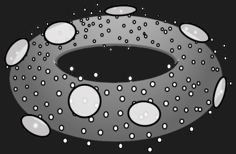


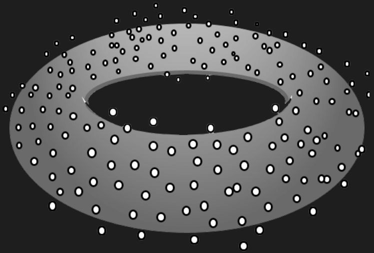
Tangent Space and Dimension Estimation with the Wasserstein Distance

Uzu Lim
with Harald Oberhauser and Vidit Nanda

University of Oxford

6 Sep 2022

Introduction



- Consider data given as points on $\mathbb{R}^D$.
- Good local parametrization $\implies$ Data lies on manifold.
- Assuming that such a manifold exists: **Manifold hypothesis**.
- Inferring properties of such manifold: **Manifold learning**.

Introduction

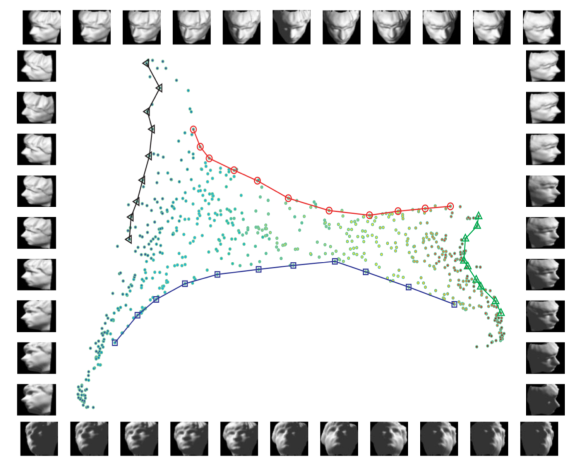
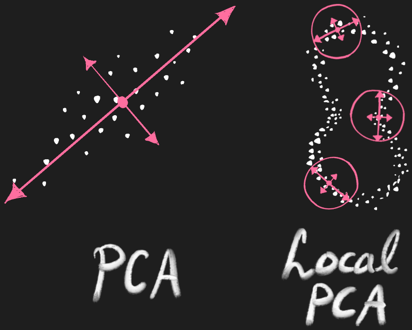


Figure: Synthetic images of face (embedded to $\mathbb{R}^{64^2}$ from 64×64 images), projected to $\mathbb{R}^2$ using the LTSA algorithm.

Introduction



- PCA (*principal component analysis*): Optimal linear regression
- Local PCA: Optimal local linear regression
- Local PCA on manifold $\rightarrow$ Tangent space & dimension

Introduction

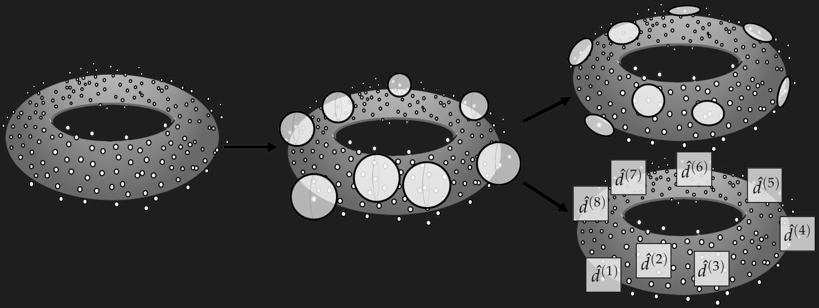


Figure: Local PCA estimates tangent spaces and intrinsic dimension(=2)

Introduction

- **Question**: How to quantify accuracy of estimating tangent space and intrinsic dimension with Local PCA?
- **Answer**: Use a matrix concentration inequality and a transportation plan.

PCA (principal component analysis)

- If $\underline{x} = \{x_1, \ldots x_m\} \subseteq \mathbb{R}^D$ then PCA is the diagonalization:

$$\Sigma[\delta_{\underline{x}}] = \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x})^\top = U\Lambda U^\top$$

  where $\bar{x} = \frac{1}{m}\sum_i x_i$, U is orthogonal and $\Lambda = \mathrm{diag}(\lambda_1, \cdots \lambda_D)$.
- Main interest: largest eigenvalues and the corresponding eigenvectors.

