Structure-Adaptive Manifold Estimation

Zhenman Yuan

Department of Statistics
University of Washington

Table of contents

Motivation, Problem Setup, Background

Manifold Estimation

We consider this problem: recover a low dimension manifold from a cloud of noise observation $\mathbb{X}_n = (X_1, ..., X_n)$ in high dimensional space. Several quantities are of great interest:

1) Manifold itself $\mathcal{M}^*$
2) Denoising estimates $\hat{X}_1, ..., \hat{X}_n$
3) Tangent Space Π_x for every $x \in \mathcal{M}^*$

Previous Work

1. Aamari and Levrard[2018, 2019], Maggioni et al. [2016]:
   The model is noise-free or with noise of small magnitude, shrinks to zero as the sample size n tends to infinity.
2. Genovese et al. [2012a]:
   The noise has a uniform distribution and in the direction orthogonal to the manifold tangent space.
3. Fefferman et al. [2018]:
   The noise is Gaussian noise, but the magnitude could not exceed the reach $\varkappa$ of the manifold.
4. ....

In summary, there are two cases that are well studied:

- The noise is totally unknown, but extremely small noise magnitude;
- The noise is large, but with completely known distribution.

The problem is, these assumptions are too restrictive and unlikely to hold in practice. Could we solve the problem of manifold recovering under weaker and more realistic assumptions on the noise?

Model and Notations

Model:

Given an i.i.d smaple $\mathbb{Y}_n = (Y_1, ..., Y_n)$, where Y_i are independent copies of a random vector Y in $\mathbb{R}^D$, generated from the model

$$Y = X + \epsilon$$

Here X is a random element whose distribution is supported on a low-dim manifold $\mathcal{M}^* \subset \mathbb{R}^D, \dim(\mathcal{M}^*) = d < D$, and ϵ is a full dimensional noise.

Notations:

1. Hausdorff distance $d_H(\cdot, \cdot)$:

$$d_H(\mathcal{M}_1, \mathcal{M}_2) = \inf\{\epsilon > 0 : \mathcal{M}_1 \subseteq \mathcal{M}_2 \oplus \mathcal{B}(0, \epsilon), \mathcal{M}_2 \subseteq \mathcal{M}_1 \oplus \mathcal{B}(0, \epsilon)\}$$

Model and Notations

Model:

Given an i.i.d smaple $\mathbb{Y}_n = (Y_1, ..., Y_n)$, where Y_i are independent copies of a random vector Y in $\mathbb{R}^D$, generated from the model

$$Y = X + \epsilon$$

Here X is a random element whose distribution is supported on a low-dim manifold $\mathcal{M}^* \subset \mathbb{R}^D, \dim(\mathcal{M}^*) = d < D$, and ϵ is a full dimensional noise.

Notations:

1. Hausdorff distance $d_H(\cdot, \cdot)$:

   $$d_H(\mathcal{M}_1, \mathcal{M}_2) = \inf\{\epsilon > 0 : \mathcal{M}_1 \subseteq \mathcal{M}_2 \oplus \mathcal{B}(0, \epsilon), \mathcal{M}_2 \subseteq \mathcal{M}_1 \oplus \mathcal{B}(0, \epsilon)\}$$

2. M: Magnitude of the noise

Model and Notations

Model:

Given an i.i.d smaple $\mathbb{Y}_n = (Y_1, ..., Y_n)$, where Y_i are independent copies of a random vector Y in $\mathbb{R}^D$, generated from the model

$$Y = X + \epsilon$$

Here X is a random element whose distribution is supported on a low-dim manifold $\mathcal{M}^* \subset \mathbb{R}^D, \dim(\mathcal{M}^*) = d < D$, and ϵ is a full dimensional noise.

