Nonlinear Mean Shift over Riemannian Manifolds

Paper by Raghav Subbarao and Peter Meer

Presentation by James Buenfil

Outline

1. What is the mean shift algorithm?
2. This paper: on a known manifold
3. Geometric preliminaries
4. Nonlinear mean shift
5. Theoretical properties
6. Relationship to other mean shift algorithms and related literature

Mean shift algorithm: Euclidian space

Let $\mathbf{x}_i \in \mathbb{R}^d, i = 1, \ldots, n$ be n independent, identically distributed points generated by an unknown probability distribution f. The *kernel density estimate*

$$\hat{f}_k(\mathbf{y}) = \frac{c_{k,h}}{n} \sum_{i=1}^{n} k\left(\frac{\|\mathbf{y} - \mathbf{x}_i\|^2}{h^2}\right) \tag{24}$$

based on a *profile function* k satisfying $k(z) \geq 0$ for $z \geq 0$, is a nonparametric estimator of the density $f(\mathbf{y})$ at $\mathbf{y}$. The constant $c_{k,h}$ is chosen to ensure that $\hat{f}_k$ integrates to one.

Define $g(\cdot) = -k'(\cdot)$. Taking the gradient of (24) we get

$$\mathbf{m}_h(\mathbf{y}) = C\frac{\nabla \hat{f}_k(\mathbf{y})}{\hat{f}_g(\mathbf{y})} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\|\mathbf{y}-\mathbf{x}_i\|^2/h^2\right)}{\sum_{i=1}^{n} g\left(\|\mathbf{y}-\mathbf{x}_i\|^2/h^2\right)} - \mathbf{y} \quad (25)$$

where, C is a positive constant and $\mathbf{m}_h(\mathbf{x})$ is the *mean shift* vector. The expression (25) shows that the mean shift vector is proportional to a normalized density gradient estimate.

The iteration

$$\mathbf{y}_{j+1} = \mathbf{m}_h(\mathbf{y}_j) + \mathbf{y}_j \quad (26)$$

is a gradient ascent technique converging to a stationary point of the density. Saddle points can be detected and removed, to obtain only the modes.

Properties of mean shift

Theorem 1. *If the kernel K has a convex and monotonically decreasing profile, the sequences $\{\mathbf{y}_j\}_{j=1,2...}$ and $\{\hat{f}_{h,K}(j)\}_{j=1,2...}$ converge and $\{\hat{f}_{h,K}(j)\}_{j=1,2...}$ is monotonically increasing.*

Finally, in Carreira-Perpinan (2007), it was shown that for Gaussian kernels, the mean shift step is the same as Expectation-Maximization. In the M-step of the EM-algorithm a tight lower bound on the function is computed and in the E-step this bound is maximized. For non-Gaussian kernels, mean shift is equivalent to generalized EM.

I mainly focus on application to clustering here

This paper: Assume known manifold

- The setting here is not Euclidian space, but a Riemannian manifold with a metric g.
- Briefly, the metric defines the inner products between tangent vectors which lie in the same tangent spaces.
- In many cases, we may need to specify which metric a manifold has.
- The choice of metric changes what the shortest distances between points are

What does knowing the (Riemannian) manifold give us?

- Tangent spaces
- Exponential map
- Logarithmic map
- Geodesic curves
- Geodesic distances
- Metric on tangent spaces

