A Level Set Based Variational Principal Flow Method for Nonparametric Dimension Reduction on Riemannian Manifolds

Shingyu Leung
Department of Mathematics
HKUST

Big Data Challenges for Predictive Modeling of Complex Systems, HKU, 28 Nov 2018
Collaborators

- Hao Liu (Georgia Tech)
- Zhigang Yao (NUS)
- Tony Chan (KAUST)
- Tony Wong (UBC)
• Background
• Our Proposed Model
• Some Implementation Details
• Numerical Examples

The work was supported in part by the Hong Kong RGC grants 16303114 and 16309316

Outline
Variation in a Dataset

Linear regression

Principal component analysis (PCA)

Curve fitting

Question: What about a dataset that satisfies a nonlinear constraint?
• Extend the PCA-idea to datasets that satisfy nonlinear constraints

• Determine curves on the manifold that retain their canonical interpretability as principal components

Goals
- **Principal Geodesic Analysis (PGA)**
  - first principal component passes through the mean of the data with the min average geodesic distance to the sample points (Huckemann and Ziezold 2006; Fletcher et al. 2008;...)

- **Tangent space PCA**
  - project all data onto the tangent space at the mean, find the direction of maximal variability, finally map the component back to the manifold (Goodall 1991; Fletcher and Joshi 2007;...)

...
Typical PCA vs. Curve fitting
Finding a smooth curve: its derivative is “close” to the first principal component by a local tangent PCA at that same point

Principal Flow (PF) (Panaretos, Pham and Yao, JASA, 2014)
• Where to start the marching or the ray tracing?
  • The Frechet mean might not give a good data descriptor
• Evolution: A Lagrangian formulation

Frechet mean

$$\mathbf{x} = \arg\min_{\mathbf{y}} \sum d^2(\mathbf{y}, \mathbf{x}_i)$$

Properties
Our Contributions

• Develop a fully Eulerian formulation
  • The surface for the constraint
  • The curve on the surface
• Robust to noise
• Balance the contributions from the data and from the PF
Applications of Interface Problems

Closest Point Method

with Y. Ning

with H.K. Zhao
Implicitly by a scalar function

\[ \phi(\vec{x}) : \mathbb{R}^n \rightarrow \mathbb{R} \]

\[ \Gamma = \phi^{-1}(c) = \{ \vec{x} : \phi(\vec{x}) = c \} \]

\[ \phi(x, y) = 1_{\Omega} \quad \text{Phase Field Method (Allen-Cahn 1979)} \]

\[ \phi(\vec{x}) = \pm d \quad \text{Level Set Method (Sethian-Osher 1988)} \]

Eulerian Interface Representations
Our Variation Approach

The geodesic distance from $x$ to the dataset on the manifold:

$$E(\Gamma) = \left[ \int_{\Gamma} d^p (x(s)) ds \right]^{1/p} + \lambda \left[ \int_{\Gamma} |n(x(s)) \cdot p(x(s))|^2 ds \right]^{1/2}$$

Challenge: The representation of the curve $\Gamma$

$n$: normal direction
$p$: principal direction

The geodesic distance from $x$ to the dataset on the manifold.
\[ E(\Gamma) = \left[ \int_{\Gamma} d^p(x(s)) ds \right]^{1/p} + \lambda \left[ \int_{\Gamma} |n(x(s)) \cdot p(x(s))|^2 \, ds \right]^{1/2} \]

\[ \Gamma = \left\{ x \in \mathbb{R}^m : \psi(x) = \phi(x) = 0 \right\} \]

for the manifold

\[ \phi(x, y, z) = x^2 + y^2 - \frac{1}{2^2} \]

\[ \psi(x, y, z) = x^2 + y^2 + z^2 - 1 \]