Notations:

1. Hausdorff distance $d_H(\cdot, \cdot)$:

$$d_H(\mathcal{M}_1, \mathcal{M}_2) = \inf\{\epsilon > 0 : \mathcal{M}_1 \subseteq \mathcal{M}_2 \oplus \mathcal{B}(0, \epsilon), \mathcal{M}_2 \subseteq \mathcal{M}_1 \oplus \mathcal{B}(0, \epsilon)\}$$

2. M: Magnitude of the noise
3. b: Maximal deviation of the noise in tangent direction.

Model and Notations

Model:
Given an i.i.d smaple $\mathbb{Y}_n = (Y_1, ..., Y_n)$, where Y_i are independent copies of a random vector Y in $\mathbb{R}^D$, generated from the model

$$Y = X + \epsilon$$

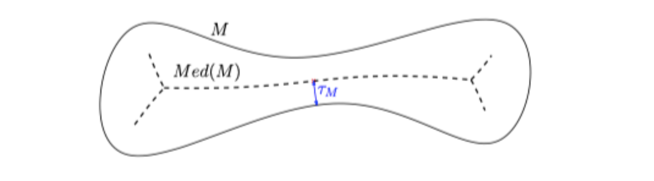
Here X is a random element whose distribution is supported on a low-dim manifold $\mathcal{M}^* \subset \mathbb{R}^D, \dim(\mathcal{M}^*) = d < D$, and ϵ is a full dimensional noise.

Notations:

1. Hausdorff distance $d_H(\cdot, \cdot)$:

$$d_H(\mathcal{M}_1, \mathcal{M}_2) = \inf\{\epsilon > 0 : \mathcal{M}_1 \subseteq \mathcal{M}_2 \oplus \mathcal{B}(0, \epsilon), \mathcal{M}_2 \subseteq \mathcal{M}_1 \oplus \mathcal{B}(0, \epsilon)\}$$

2. M: Magnitude of the noise
3. b: Maximal deviation of the noise in tangent direction.
4. $\varkappa$: The reach of the manifold $\mathcal{M}^*$.



labelformat=empty

For a manifold M, denote

$$\mathrm{Med}(M) = \{z \in \mathbb{R}^D | \exists\ p \neq q \in M, ||z - p|| = ||z - q|| = d(z, M)\}$$

The reach is defined by

$$\tau_M = \inf_{p \in M} d(p, \mathrm{Med}(M)) = \inf_{z \in \mathrm{Med}(M)} d(z, M)$$

Assumptions – A1

Regularity of the underlying manifold $\mathcal{M}^*$

$$\begin{aligned}\mathcal{M}^* \in \mathcal{M}^{\mathrm{d}}_{\varkappa} = \{&\mathcal{M} \subset \mathbb{R}^{\mathrm{D}} : \mathcal{M} \text{ is a compact, connected manifold} \\ &\text{without a boundary}, \mathcal{M} \in \mathcal{C}^2, \mathcal{M} \subseteq \mathcal{B}(0, \mathrm{R}), \\ &\text{reach}(\mathcal{M}) \geq \varkappa, \dim(\mathcal{M}) = \mathrm{d} < \mathrm{D}\}\end{aligned}$$

Assumption – A2

Density of X on the manifold $\mathcal{M}^*$

Denote $\mathrm{p}(\mathrm{x})$ to be the density of X:

$$\exists\, \mathrm{p}_1 \geq \mathrm{p}_0 > 0 : \forall \mathrm{x} \in \mathcal{M}^* \quad \mathrm{p}_0 \leq \mathrm{p}(\mathrm{x}) \leq \mathrm{p}_1,$$

$$\exists\, \mathrm{L} > 0 : \forall \mathrm{x}, \mathrm{x}^{'} \in \mathcal{M}^* \quad |\mathrm{p}(\mathrm{x}) - \mathrm{p}(\mathrm{x}^{'})| \leq \frac{\mathrm{L}||\mathrm{x} - \mathrm{x}^{'}||}{\varkappa}.$$

Assumption – A3

Noise Magnitude/Direction and Reach

There exist $0 \leq \mathrm{M} < \varkappa$ and $0 \leq \mathrm{b} \leq \varkappa$, such that

$$\mathbb{E}(\epsilon|\mathrm{X}) = 0,\ ||\epsilon|| \leq \mathrm{M} < \varkappa$$
$$||\Pi(\mathrm{X})\epsilon|| \leq \frac{\mathrm{Mb}}{\varkappa}\ \mathbb{P}(\cdot|\mathrm{X}) - \text{almost surely},$$

Assumption – A4

Upper bounds for pairs (M, b)

$$\begin{cases} M \leq A n^{-\frac{2}{3d+8}}, \\ M^3 b^2 \leq \alpha \varkappa \left[\left(\frac{D \log n}{n} \right)^{\frac{4}{d}} \vee \left(\frac{D M^2 \varkappa^2 \log n}{n} \right)^{\frac{4}{d+4}} \right] \end{cases}$$

where A and α are some positive constants.