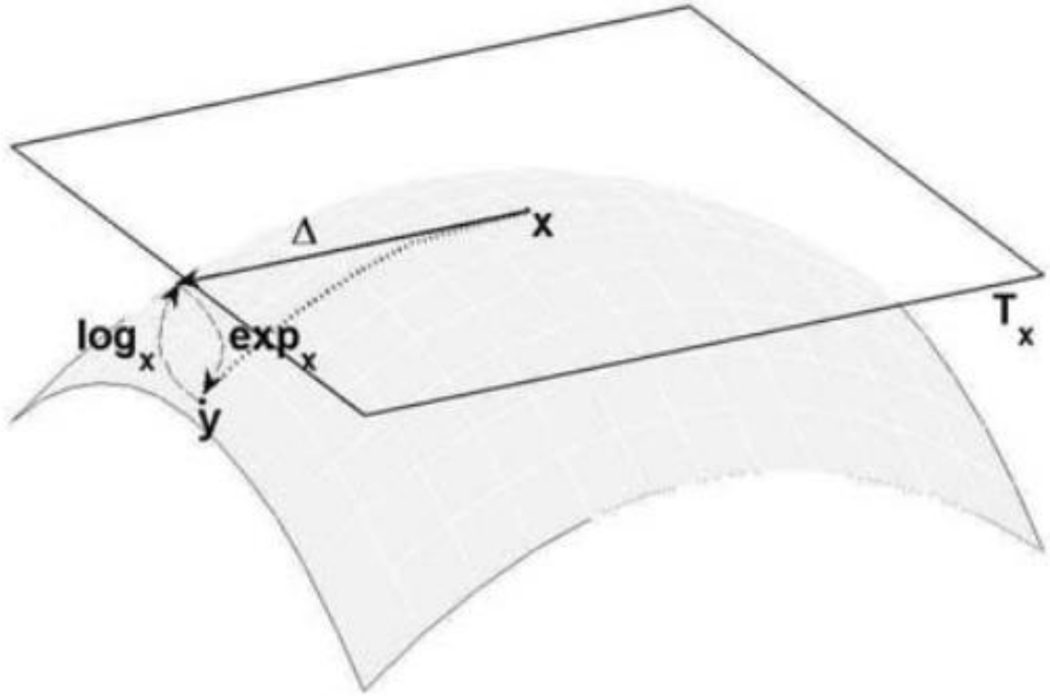

Fig. 2 Example of a two-dimensional manifold and the tangent space at the point **x**

Geometric Preliminaries

- Tangent spaces: Each point $x \in \mathcal{M}$ defines a vector space $T_x\mathcal{M}$
- Metric: For $v, w \in T_x\mathcal{M}$ we have an inner product $g(v, w) \in \mathbb{R}$
- Distance: The metric g allows us to define distances of curves c :$(0, 1) \rightarrow \mathcal{M}$
- Geodesic distance: We can then define the shortest distance between two points x and y on the manifold, denoted $d(x, y)$
- Exponential map: $exp_x(v) : T_x\mathcal{M} \rightarrow \mathcal{M}$ maps tangent vectors to the point on manifold which is the endpoint of the unique geodesic γ starting at x with initial velocity v: $exp_x(v) = \gamma(1)$
- Logarithmic map: $log_x(y) : \mathcal{M} \rightarrow T_x\mathcal{M}$ is the inverse of exp at x

Nonlinear Mean shift

Consider a Riemannian manifold with a metric d. Given n points on the manifold, $\mathbf{x}_i, i = 1, \ldots, n$, the kernel density estimate with profile k and bandwidth h is

$$\hat{f}_k(\mathbf{y}) = \frac{c_{k,h}}{n} \sum_{i=1}^{n} k\left(\frac{d^2(\mathbf{y}, \mathbf{x}_i)}{h^2}\right).$$

Calculating the gradient of $\hat{f}_k$ at $\mathbf{y}$, we get

$$\begin{aligned}
\nabla \hat{f}_k(\mathbf{y}) &= \frac{1}{n} \sum_{i=1}^{n} \nabla k\left(\frac{d^2(\mathbf{y}, \mathbf{x}_i)}{h^2}\right) \\
&= -\frac{1}{n} \sum_{i=1}^{n} g\left(\frac{d^2(\mathbf{y}, \mathbf{x}_i)}{h^2}\right) \frac{\nabla d^2(\mathbf{y}, \mathbf{x}_i)}{h^2} \\
&= \frac{2}{n} \sum_{i=1}^{n} g\left(\frac{d^2(\mathbf{y}, \mathbf{x}_i)}{h^2}\right) \frac{\log_{\mathbf{y}}(\mathbf{x}_i)}{h^2}
\end{aligned}$$

where, $g(\cdot) = -k'(\cdot)$, and in the final step we use (10).