Local PCA - Tangent space

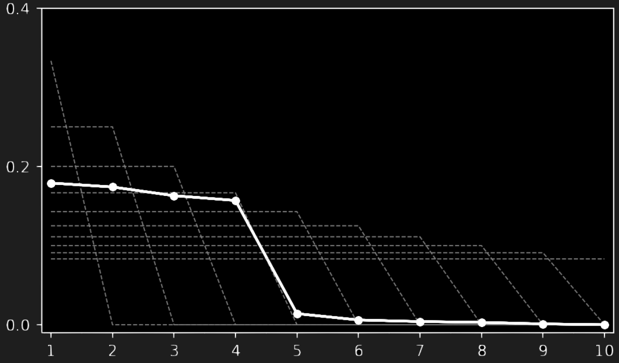
- If $y \in \mathbb{R}^D$ and $r > 0$, Local PCA performs PCA on:

$$\{x_1, \dots x_m\} \cap B_r(y)$$

- Let $M \subseteq \mathbb{R}^D$ be a compact smooth d-dim. submanifold. If $\underline{X} = (X_1, \cdots X_m)$ is drawn from the uniform distribution on M and tiny r, we should have:

$$\hat{T}_i := \pi_d[\underline{X}_i] \approx T_{X_i}M, \text{ where } \underline{X}_i = \underline{X} \cap B_r(X_i)$$

Local PCA - Dimension



$$\hat{d}_\eta = \min \left\{ k \,\middle|\, (\lambda_{k+1} + \cdots + \lambda_D) \leq \eta \cdot (\lambda_1 + \cdots + \lambda_D) \right\}$$

Theorem A - Tangent Space

Let $X_1, \dots X_m$ be an iid sample from μ, with $\mu = Law(X+Y)$.
X has probability density $\varphi : M \to \mathbb{R}$ and $\|Y\| \leq s$.

Given $\theta, \delta > 0$, suppose that:

$$\sqrt{2\tau s} \leq r \leq S_1 \quad \text{and} \quad \frac{m(r-2s)^d}{\log m} \geq S_2$$

Then with probability at least $1-\delta$,

$$\max_i \measuredangle \left(\widehat{T}_i, T_i \right) \leq \theta$$

Here S_1, S_2 are:

$$S_1 = \frac{c_1 \tau \sin\theta}{(d+2)} \cdot \frac{\varphi_{\min}}{3\varphi_{\min} + 8d\varphi_{\max} + 5\alpha\tau}$$

$$S_2 = \frac{c_2 (d+2)^2}{\omega_d \varphi_{\min} \sin^2\theta} \log\left(\frac{c_3 D}{\delta}\right)$$

and $c_1 = 1/16$, $c_2 = 4642$, $c_3 = 14$.

Theorem A - Tangent Space

Let $X_1, \dots X_m$ be an iid sample from μ, with $\mu = Law(X+Y)$.
X has probability density $\varphi : M \to \mathbb{R}$ and $\|Y\| \le s$.
Given $\theta, \delta > 0$, suppose that:

$$\sqrt{2\tau s} \le r \le S_1 \quad \text{and} \quad \frac{m(r-2s)^d}{\log m} \ge S_2$$

Then with probability at least $1-\delta$,

$$\max_i \measuredangle\left(\widehat{T}_i, T_i\right) \le \theta$$

Here S_1, S_2 are:

$$S_1 = \frac{c_1 \tau \sin\theta}{(d+2)} \cdot \frac{\varphi_{\min}}{3\varphi_{\min} + 8d\varphi_{\max} + 5\alpha\tau}$$

$$S_2 = \frac{c_2 (d+2)^2}{\omega_d \varphi_{\min} \sin^2\theta} \log\left(\frac{c_3 D}{\delta}\right)$$

and $c_1 = 1/16$, $c_2 = 4642$, $c_3 = 14$.

Theorem A - Tangent Space

Let $X_1, \ldots X_m$ be an iid sample from μ, with $\mu = Law(X+Y)$.
X has probability density $\varphi : M \to \mathbb{R}$ and $\|Y\| \le s$.
Given $\theta, \delta > 0$, suppose that:

$$\sqrt{2\tau s} \le r \le S_1 \quad \text{and} \quad \frac{m(r-2s)^d}{\log m} \ge S_2$$

Then with probability at least $1-\delta$,

$$\max_i \measuredangle \left(\widehat{T}_i, T_i \right) \le \theta$$

Here S_1, S_2 are:

$$S_1 = \frac{c_1 \tau \sin\theta}{(d+2)} \cdot \frac{\varphi_{\min}}{3\varphi_{\min} + 8d\varphi_{\max} + 5\alpha\tau}$$

$$S_2 = \frac{c_2 (d+2)^2}{\omega_d \varphi_{\min} \sin^2\theta} \log\left(\frac{c_3 D}{\delta}\right)$$

and $c_1 = 1/16$, $c_2 = 4642$, $c_3 = 14$.