Assumption – A4

There are two specific cases of interest:

Maximal Admissible Magnitude:

$$M = M(n) \leq A n^{-\frac{2}{3d+8}}$$

$$b = b(n) \leq \frac{\sqrt{\alpha \varkappa}}{A^{3/2}} \left[\left(\frac{D \log n}{n} \right)^{\frac{1}{d}} \vee \left(\frac{D M^2 \varkappa^2 \log n}{n} \right)^{\frac{1}{d+4}} \right]$$

Maximal Admissible Angle:

$$b = \varkappa, M = M(n) \leq \left(\frac{D^4 \alpha^{d+4}}{\varkappa^{d-4}} \right)^{\frac{1}{3d+4}} n^{-\frac{4}{3d+4}}$$

Algorithm and Theoretical Guarantee

Nadaraya-Watson estimator:

$$\hat{X}_i^{(NW)} = \frac{\sum_{j=1}^{n} \omega_{ij}^{(NW)} Y_j}{\sum_{j=1}^{n} \omega_{ij}^{(NW)}}$$

and $\omega_{ij}^{(NW)}$ are the smoothing weights defined by

$$\omega_{ij}^{(NW)} = \mathcal{K}\left(\frac{||Y_i - Y_j||^2}{h^2}\right), \quad 1 \leq i, j \leq n,$$

where $\mathcal{K}(\cdot)$ is a smoothing kernel (could be asymmetric) and the bandwidth $h = h(n)$ is the hyper-parameter of the kernel.

Algorithm – Structure-adaptive manifold estimator

SAME

1. Assume d is known. The initial guess $\hat{\Pi}_1^{(0)}, ..., \hat{\Pi}_n^{(0)}$ of $\Pi(X_1), ..., \Pi(X_n)$, the number of iterations $K+1$, an initial bandwidth h_0, the threshold τ and constant $a > 1$ and $\gamma > 0$ are given.
2. for k from 0 to K do
3. Compute the weights $\omega_{ij}^{(k)}$ according to the formula

$$\omega_{ij}^{(k)} = \mathcal{K}\left(\frac{||\hat{\Pi}_i^{(k)}(Y_i - Y_j)||^2}{h_k^2}\right)1(||Y_i - Y_j|| \le \tau), \;\; 1 \le i, j \le n.$$

4. Compute the Nadaraya-Watson estimates

$$\hat{X}_i^{(k)} = \sum_{j=1}^{n} \omega_{ij}^{(k)} Y_j / \left(\sum_{j=1}^{n} \omega_{ij}^{(k)}\right), \;\; 1 \le i \le n.$$

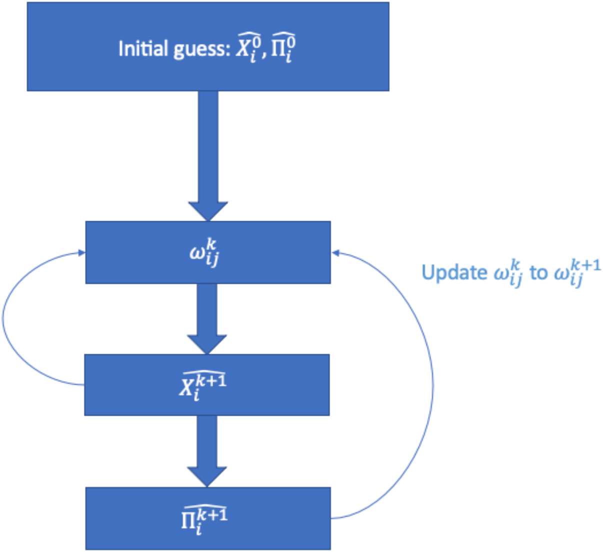
5. If $k < K$, for each i from 1 to n, define a set $\mathcal{J}_i^{(k)} = \{j : ||\hat{X}_j^{(k)} - \hat{X}_i^{(k)}|| \le \gamma h_k\}$ and compute the matrices

$$\hat{\Sigma}_i^{(k)} = \sum_{j \in \mathcal{J}_i^{(k)}} (\hat{X}_i^{(k)} - \hat{X}_j^{(k)})(\hat{X}_i^{(k)} - \hat{X}_j^{(k)})^T, \;\; 1 \le i \le n.$$

6. If $k < K$, for each i from 1 to n, define $\hat{\Pi}_i^{(k+1)}$ as a projector onto a linear span of eigenvectors of $\hat{\Sigma}_i^{(k)}$, corresponding to the largest d eigenvalues.
7. If $k < K$, set $h_{k+1} = h_k / a$.