- Important note: the KDE is only defined for points on the manifold

Theorem just used

Theorem 1 *The gradient of the Riemann squared distance is given by*

$$\nabla f(\mathbf{x}) = \nabla_{\mathbf{x}} d^2(\mathbf{x}, \mathbf{y}) = -2 \log_{\mathbf{x}}(\mathbf{y}). \quad (10)$$

Nonlinear mean shift algorithm

$$\mathbf{m}_h(\mathbf{y}) = \frac{\sum_{i=1}^{n} g\left(\frac{d^2(\mathbf{y},\mathbf{x}_i)}{h^2}\right) \log_{\mathbf{y}}(\mathbf{x}_i)}{\sum_{i=1}^{n} g\left(\frac{d^2(\mathbf{y},\mathbf{x}_i)}{h^2}\right)}.$$

$$\mathbf{y}_{j+1} = \exp_{\mathbf{y}_j}\left(\mathbf{m}_h(\mathbf{y}_j)\right).$$

Nonlinear mean shift algorithm

MEAN SHIFT OVER RIEMANNIAN MANIFOLDS

Given: Points on a manifold $\mathbf{x}_i, i = 1, \ldots, n$

for $i \leftarrow 1 \ldots n$

$\mathbf{y} \leftarrow \mathbf{x}_i$

repeat

$$\mathbf{m}_h(\mathbf{y}) \leftarrow \frac{\sum_{i=1}^{n} g\left(d^2(\mathbf{y}, \mathbf{x}_i)/h^2\right) log_{\mathbf{y}}(\mathbf{x}_i)}{\sum_{i=1}^{n} g\left(d^2(\mathbf{y}, \mathbf{x}_i)/h^2\right)}$$

$\mathbf{y} \leftarrow exp_{\mathbf{y}}\left(\mathbf{m}_h(\mathbf{y})\right)$

until $\|\mathbf{m}_h(\mathbf{y})\| < \epsilon$

Retain $\mathbf{y}$ as a local mode

Report distinct local modes.

Proven properties of nonlinear mean shift

Theorem 2 *If the kernel K has a convex and monotonically decreasing profile and the bandwidth h is less than the injectivity radius $i(\mathcal{M})$ of the manifold, the sequence $\{f(\mathbf{y}_j)\}_{j=1,2,\ldots}$ is convergent and monotonically non-decreasing.*

The authors state that the proof that this algorithm is equivalent to the EM algorithm is essentially the same as the one in Carreira-Perpinan (2007) (for Lie groups).

Background of development of nonlinear mean shift

- Mean shift originally developed by Fukunaga, K., & Hostetler, L. D. (1975)
- Popularized by Comaniciu and Meer (2002)
- A mean shift algorithm for Lie groups was developed by Tuzel, Subbarao, and Meer (2005)
- Nonlinear Mean Shift for Clustering over Analytic Manifolds by Subbarao and Meer (2006) introduced the algorithm presented here, which generalized the Lie group version

The main contribution of this paper is to derive theoretical properties, as well as to apply it to Motion Segmentation and Discontinuity Preserving Filtering. They also provide the numerical details for several specific manifolds.

Will refer to the algorithm presented here as RMS (Riemannian mean shift).

Work which followed

- Intrinsic Mean Shift for Clustering on Stiefel and Grassmann Manifolds by Cetingul and Vidal (2009) refer to the mean shift algorithm developed here as the extrinsic mean shift algorithm, and develop an "intrinsic" one which avoids the computing the exponential map at every iteration, but only on the Stiegel and Grassmann manifolds

- In Clustering via Mode Seeking by Direct Estimation of the Gradient of a Log-Density by Sasaki et al. (2014), the idea is that just because you have a good density estimate does not mean that you have a good estimate of the gradient. They model the gradient directly. (Euclidian mean shift)

- In Least-Squares Log-Density Gradient Clustering for Riemannian Manifolds by Ashizawa et al. (2017), the approach of Sasaki et al. (2014) is adapted the Riemannian manifold setting and is shown empirically to have better performance than the RMS. No theoretical analysis performed

Work which followed cont.

- In Mode estimation on matrix manifolds: Convergence and robustness by Sasaki et al. (2022), they follow the approach of Ashizawa et al. (2017) but avoid needing to compute the exponential and logarithmic maps at every iteration, only on 4 matrix manifolds (Stiefel, oblique, Grassmann manifolds and the set of symmetric positive definite matrices). (Euclidian metric)
- Yikun and Yen-chi has 4 papers where the setting is the directional mean shift! (the manifold is the hyper-sphere)

Thanks for listening!