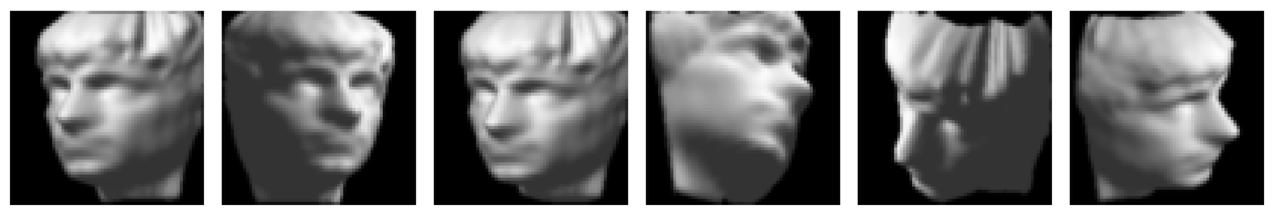
Dimension detection via slivers

Siu-Wing Cheng and Man-Kwun Chiu

scheng@cse.ust.hk, chiumk@cse.ust.hk
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong

Introduction

- A lot of data have very high extrinsic dimension. For example, the following images can be viewed as points in a 4096-dimensional space.



- When the images are related, it is often postulated that they live on a manifold M of much lower dimension.
- Knowing the manifold dimension can help in Manifold Embedding [1] and Manifold Reconstruction [2].
- We address the dimension detection problem
  - Compute the dimension of a manifold given a set of point samples drawn from the manifold.

Manifold

- An m-dimensional manifold M $\subseteq \mathbb{R}^d$
  - Every point in M has a neighborhood which resembles $\mathbb{R}^m$.
  - E.g. 1-manifolds : curve. 2-manifolds : surface.

Theoretical Result

- Proposed Method to estimate the manifold dimension by
  - Analyzing the shape of simplices formed by point samples in a neighborhood
  - Detecting slivers in some neighborhoods
- Given a sample set P
  - an m-dimensional manifold M in $\mathbb{R}^d$ with positive reach
  - Poisson process with parameter λ
- For sufficiently large λ such that $\lambda = \Omega(2^{\Theta(m^6)} + 2^{\Theta(km^2)})$,
  - Time : $O(kd|P|^{1+1/k})$
  - Probability of success: $1 - 2^{-k}$

Sliver

- Slivers are simplices with very little volume
  - Vertices and edges are not σ-slivers for any $\sigma \in (0,1)$.
  - For $2 \leq j \leq d$, a j-simplex τ is a σ-sliver if

$$\mathrm{vol}(\tau) \leq \frac{\sigma^j L_\tau^j}{j!},$$

where L_τ is the longest edge length of τ.