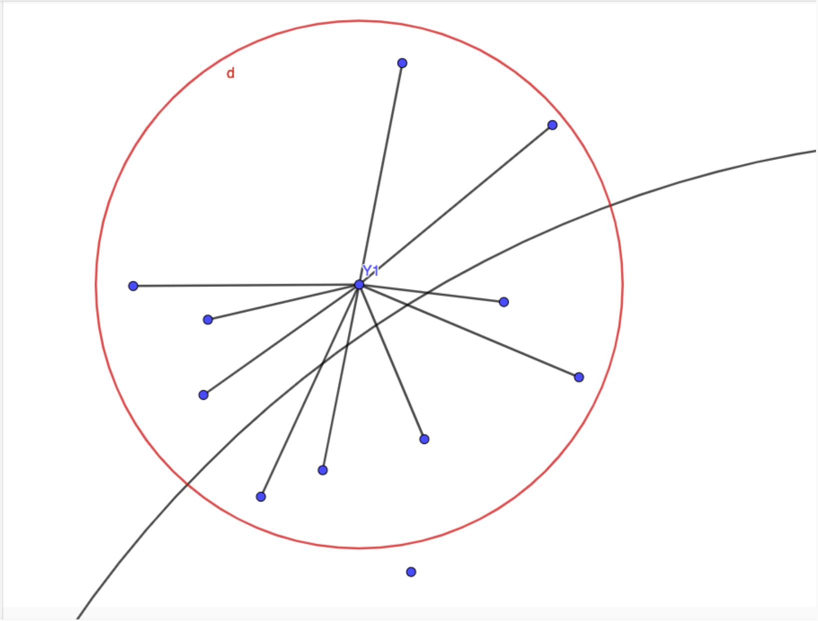
$\mathbf{return}$ the estimates $\hat{X}_1 = \hat{X}_1^{(K)}, ..., \mathrm{hat}X_n = \hat{X}_n^{(K)}$.

Algorithm

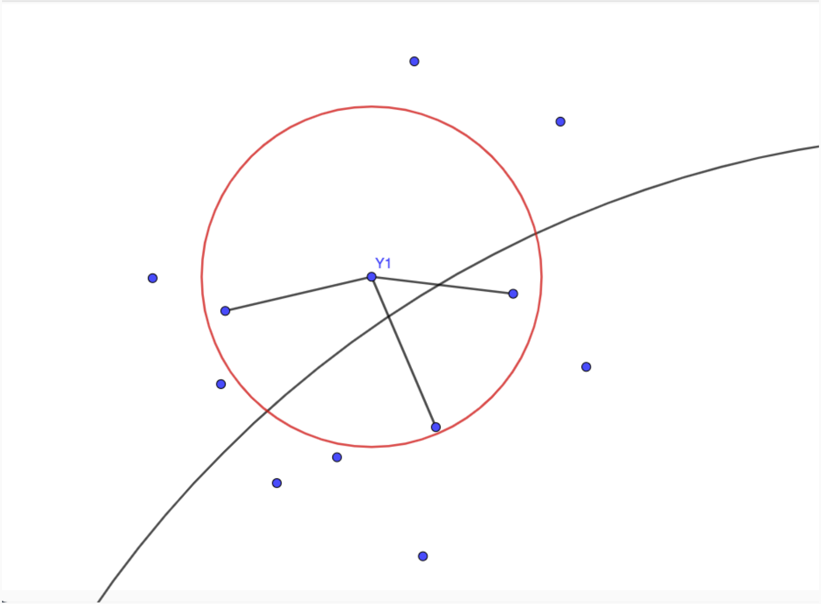


Initial guess: $\widehat{X_i^0}, \widehat{\Pi_i^0}$
ω_{ij}^k
$\widehat{X_i^{k+1}}$
$\widehat{\Pi_i^{k+1}}$
Update ω_{ij}^k to ω_{ij}^{k+1}