Our Eulerian Formulation
\[
E(\Gamma) = \left[ \int_{\Gamma} d^p(x(s)) \, ds \right]^{\frac{1}{p}} + \lambda \left[ \int_{\Gamma} |n(x(s)) \cdot p(x(s))|^2 \, ds \right]^{\frac{1}{2}}
\]

\[
E(\phi; \lambda) = \left[ \int_{\mathbb{R}^m} d^p(x)\delta(\phi)\delta(\psi)\|\nabla \phi\| \, dx \right]^{\frac{1}{p}} + \lambda \left[ \int_{\mathbb{R}^m} |n(x) \cdot p(x)|^2 \delta(\phi)\delta(\psi)\|\nabla \phi\| \, dx \right]^{\frac{1}{2}}
\]

\[
= \left[ \int_{M_e} d^p(x)\delta(\phi)\delta(\psi)\|\nabla \phi\| \, dx \right]^{\frac{1}{p}} + \lambda \left[ \int_{M_e} |n(x) \cdot p(x)|^2 \delta(\phi)\delta(\psi)\|\nabla \phi\| \, dx \right]^{\frac{1}{2}}
\]

\[
\left( \frac{\delta E}{\delta \phi}, h \right) = -\gamma_1 I_1 - \lambda \gamma_2 (I_2 + I_3)
\]

\[
P = I - n \otimes n\] projection onto the direction perpendicular to \(n\)
\[
\frac{\partial \phi}{\partial t} = \nabla \cdot \left[ \gamma_1 d^p(x) \nabla \phi + \lambda \gamma_2 (p + Pp)(p \cdot \nabla \phi) \right]
\]

Some Properties

First term: drives the curve towards the data
Second term:
1) free \( p \), replace \( p \) by \(-p\)
2) \( P \) always contributes to the motion until \( p \) is perpendicular to \( n \)

\[ P = I - n \otimes n \]
Some Implementation Details

I. The geodesic distance

II. An approximation to the principal direction

III. Level set regularizations

\[ E(\phi; \lambda) = \left[ \int_{M_e} d^p(x) \delta(\phi) \delta(\psi) \left\| \nabla \phi \right\| dx \right]^{\frac{1}{p}} + \lambda \left[ \int_{M_e} \left| n(x) \cdot p(x) \right|^2 \delta(\phi) \delta(\psi) \left\| \nabla \phi \right\| dx \right]^{\frac{1}{2}} \]

\[ \frac{\partial \phi}{\partial t} = \nabla \cdot \left[ \gamma_1 d^p(x) \nabla \phi + \lambda \gamma_2 (p + Pp)(p \cdot \nabla \phi) \right] \]

I. The geodesic distance

II. An approximation to the principal direction

III. Level set regularizations

Some Implementation Details
I. Geodesic Distance – A First Attempt

Based on the Fast Marching Method, Memoli and Sapiro, JCP, 01.
Fast Sweeping Method

\[ |\nabla u(x)| = 1 \]

Godunov Hamiltonian:
(Rouy and Tourin 92)

\[
\left[ \left( u_{i,j} - u_{\text{min}}^x \right)^+ \right]^2 + \left[ \left( u_{i,j} - u_{\text{min}}^y \right)^+ \right]^2 = 1
\]

Fast sweeping: Iteratively update the solution based on the local solver according to different sweeping direction

- \( i=1:M \) and \( j=1:N \)
- \( i=1:M \) and \( j=N:-1:1 \)
- \( i=M:-1:1 \) and \( j=1:N \)
- \( i=M:-1:1 \) and \( j=N:-1:1 \)

\[
\begin{cases}
\min(u_{\text{min}}^x, u_{\text{min}}^y) + \Delta x & \text{if } |u_{\text{min}}^x - u_{\text{min}}^y| > \Delta x \\
\frac{1}{2} \left( u_{\text{min}}^x + u_{\text{min}}^y + \sqrt{2\Delta x^2 - \left( u_{\text{min}}^x - u_{\text{min}}^y \right)^2} \right) & \text{otherwise}
\end{cases}
\]

Algorithm 1: The fast sweeping method for the eikonal equation \( |\nabla u| = 1 \) or the advection equation \( \mathbf{v} \cdot \nabla u = (p, q) \cdot (u_x, u_y) = 0 \).

**Data:** The set of boundary points \((x, y)_a\), the mesh size \(\Delta x\).

**Result:** \(u_{i,j}\) in the computational domain.

**Initialization:** Assign the boundary condition. For the eikonal equation, set \(u_{i,j} = 0\) if \((x_i, y_j)\) is a point in the boundary, otherwise \(u_{i,j} = \infty\);

**while not converges do**

**for each of the four sweeping directions do**

if \(u_{i,j} \neq 0\) then

- update \(u_{i,j}\) according to the local solution to the corresponding update formula associated to the numerical discretization;

end

end

end
Algorithm 2: The first attempt to solve the surface eikonal equation $|\nabla \Sigma u| = 1$ using the Godunov fast sweeping method.

