Mixed dimensions estimation and clustering in high dimensional noisy point clouds

Gloria Haro°, Gregory Randall* and Guillermo Sapiro◇

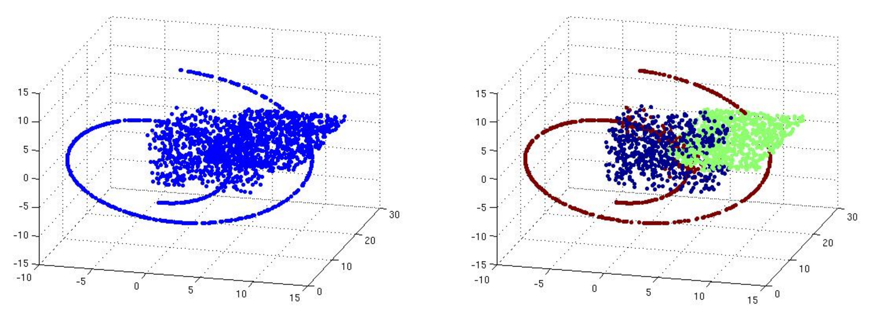
°Universitat Politècnica de Catalunya (Spain), *Universidad de la Republica (Uruguay), ◇University of Minnesota (USA)

Abstract

A framework for the regularized and robust estimation of non-uniform dimensionality and density in high dimensional noisy data is introduced in this work. This leads to learning stratifications, that is, mixture of manifolds representing different characteristics and complexities in the data set. The basic idea relies on modeling the high dimensional sample points as a process of Translated Poisson mixtures, with regularizing restrictions, leading to a model which incorporates the modeling of noise.

Introduction

Motivation: Detect and estimate different dimensions/complexities and densities in the same **noisy** point cloud data and cluster the points according to the dimensionality and density of the underlying possible multiple manifolds.



Local dimension estimation

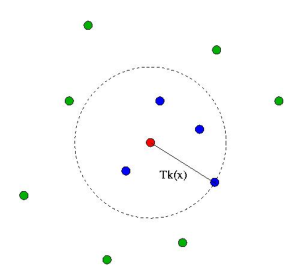
Levina and Bickel's approach [4] (Takens' estimator [6]).

Basic idea: proportion of points falling into a ball.

$$\frac{k}{n} \approx f(x)V(m)R_k(x)^m$$

where:

- k: number of points inside ball.
- n: total number of points.
- $f(x)$: local density at point x.
- $V(m)$: volume of the unit sphere in $\mathbb{R}^m$.
- $R_k(x)$: Euclidean distance from x to its k-th nearest neighbor.



Observable event: $N(R, x)$, Number of points falling into a small sphere $B(R, x)$ (radius R, centered at x).

Making the **approximations**:

- Binomial process by a **Poisson process** ($n \to \infty$, k moderate, and $k/n \to 0$).
- $f(x) \approx const.$ in a small sphere.

then, the rate λ of the counting process N

$$\lambda(r, x) = f(x)V(m)mr^{m-1}$$

Maximize the Log-likelihood of the observed process $N(R, x)$.

Local dimension estimator:

$$m(x) = \left[\frac{1}{k-2}\sum_{j=1}^{k-1}\log\frac{R_k(x)}{R_j(x)}\right]^{-1}$$

Translated Poisson Model

Theorem (Snyder & Miller [5]). *Let $\{N(A) : A \subseteq X\}$ be a Poisson process with an integrable intensity function $\{\lambda(x) : x \in X\}$. Points of this input point process are translated to the output space Z to form the output point process $\{M(B) : B \subseteq Z\}$, where each point is independently translated according to the* **transition density** *$f(z|x)$. Then, if there are no insertions and deletions, $\{M(B) : B \subseteq Z\}$ is a Poisson process with intensity*

$$\mu(z) = \int_X f(z|x)\lambda(x)dx.$$

$\lambda(r, x)$ is parametrized by the Euclidean distances r of the points. We consider a random translation $f(s|r)$ in the distances. The intensity of the Poisson process in the output space, is given by

$$\mu(s, x_t) = \int_0^{R'} f(s|r)\lambda(r, x_t)dr.$$

Maximize the Likelihood of the new Translated Poisson process.

$$m(x_t) = \left[\frac{1}{k-1}\sum_{i=1}^{k-1}\frac{\int_0^{R'} f(R_i(x_t)|r)r^{m-1}\log\frac{R_k(x_t)}{r}dr}{\int_0^{R'} f(R_i(x_t)|r)r^{m-1}dr}\right]^{-1}$$

We substitute r^{m-1} by its Taylor expansion around R_i.

Local dimension estimator (noisy case):

$$m(x_t) \approx \left[\frac{1}{k-1}\sum_{i=1}^{k-1}\frac{\int_0^{R'} f(R_i|r)\log\frac{R_k}{r}dr}{\int_0^{R'} f(R_i|r)dr}\right]^{-1}$$

If $f(s|r) = \delta(s - r)$ (no noise) $\to$ Levina and Bickel estimator.