Algorithm



Algorithm



Theoretical Guarantee

Theorem 1

Assume A_1, S_2, A_3, A_4 hold Let the initial guesses $\hat{\Pi}_1^{(0)}, \dots, \hat{\Pi}_n^{(0)}$ of $\Pi(X_1), \dots, \Pi(X_n)$ be such that on an event with prob at least $1 - n^{-1}$, it holds

$$\max_{1 \le i \le n} \|\hat{\Pi}_i^{(0)} - \Pi(X_i)\| \le \frac{\Delta h_0}{\varkappa}$$

with a constant Δ, such that $\Delta h_0 \le \varkappa/4$, and $h_0 = C_0 / \log n$, where $C_0 > 0$ is an absolute constant. Choose $\tau = 2C_0/\sqrt{\log n}$ and set any $a \in (1, 2]$. If n is larger than a constant N_Δ, depending on Δ, then there exists a choice of γ, such that after K iterations SAME produces estimates $\hat{X}_1, \dots, \hat{X}_n$, such that, with prob at least $1 - \frac{5K+4}{n}$, it holds

$$\max_{1 \le i \le n} \|\hat{X}_i - X_i\| \lesssim \frac{Mb \vee Mh_K \vee h_K^2}{\varkappa} + \sqrt{\frac{D(h_K^2 \vee M^2)\log n}{n h_K^d}},$$

$$\max_{1 \le i \le n} \|\hat{\Pi}_i^{(K)} - \Pi(X_i)\| \lesssim \frac{h_K}{\varkappa} + h_K^{-1}\sqrt{\frac{D(h_K^2/\varkappa^2 \vee M^2)\log n}{n h_K^d}},$$

provided that $h_K \gtrsim ((D \log n/n)^{1/d} \vee (DM^2\varkappa^2 \log n/n)^{1/(d+4)})$ (with a sufficiently large hidden constant, which is greater than 1). In particular, if one choose the parameter a and the number of iterations K in such a way that $h_K \asymp ((D\varkappa^2 \log n/n)^{1/(d+2)} \vee (DM^2\varkappa^2 \log n/n)^{1/(d+4)})$ then

$$\max_{1 \le i \le n} \|\hat{X}_i - X_i\| \lesssim \frac{Mb}{\varkappa} + \frac{1}{\varkappa}(D\varkappa^2 \log n/n)^{2/(d+2)} \vee \frac{M}{\varkappa}(DM^2\varkappa^2 \log n/n)^{1/(d+4)}.$$

If $h_K \asymp ((D \log n/n)^{1/d} \vee (DM^2\varkappa^2 \log n/n)^{1/(d+4)})$ then

$$\max_{1 \le i \le n} \|\hat{\Pi}_i^{(K)} - \Pi(X_i)\| \lesssim \frac{1}{\varkappa}\left(\frac{D \log n}{n}\right)^{\frac{1}{d}} \vee \frac{1}{\varkappa}\left(\frac{DM^2\varkappa^2 \log n}{n}\right)^{\frac{1}{d+4}}$$

How to Choose Initial Guess?

The estimates suggested by Aamari and Levrard[2018] is:

1. For each i from 1 to n introduce
$$\hat{\Sigma}_i^{(0)} = \frac{1}{n-1} \sum_{j \neq i} (Y_j - \bar{X}_i)(Y_j - \bar{X}_i)^T 1(Y_j \in \mathcal{B}(Y_i, h_0))$$
2. Let $\hat{\Pi}_i^{(0)}$ be the projector onto the linear span of the d largest eigenvalues of $\hat{\Sigma}_i^{(0)}$.

Proposition 5.1 in Aamari and Levrard [2018]

Assume A_1, A_2, A_3 hold. Set $h_0 \gtrsim (\log n/n)^{1/d}$ for large enough hidden constant. Let $M/h_0 \leq \frac{1}{4}$ and let $h_0 = h_0(n) = o(1)$, as $n \to \infty$. Then for n large enough, with prob larger than $1 - n^{-1}$, it holds

$$\max_{1 \leq i \leq n} ||\hat{\Pi}_i^{(0)} - \Pi(X_i)|| \lesssim \frac{h_0}{\varkappa} + \frac{M}{h_0}.$$

