Diffusion Processes and Dimensionality Reduction on Manifolds
ESI, Vienna, Feb. 2015

Stefan Sommer
Department of Computer Science, University of Copenhagen
Outline

- Dimensionality Reduction
- Diffusion PCA
- Development and Anisotropic Diffusions
- Examples
Dimensionality Reduction in Non-Linear Manifolds

- **dim. reduction and linearizations** - mappings from non-linear manifolds to low dimensional Euclidean space that preserves structure of data
- Non-Euclidean generalizations of *PCA*:
  - Principal Geodesic Analysis (PGA, Fletcher et al., ’04)
  - Geodesic PCA (GPCA, Huckeman et al., ’10)
  - Horizontal Component Analysis (HCA, Sommer, ’13)
  - Principal Nested Spheres ((C)PNS, Jung et al., ’12)
  - Barycentric Subspaces (BS, Pennec, ’15)
Dimensionality Reduction in Non-Linear Manifolds

PGA: analysis relative to the data mean
Dimensionality Reduction in Non-Linear Manifolds

PGA: analysis relative to the data mean

data points on non-linear manifold
Dimensionality Reduction in Non-Linear Manifolds

- **dim. reduction and linearizations**: Mappings from non-linear manifolds to low dimensional Euclidean space that preserves structure of data

- Non-Euclidean generalizations of PCA:
  - Principal Geodesic Analysis (PGA, Fletcher et al., '04)
  - Geodesic PCA (GPCA, Huckeman et al., '10)
  - Horizontal Component Analysis (HCA, Sommer, '13)
  - Principal Nested Spheres ((C)PNS, Jung et al., '12)
  - Barycentric Subspaces (BS, Pennec, '15)

 PGA: analysis relative to the data mean

Intrinsic mean $\mu$
Dimensionality Reduction in Non-Linear Manifolds

PGA: analysis relative to the data mean

tangent space $T_{\mu}M$
Dimensionality Reduction in Non-Linear Manifolds

PGA: analysis relative to the data mean

projection of data point to $T_{\mu}M$
Dimensionality Reduction in Non-Linear Manifolds

 PGA: analysis relative to the data mean

 Euclidean PCA in tangent space
Dimensionality Reduction in Non-Linear Manifolds

- *dim. reduction and linearizations* - mappings from non-linear manifolds to low dimensional Euclidean space that preserves structure of data

- Non-Euclidean generalizations of *PCA*:
  - Principal Geodesic Analysis (PGA, Fletcher et al., ’04)
  - Geodesic PCA (GPCA, Huckeman et al., ’10)
  - Horizontal Component Analysis (HCA, Sommer, ’13)
  - Principal Nested Spheres ((C)PNS, Jung et al., ’12)
  - Barycentric Subspaces (BS, Pennec, ’15)
Dimensionality Reduction in Non-Linear Manifolds

- Dimensionality reduction and linearizations: Mappings from non-linear manifolds to low-dimensional Euclidean space that preserves the structure of data.

- Non-Euclidean generalizations of PCA:
  - Principal Geodesic Analysis (PGA, Fletcher et al., '04)
  - Geodesic PCA (GPCA, Huckeman et al., '10)
  - Horizontal Component Analysis (HCA, Sommer, '13)
  - Principal Nested Spheres ((C)PNS, Jung et al., '12)
  - Barycentric Subspaces (BS, Pennec, '15)

Stefan Sommer (sommer@diku.dk) (Department of Computer Science, University of Copenhagen) — Diffusion Processes and Dimensionality Reduction on Manifolds

PGA:

GPCA:

HCA:
PGA, GPCA, HCA, PNS, …

- search for explicitly constructed parametric subspaces: geodesic sprays, geodesics, iterated development, …
- in general manifolds, these subspaces are not totally geodesic
- projections to subspaces are problematic: geodesics may be dense on tori
## Generalizing Linear Statistics

<table>
<thead>
<tr>
<th>Euclidean</th>
<th>Riemannian</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm $| x - y |$</td>
<td>distances $d(x, y)$</td>
</tr>
<tr>
<td>vectors</td>
<td>$v_0$ for geodesics</td>
</tr>
<tr>
<td>linear subspaces</td>
<td>geodesic sprays</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

why are geodesics fundamental when estimating covariance?