Translated Poisson Mixture Model

Consider J **mixture components**:
vector of parameters $\psi = \{\psi^j = (\pi^j, \theta^j, m^j); j = 1, \ldots, J\}$ where

- π^j is the mixture coefficient for class j,
- θ^j is the density parameter ($f^j = e^{\theta^j}$)
- m^j is the dimension.

Observable event: $y = N(R, x)$, # points inside ball $B(R, x)$.
Observation sequence: $Y = \{y_t; t = 1, \ldots, T\}$.

Density function:

$$p(y_t|\psi) = \sum_{j=1}^{J}\pi^j p(y_t|m^j, \theta^j)$$

Hidden-state information: $Z = \{z_t \in C; t = 1 \ldots T\}$, where $z_t = C^j$ means that the j-th mixture generates y_t.

Regularized TPMM [1, 3]:

$$F(\psi, H) = L(Y|\psi, H) + \alpha S(H)$$

where $H = \{h^j(y_t) \leq 1; t = 1, \ldots, T, j = 1, \ldots, J\}$, $h^j(y_t)$ is the probability that observation t belongs to mixture j and

$$S(H) = -\sum_{t=1}^{T}\sum_{j=1}^{J} h^j(y_t)\mathcal{D}(t, j, X, H)$$

As **dissimilarity measure** we use $\mathcal{D}_R := \sum_{l\sim t}(1 - h^j(y_l))^2$

Algorithm R-TPMM

1. Compute the **local estimators** $m(x_t)$ and $f(x_t)$.
2. **Initialization** of $\psi_0^j = \{\pi_0^j, m_0^j, \theta_0^j\}$ for all $j = 1, \ldots, J$.
3. **Iterations** (until convergence of ψ_n^j):
   For each class $j = 1, \ldots, J$,
   - Compute $h_n^j(y_t)$.

$$h_{n+1}^j(y_t) = \frac{p(y_t|m_n^j, \theta_n^j)\pi_n^j e^{-\alpha\mathcal{D}(t,j,X,H_n)}}{\sum_{l=1}^{J} p(y_t|m_n^l, \theta_n^l)\pi_n^l e^{-\alpha\mathcal{D}(t,l,X,H_n)}}$$

   - Compute ψ_{n+1}^j.

$$m_{n+1}^j = \left[\frac{\sum_{t=1}^{T} h_n^j(y_t)m(x_t)^{-1}}{\sum_{t=1}^{T} h_n^j(y_t)}\right]^{-1}$$

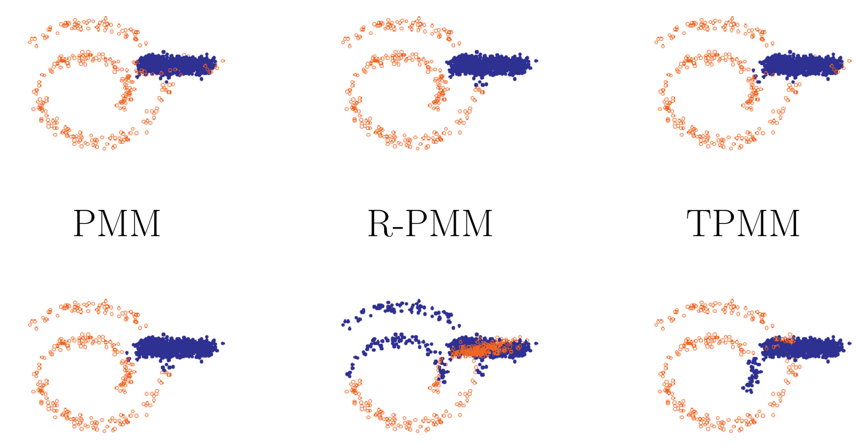
$$f_{n+1}^j = \left[\frac{\sum_{t=1}^{T} h_n^j(y_t)f(x_t)^{-1}}{\sum_{t=1}^{T} h_n^j(y_t)}\right]^{-1} \qquad \pi_{n+1}^j = \frac{1}{T}\sum_{t=1}^{T} h_n^j(y_t)$$

Notation:

	$\sigma = 0$	$\sigma > 0$
$\alpha = 0$	PMM [2]	TPMM
$\alpha > 0$	R-PMM [3]	R-TPMM

Experiments

Synthetic Data 1

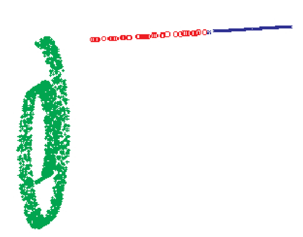


Clustering of a spiral and a plane with noise. Results with different algorithms.