Theorem 2

Assume A_1, A_2, A_3, A_4 hold. Consider the piecewise linear manifold estimate

$$\hat{\mathcal{M}} = \{\hat{X}_i + h_K \hat{\Pi}_i^{(K)} u : 1 \le i \le n, u \in \mathcal{B}(0,1) \subset \mathbb{R}^D\}$$

where $\hat{\Pi}_i^{(K)}$ is a projector onto d-dimensional space obtained on the K-th iteration of SAME. Then, as long as $h_K (\log n/n)^{1/d}$, on the event with prob at least $1 - \frac{5K+5}{n}$, it holds

$$d_H(\hat{\mathcal{M}}, \mathcal{M}^*) \lesssim \frac{h_K^2}{\varkappa} + \sqrt{\frac{D(h_K^4/\varkappa^2 \vee M^2) \log n}{n h_K^d}}$$

In particular, if a and K are chosen such that
$h_K \asymp ((D \log n/n)^{1/d} \vee (DM^2 \varkappa^2 \log n/n)^{1/(d+4)})$, then

$$d_H(\hat{\mathcal{M}}, \mathcal{M}^*) \lesssim \varkappa^{-1} (\frac{D \log n}{n})^{\frac{2}{d}} \vee \varkappa^{-1} (\frac{DM^2 \varkappa^2 \log n}{n})^{\frac{2}{d+4}}.$$

Theorem 3

Suppose that the sample $\mathbb{Y}_n = \{Y_1, ..., Y_n\}$, where $\mathcal{M}^* \in \mathcal{M}^d_\varkappa$, the density $q(x)$ of X fulfils assumption A_2 and the noise ϵ satisfies assumption A_3 with
$b \asymp ((D \log n/n)^{1/d} \vee (DM^2\varkappa^2 \log n/n)^{1/(d+4)})$. Then, if n is sufficiently large and $M\varkappa \gtrsim (\log n/n)^{2/d}$, it holds

$$\inf_{\hat{\mathcal{M}}} \sup_{\mathcal{M}^* \in \mathcal{M}^d_\varkappa} \mathbb{E}_{\mathcal{M}^*} d_H(\hat{\mathcal{M}}, \mathcal{M}^*) \gtrsim \frac{1}{\varkappa}\left(\frac{M^2\varkappa^2 \log n}{n}\right)^{\frac{2}{d+4}},$$

Simulation Studies

For all the following simulation study, we use $\hat{\Pi}_i^{(0)} = I$ and kernel $K(t) = e^{-t}$

$$n = 1500, M = 0.2, h_k = 0.6 \cdot 1.25^{-k}, 0 \leq k \leq 7, \tau = 0.9, \gamma = 4$$

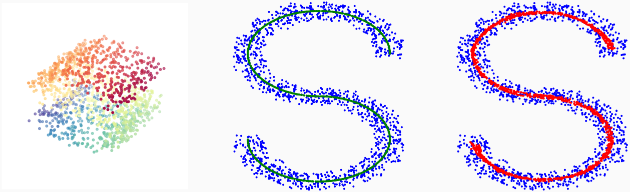


Figure 4: S-shaped curve

$$n = 2500, M = 1.25, h_k = 3.5 \cdot 1.25^{-k}, 0 \leq k \leq 3, \tau = 3.5, \gamma = 4$$

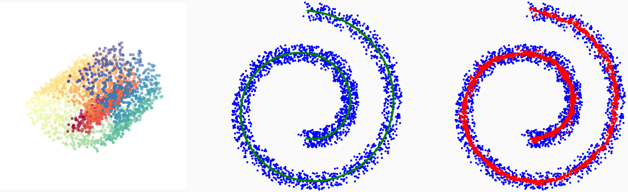


Figure 5: Swiss Roll

Noised Circle

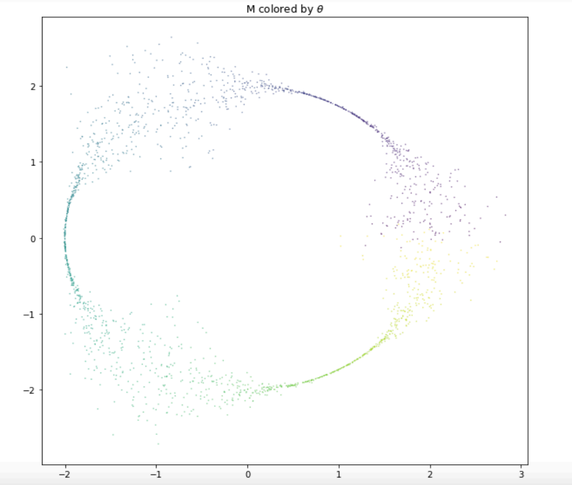
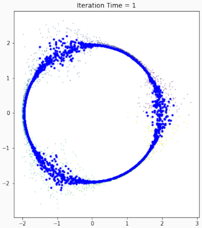