Theorem B - Dimension

Let $X_1, \ldots X_m$ be an iid sample from μ, with $\mu = Law(X+Y)$.
X has probability density $\varphi : M \to \mathbb{R}$ and $\|Y\| \leq s$.
Given $\eta, \delta > 0$, suppose that:

$$\sqrt{2\tau s} \leq r \leq S_1 \quad \text{and} \quad \frac{m(r-2s)^d}{\log m} \geq S_2$$

Then with probability at least $1-\delta$,

$$\forall i, \quad \hat{d}_i = d$$

Here S_1, S_2 are:

$$S_1 = \frac{c_1 \tau}{(d+2)D(1+\eta^{-1})} \cdot \frac{\varphi_{\min}}{3\varphi_{\min} + 8d\varphi_{\max} + 5\alpha\tau}$$

$$S_2 = \frac{c_2(d+2)^2 D^2 (1+\eta^{-1})^2}{\omega_d \varphi_{\min}} \log\left(\frac{c_3 D}{\delta}\right)$$

and $c_1 = 1/48$, $c_2 = 41778$, $c_3 = 14$.

Theorem B - Dimension

Let $X_1, \ldots X_m$ be an iid sample from μ, with $\mu = Law(X + Y)$.
X has probability density $\varphi : M \to \mathbb{R}$ and $\|Y\| \leq s$.
Given $\eta, \delta > 0$, suppose that:

$$\sqrt{2\tau s} \leq r \leq S_1 \quad \text{and} \quad \frac{m(r-2s)^d}{\log m} \geq S_2$$

Then with probability at least $1 - \delta$,

$$\forall i, \quad \hat{d}_i = d$$

Here S_1, S_2 are:

$$S_1 = \frac{c_1 \tau}{(d+2)D(1+\eta^{-1})} \cdot \frac{\varphi_{\min}}{3\varphi_{\min} + 8d\varphi_{\max} + 5\alpha\tau}$$

$$S_2 = \frac{c_2 (d+2)^2 D^2 (1+\eta^{-1})^2}{\omega_d \varphi_{\min}} \log\left(\frac{c_3 D}{\delta}\right)$$

and $c_1 = 1/48$, $c_2 = 41778$, $c_3 = 14$.

Strategy of proof

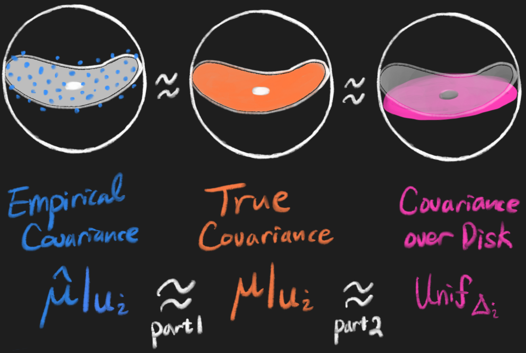
Total estimation error is allocated to two approximations:

1. Empirical covariance $\approx$ True covariance
2. Covariance over curvy disk $\approx$ Covariance over flat disk.

Part 1 is a modified matrix Hoeffding inequality.

Part 2 is measured using the Wasserstein distance. This is translated to matrix norm using a Lipschitz relation.

Strategy of proof

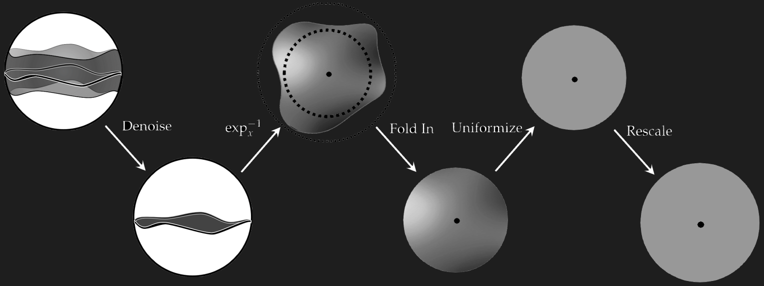


Part 1: Matrix Hoeffding: $\Sigma[\hat{\mu}|_{U_i}] \approx \Sigma[\mu|_{U_i}]$

Part 2: Wasserstein distance and Lipschitz relation

$W_1(\mu|_{U_i}, \text{Unif}_{\Delta_i}) \approx 0$ and thus $\Sigma[\mu|_{U_i}] \approx \Sigma[\text{Unif}_{\Delta_i}]$.

Transportation plan



Flattening a manifold using a transportation plan.

$$
\begin{aligned}
Q =& 3\sigma + (\rho + 2\sigma)^2 + \frac{1.18\varphi_{\max}}{\Phi}(2\rho + (\rho + 2\sigma)^2)(1 - \Omega^d) \\
&+ \frac{2.18\rho}{\Phi}(\varphi_{\max} - \varphi_{\min}) + 1.38\rho^3 \\
\leq& 3 + \frac{8\varphi_{\max} d + 5\alpha\tau}{\varphi_{\min}}
\end{aligned}
$$

Thank you!