	PMM		R-PMM		TPMM		R-TPMM	
	Estimated parameters							
m	2.47	1.51	2.48	1.43	1.86	1.35	1.87	1.32
θ	0.13	0.03	0.15	0.03	0.87	0.34	0.83	0.40
	Points in each class							
Pl.	764	36	800	0	784	16	800	0
Sp.	22	278	25	275	27	273	29	271

Estimated parameters and clustering results of a spiral and a plane with noise ($k = 40$, $J = 2$).

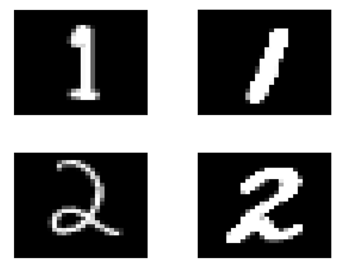
Synthetic Data 2



m: 1.98, 1.02, 0.99.
θ: 0.49, 0.53, 6.89.

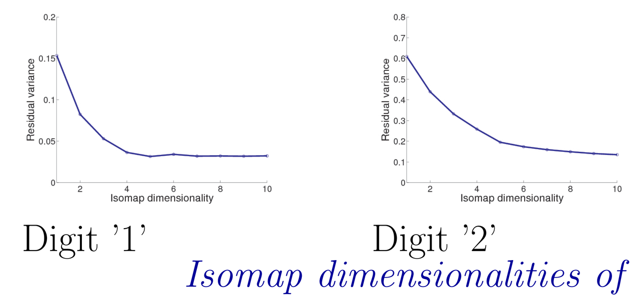
Clustering of a Swiss roll and a line with two different densities with R-TPMM, $k = 20$, $J = 3$, and $\alpha = 2$.

MNIST database of handwritten digits



	PMM		R-TPMM	
	Estimated parameters			
m	7.33	12.79	2.88	7.24
θ	-7.38	-23.99	-1.62	-12.90
	Points in each class			
'1'	1032	0	1029	3
'2'	70	1065	17	1118

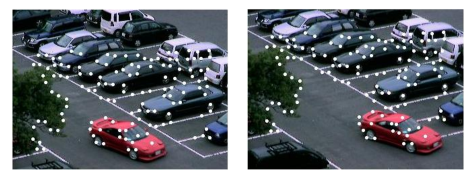
Results for digits 1 and 2 ($k = 30$, $J = 2$, $\alpha = 2$, $\sigma = 1.5$).



The graph show the residual variance of the first ten Isomap embedding dimensionalities.

Isomap dimensionalities of digits '1' and '2'.

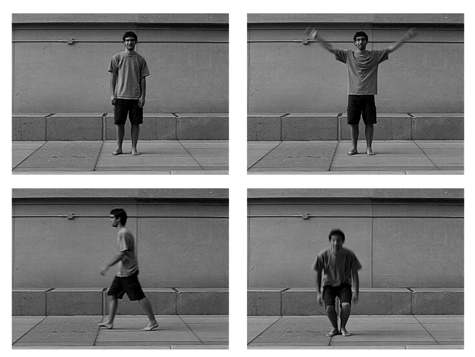
Motion segmentation



Method	Effectiveness
Costeira-Kanade	60.3%
Ichimura	92.6%
Kanatani-Sugaya	100%
Souvenir-Pless	93.38%
GPCA	100%
R-TPMM	100%

Classification rates, using different methods, for the motion segmentation in the Kanatani Laboratory sequence. R-TPMM with $k = 10$, $\alpha = 2$ and $\sigma = 0.05$.

Activities in video



	Samples in each cluster			
	C1	C2	C3	C4
Standing	505	0	6	0
Walking	0	464	45	14
Waving	1	1	430	0
Jumping	0	0	0	207

Classifying human activities in video with the R-TPMM algorithm ($k = 10$, $J = 4$, $\alpha = 40$, $\sigma = 0.25$). We use the 6 previous and 6 posterior frames as neighbors in $\mathcal{D}_R$, which results in a temporal regularization. The global classification is 96% accurate.

References

[1] C. Ambroise and G. Govaert. Convergence of an EM-type algorithm for spatial clustering. *Pattern Recognition Letters*, 19(10):919–927, 1998.

[2] G. Haro, G. Randall, and G. Sapiro. Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds. In *Advances in NIPS 19, Vancouver, Canada*, 2006.

[3] G. Haro, G. Randall, and G. Sapiro. Regularized mixed dimensionality and density learning in computer vision. In *Proceedings of 1st Workshop on Component Analysis Methods for C.C.M. and E.P. in C.V., in conjunction with CVPR*, 2007.

[4] E. Levina and P.J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in NIPS 17, Vancouver, Canada*, 2005.

[5] D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space*. Springer-Verlag, 1991.

[6] F. Takens. On the numerical determination of the dimension of an attractor. *Lecture notes in mathematics. Dynamical systems and bifurcations*, 1125:99–106, 1985.