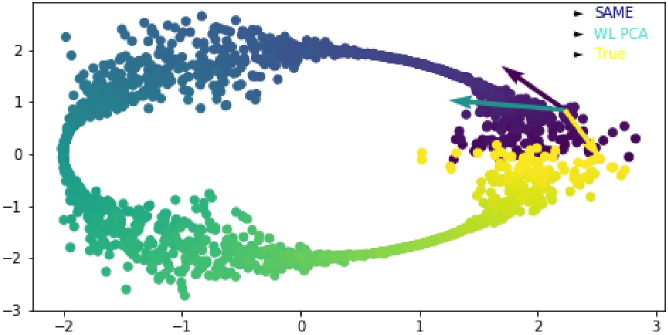
Figure 6: Noised Observations around Circle

Estimation of X_i

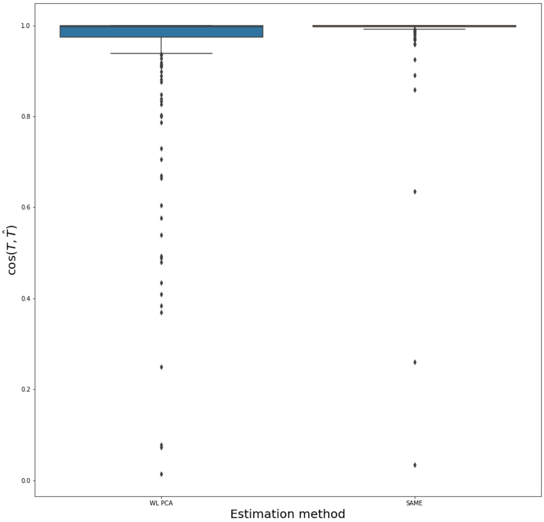
$$n = 2000, h_k = 0.6 \cdot 1.25^{-k}, \tau = 0.9, \text{gamma} = 4$$



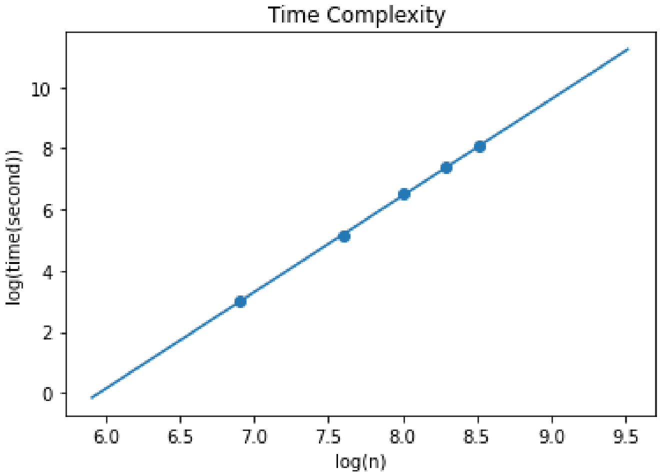
Estimation of Tangent space



Box Plot



Slope $s \approx 3.13 \Rightarrow$ Cost time $\approx \Theta(n^{3.13})$



Summary and Conclusion

Advantages and Challenges

- Advantage
    1. Concise and simple algorithm
- Challenge

Advantages and Challenges

- Advantage
    1. Concise and simple algorithm
    2. Asymptotic optimal if the assumptions are satisfied
- Challenge

Advantages and Challenges

- Advantage
  1. Concise and simple algorithm
  2. Asymptotic optimal if the assumptions are satisfied
- Challenge
  1. Polynomial time with respect to size n.

Advantages and Challenges

- Advantage
  1. Concise and simple algorithm
  2. Asymptotic optimal if the assumptions are satisfied
- Challenge
  1. Polynomial time with respect to size n.
  2. Take care of the chosen parameters, including $\hat{\Pi}^{(0)}$, h_0, τ and γ.

Thanks for listening!

Any Question?