- Euclidean space analogies can lead to non-local constructions
- to goal of this talk is to get closer to constructions defined “infinitesimally”
Euclidean PCA

Usual formulation:

- eigendecomposition \((u_1, \lambda_1), \ldots, (u_d, \lambda_d)\) of sample covar. matrix \(C\)
- principal components: \(x_n = U^T(y_n - \mu)\)

Probabilistic interpretation (Tipping, Bishop, ’99):

- latent variable model
  \[
  y = Wx + \mu + \epsilon, \quad x \sim \mathcal{N}(0, I), \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)
  \]
- marginal distribution \(y \sim \mathcal{N}(\mu, C_\sigma), \quad C_\sigma = WW^T + \sigma^2 I\)
- MLE of \(W\): \(W_{ML} = U(\Lambda - \sigma^2 I)^{1/2} + \text{rotation}\)
  \[
  \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)
  \]
- principal components:
  \[
  E[x_n | y_n] = (W^T W + \sigma^2 I)^{-1} W_{ML}^T (y_n - \mu)
  \]
Diffusion PCA

- probabilistic PCA does not explicitly use subspaces
- on Riemannian manifolds, the Eells-Elworthy-Malliavin construction gives a map
  \[
  \int_{\text{Diff}} : FM \rightarrow \text{Dens}(M)
  \]
- \( \Gamma \subset \text{Dens}(M) \): the image \( \int_{\text{Diff}}(FM) \), the set of (normalized) densities resulting from diffusions in \( FM \)
- \( \mu \in \Gamma \approx \) anisotropic normal distribution
- with \( \mu = \int_{\text{Diff}}(x, X_\alpha) = p_\mu \mu_0 \), define the log-likelihood
  \[
  \ln \mathcal{L}(x, X_\alpha) = \ln \mathcal{L}(\mu) = \sum_{i=1}^{N} \ln p_\mu(y_i)
  \]
- Diffusion PCA: maxim. \( \ln \mathcal{L}(x, X_\alpha) \) for \( (x, X_\alpha) \in FM \)
- MLE of data \( y_i \) under the assumption \( y \sim \mu \in \Gamma \)
Diffusion PCA

- all geometric complexities are hidden in the diffusion processes
- no parametric subspaces, no projections to dense geodesics
- principal components: mean sample paths reaching $y_i$

$$\hat{x}_i(t) = E[x(t)|x(1) = y_i]$$

- path dependency can be integrated out

$$\tilde{x}_i = \int_0^1 \frac{d}{dt} \hat{x}_i(t) dt = \hat{x}_i(1)$$

$\tilde{x}_i$ provides a linear view of the data
Diffusion PCA

- All geometric complexities are hidden in the diffusion processes.
- No parametric subspaces, no projections to dense geodesics.
- Principal components: mean sample paths reaching $y_i$
  $$\hat{x}_i(t) = \mathbb{E}[x(t) | x(1) = y_i]$$
- Path dependency can be integrated out.
  $$\tilde{x}_i = \int_0^1 \hat{x}_i(t) \, dt = \hat{x}_i(1)$$

$x_i$ provides a linear view of the data.
Statistical Manifold: Geometry of $\Gamma$

- Densities

$$\text{Dens}(M) = \{ \mu \in \Omega^n(M) : \int_M \mu = 1, \mu > 0 \}$$

- Fisher-Rao metric: $$G_{\mu}^{FR}(\alpha, \beta) = \int_M \frac{\alpha}{\mu} \frac{\beta}{\mu} \mu$$

- $\Gamma$ finite dim. subset of $\text{Dens}(M)$

- properties of $\int_{\text{Diff}} : FM \to \text{Dens}(M)$

- naturally defined on bundle of symmetric positive $T^0_2$ tensors
SDEs On Manifolds

- stationary driftless diffusion SDE in $\mathbb{R}^n$:
  \[ dX_t = \sigma dW_t, \quad \sigma \in M^{n \times d} \]

- diffusion field $\sigma$, infinitesimal generator $\sigma \sigma^T$

- curvature: stationary field/generator cannot be defined due to holonomy
SDEs On Manifolds

- stationary driftless diffusion SDE in $\mathbb{R}^n$:
  \[ dX_t = \sigma dW_t , \quad \sigma \in M^{n \times d} \]

- diffusion field $\sigma$, infinitesimal generator $\sigma \sigma^T$

- curvature: stationary field/generator cannot be defined due to *holonomy*
The Frame Bundle

