlin Heidelberg, January 2013, pp. 76–83.
[5] Stephan Huckemann, Thomas Hotz, and Axel Munk, *Rejoinder to the discussion of intrinsic shape analysis: Geodesic principal component analysis for riemannian manifolds under isometric lie group actions*, Statistica Sinica **20** (2010), no. 1.
[6] Michael E. Tipping and Christopher M. Bishop, *Probabilistic principal component analysis*, Journal of the Royal Statistical Society. Series B (Statistical Methodology) **61** (1999), no. 3, 611–622.
[7] Elton P. Hsu, *Stochastic analysis on manifolds*, American Mathematical Soc., 2002 (en).
[8] Richard S. Hamilton, *The inverse function theorem of nash and moser*, Bulletin of the American Mathematical Society **7** (1982), no. 1, 65–222.
[9] Klas Modin, *Information geometry and the Fisher-Rao metric on the space of probability distributions*, Math in the Cabin - Shape Analysis Workshop in Bad Gastein (Bad Gastein, Austria), July 2014, `https://hal.archives-ouvertes.fr/hal-01076953`.

Reporter: Stephan Huckemann