**Data:** The source location $(x_s, y_s)$, the mesh size $\Delta x$, the level set representation of the surface $\phi_{i,j}$ and the tube radius $h = O(\Delta x)$.

**Result:** $u_{i,j}$ in the computational tube.

Initialization: Set $u_{i,j} = 0$ if $(x_i, y_j)$ is at the point source, otherwise $u_{i,j} = \infty$;

while not converges do
  for each of the four sweeping directions do
    if $|\phi_{i,j}| \leq h$ and $u_{i,j} \neq 0$ then
      update $u_{i,j}$ using (3);
    end
  end
end

$|\nabla u(x)| = 1, \; x \in \Omega \setminus P$

$u(x_s) = 0, \; x \in P$

Godunov numerical Hamiltonian

$$u_{i,j} \left\{ \begin{array}{ll} \min(u_{\min}^x, u_{\min}^y) + \Delta x & \text{if } |u_{\min}^x - u_{\min}^y| > \Delta x \\ \frac{1}{2} \left[ u_{\min}^x + u_{\min}^y + \sqrt{2\Delta x^2 - (u_{\min}^x - u_{\min}^y)^2} \right] & \text{otherwise} \end{array} \right.$$
Exact solution $\sqrt{2}$

\[
\frac{h}{\Delta x} = \frac{2.5}{\sqrt{2}}
\]

\[
\frac{h}{\Delta x} = \frac{3.5}{\sqrt{2}}
\]

$\Delta x \to 0$

$h = O(\Delta x^\gamma)$

Characteristic direction

Numerical upwind direction
The Geodesic Distance

\[(p, q) \cdot (u_x, u_y) = 0\]

Characteristic direction

\[|\nabla u(x)| = 1\]

\[(p, q) \cdot (u_x, u_y) = 0\]

\[
u_{i,j} \leftarrow \frac{p^+ u_{i-1,j} - p^- u_{i+1,j} + q^+ u_{i,j-1} - q^- u_{i,j+1}}{|p| + |q|}
\]

\[|\nabla_{\Sigma} u(x)| = 1, \quad x \in \Sigma \setminus P\]

\[u(x_s) = 0, \quad x \in P\]

\[|\nabla u(x)| = 1, \quad x \in \Omega \setminus P\]

\[u(x_s) = 0, \quad x \in P\]
II. Approximating the Principal Direction

$p_{\log}$: tangent plane PCA based on log-map

$p_{\text{orth}}$: tangent plane PCA based on orthogonal projection

$p$: Euclidean PCA

\[ \sum_{x_i \in S_x} (x_i - \mathbf{x})(x_i - \mathbf{x})^T \]
Theorem 4.1. Given $x, x_1, x_2, ..., x_n \in \mathcal{M} \subset \mathbb{R}^3$ for $n > 0$ and denote $y_i = (x_i - x)$. Let $n$ be the unit normal of the tangent plane at $x$, and $t_i$ be the unit tangent component of $y_i$ on the tangent plane. If $y_i \cdot t_i > \sqrt{2} (y_i \cdot n)$ for all data points, we have $p = p_{ort}$. 

$p_{ort}$: tangent plane PCA based on orthogonal projection

$p$: standard Euclidean PCA

$$\sum_{x_i \in \mathcal{S}_x} (x_i - x)(x_i - x)^T$$
III. Regularizations

- Reinitialization

\[ \phi_t + \text{sgn}(\phi^n)(\|\nabla \phi\| - 1) = 0 \]

- Orthogonalization

\[ \phi_t + \text{sgn}(\phi^n) \nabla \psi \cdot \nabla \phi = 0 \]

First order Hamilton-Jacobi equations: TVDRK-WENO
Examples
... with a topological change
\[ E(\Gamma) = \left[ \int_{\Gamma} d^p (x(s)) \, ds \right]^{\frac{1}{p}} + \lambda \left[ \int_{\Gamma} |n(x(s)) \cdot p(x(s))|^2 \, ds \right]^{\frac{1}{2}} \]
Reconstructing the tectonic plates from earthquake epicenter data
Conclusions

• An energy minimization algorithm
• Limitations
  • computational efficiency
  • dimensionality
• Unsupervised algorithms → manifold learning

Reference

Thank You