- the manifold and frames (bases) for the tangent spaces $T_pM$
- $F(M)$ consists of pairs $u = (x, X_\alpha)$, $x \in M$, $X_\alpha$ frame for $T_xM$
- curves in the horizontal part of $F(M)$ correspond to curves in $M$ and parallel transport of frames
The Frame Bundle

- the manifold and frames (bases) for the tangent spaces $T_x M$
- $F(M)$ consists of pairs $u = (x, X^\alpha)$, $x \in M$, $X^\alpha$ frame for $T_x M$
- curves in the horizontal part of $F(M)$ correspond to curves in $M$ and parallel transport of frames
The Frame Bundle

- the manifold and frames (bases) for the tangent spaces $T_pM$
- $F(M)$ consists of pairs $u = (x, X_\alpha)$, $x \in M$, $X_\alpha$ frame for $T_xM$
- curves in the horizontal part of $F(M)$ correspond to curves in $M$ and parallel transport of frames
SDEs On Manifolds:
Eells-Elworthy-Malliavin construction

- $H_i, i = 1 \ldots, n$ horizontal vector fields on $F(M)$:
  \[
  H_i(u) = \pi_*^{-1}(u_i)
  \]

- SDE in $\mathbb{R}^n$:
  \[
  dX_t = \sigma(X_t) dW_t, \quad \sigma(X_t) \in M^{n \times n}
  \]

- SDE in $F(M)$:
  \[
  dU_t = H_i(U_t) \circ dX^i_t = H_i(U_t) \circ \sigma(X_t)^i_j dW^j_t
  \]

- SDE on $M$:
  \[
  \pi_{F(M)}(U_t)
  \]
SDEs on Manifolds:

Eells-Elworthy-Malliavin construction

- $H_i, i = 1 \ldots n$ horizontal vector fields on $F(M)$:

\[ H_i(u) = \pi^{-1} \star (u_i) \]

- SDE in $\mathbb{R}^n$:

\[ dX_t = \sigma(X_t) \, dW_t, \quad \sigma(X_t) \in \mathbb{R}^{n \times n} \]

- SDE in $F(M)$:

\[ dU_t = H_i(U_t) \circ dX_t^i = H_i(U_t) \circ \sigma(X_t) \, dW_t^i \]

- SDE on $\pi F(M)$:

\[ \pi F(M)(U_t) \]
Anisotropic Diffusions

\[ \int_{\text{Diff}} (x, X_\alpha) = \pi_{F(M)}(U_1) \]

- stochastic development / “rolling without slipping”
- in \( \mathbb{R}^n \), sample path increments \( \Delta x_{t_i} = x_{t_{i+1}} - x_{t_i} \) are normally distributed \( \mathcal{N}(0, (\Delta t)^{-1} \Sigma) \) with log-probability

\[ \ln \tilde{p}_\Sigma(x_t) \propto -\frac{1}{\Delta t} \sum_{i=1}^{N-1} \Delta x_{t_i}^T \Sigma^{-1} \Delta x_{t_i} + c \]

- Formally, we can set

\[ \ln \tilde{p}_\Sigma(x_t) \propto -\int_0^1 \| \dot{x}_t \|^2 \Sigma dt + c \]
“Most Probable Paths”

- Let $\phi$ be path development and $U_t$ a heat diffusion. As $N \to \infty$, the finite path measure $\to$ pullback $\phi^* \nu$ of Wiener measure $\nu$ on $W([0, 1], M)$ (cont. paths from $x$) (Andersson, Driver, ’99)

- geodesics and be formally viewed as “most probable paths” (MPPs) for the pull-back path density, i.e. MPPs for the Euclidean $\mathbb{R}^n$ diffusion

- Anisotropic case: $(x_t, X_{\alpha, t})$ path in $FM$, $X_{\alpha, t}$ represents $\Sigma^{1/2}$ and defines invertible map $\mathbb{R}^n \to T_{x_t}M$. Inner product on $T_{x_t}M$

$$\langle v, w \rangle_{X_{\alpha, t}} = \langle X_{\alpha, t}^{-1} v, X_{\alpha, t}^{-1} w \rangle_{\mathbb{R}^n}$$
Sub-Riemannian Geometry

- optimal control problem with nonholonomic constraints

\[ x_t = \arg \min_{c_t, c_0=x, c_1=y} \int_0^1 \| \dot{c}_t \|_{X_{\alpha, t}}^2 \, dt \]

- let

\[ \langle \tilde{v}, \tilde{w} \rangle_{HFM} = \langle X^{-1}_{\alpha, t} \pi_*(\tilde{v}), X^{-1}_{\alpha, t} \pi_*(\tilde{w}) \rangle_{\mathbb{R}^n} \]

on \( H(x_t, x_{\alpha, t}) \) FM. This defines a sub-Riemannian metric \( G \) on \( TFM \) and equivalent problem

\[ (x_t, X_{\alpha, t}) = \arg \min_{(c_t, C_{\alpha, t}), c_0=x, c_1=y} \int_0^1 \| (\dot{c}_t, \dot{C}_{\alpha, t}) \|_{HFM}^2 \, dt \]

with constraints \( (\dot{c}_t, \dot{C}_{\alpha, t}) \in H(c_t, C_{\alpha, t}) FM \)
Evolution Equations

- Geodesics satisfy the Hamilton-Jacobi equations

\[
\dot{y}^k_t = G^{kj}(y_t)\xi_{t,j}, \quad \dot{\xi}_{t,k} = -\frac{1}{2} \frac{\partial G^{pq}}{\partial y^k} \xi_{t,p} \xi_{t,q}
\]

- In coordinates \((x^i)\) for \(M\), \(X^i_\alpha\) for \(X_\alpha\), and \(W\) encoding the inner product \(W^{kl} = \delta^{\alpha\beta} X^k \alpha X^l \beta\):

\[
\dot{x}^i = W^{ij} \xi_j - W^{ih} \Gamma^j_h \xi_{j\beta}, \quad \dot{X}^i_\alpha = -\Gamma^i_h W^{hj} \xi_j + \Gamma^i_k W^{kh} \Gamma^j_h \xi_{j\beta}
\]

\[
\dot{\xi}^i = W^{hl} \Gamma^k_l \xi_h \xi_k \delta - \frac{1}{2} \left( \Gamma^h_{\gamma i} W^{kh} \Gamma^k_h \delta + \Gamma^h_{\gamma k} W^{kh} \Gamma^i_h \delta \right) \xi_h \xi_k \delta
\]

\[
\dot{\xi}^i_\alpha = \Gamma^h_{\gamma k} W^{kh} \Gamma^i_h \delta \xi_h \xi_k \delta - \left( W^{hl} \Gamma^k_l + W^{hl} \Gamma^k_{l,i\alpha} \right) \xi_h \xi_k \delta
\]

\[-\frac{1}{2} \left( W^{hk} \xi_h \xi_k + \Gamma^h_{\gamma k} W^{kh} \Gamma^i_h \delta \right) \xi_h \xi_k \delta
\]

Stefan Sommer (sommer@diku.dk) (Department of Computer Science, University of Copenhagen) — Diffusion Processes and Dimensionality Reduction on Manifolds

Slide 16/22
$S^2$

(a) $\text{cov. diag}(1, 1)$    (b) $\text{cov. diag}(2, .5)$    (c) $\text{cov. diag}(4, .25)$
$S^2$
Landmark LDDMM

- Christoffel symbols (Michelli et al. ’08)

\[ \Gamma^k_{ij} = \frac{1}{2} g_{ir} \left( g_{kl} g^{rs} \frac{\partial}{\partial x^s} - g_{sl} g^{rk} \frac{\partial}{\partial x^k} - g_{rl} g^{ks} \frac{\partial}{\partial x^s} \right) g_{sj} \]

- mix of transported frame and cometric: \( F^d M \) bundle of rank \( d \) linear maps \( \mathbb{R}^d \to T_x M \), \( \xi, \tilde{\xi} \in T^* F^d M \), cometric

\[ \langle \xi, \tilde{\xi} \rangle = \delta^{\alpha\beta} (\xi|\pi_*^{-1} X_\alpha)(\tilde{\xi}|\pi_*^{-1} X_\beta) + \lambda \langle \xi, \tilde{\xi} \rangle_{g_R} \]

- the whole frame need not be transported
Landmark LDDMM Examples

(a) cov. diag(2, .5)^2  
(b) cov. diag(1, 1)  
(c) cov. diag(.5, 2)^2

(d)  
(e)  
(f)
Summary

- Diffusion PCA: dim. reduction or linearization by MLE of densities generated by diffusion processes
- Infinitesimal definition, no subspaces, no projections
- Diffusion map $\int_{\text{Diff}} : FM \rightarrow \text{Dens} (M)$ from Eells-Elworthy-Malliavin construction of Brownian motion
- “MPPs” for anisotropic diffusions generalize geodesics