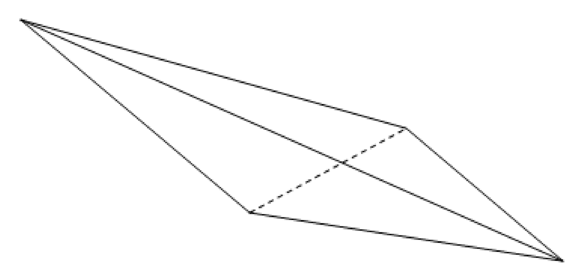


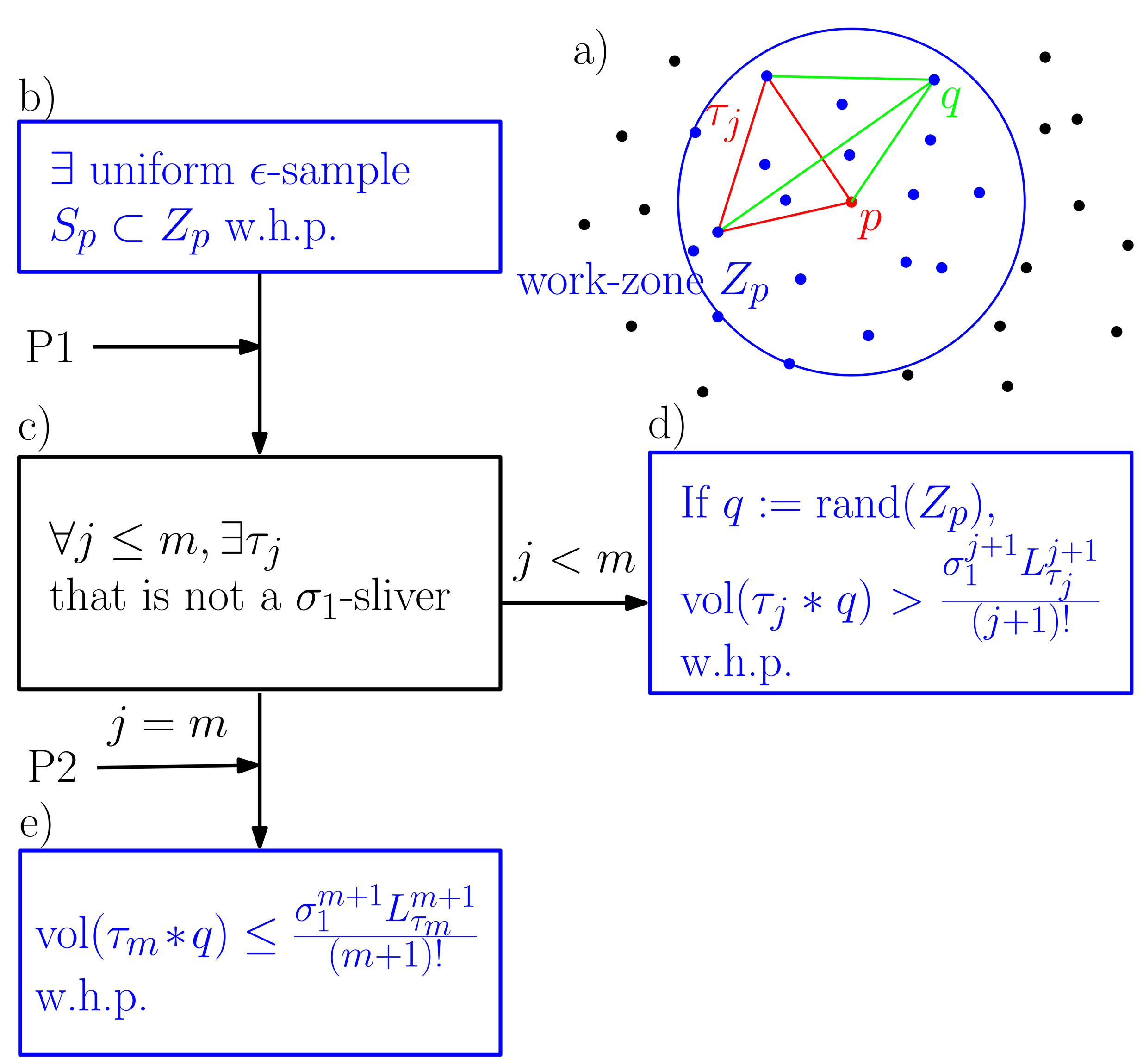
Figure 1: 3-dimensional slivers

- In [2], $\forall \sigma < \sigma_0, \forall \varepsilon < \varepsilon_0(\sigma)$, if P is an uniform ε-sample,

C1 $\exists$ sliver free, m-complex $\approx$ M.

C2 $\forall j \geq m+1$, j-simplex τ is a σ-sliver if $L_\tau = O(\varepsilon)$ and boundaries(τ) is sliver free.

Underlying Theory



a) We form a work-zone Z_p from P.

b) We prove that there exists an uniform ε-sample in which the points around p within distance $O(\varepsilon)$ are subset of Z_p with high probability.

c) Applying the result (C1), for any $j \leq m$, there exists a j-dimensional simplex τ_j that is not a σ_1-sliver.

d) When $j < m$, we can prove that, if we choose a point q inside Z_p at random, $\mathrm{vol}(\tau_j * q) > \frac{\sigma_1^{j+1} L_{\tau_j}^{j+1}}{(j+1)!}$ with high probability.

e) When $j = m$, using the result (C2), we can prove that $\mathrm{vol}(\tau_m * q) \leq \frac{\sigma_1^{m+1} L_{\tau_m}^{m+1}}{(m+1)!}$ with high probability.

Experimental Results

- n-Sphere
  The numbers of successes in each entry are in this order (Ours, MLE, MA, PN, LPCA).

	100 pts	500 pts	1000 pts
S^3	30,30,29,26,29	30,30,30,30,30	30,30,30,30,30
S^4	30,30,29, 6, 5	30,30,30, 9,23	30,30,30,13,30
S^5	27,30,21, 0, 0	30,30,30, 0, 0	30,30,30, 0, 6
S^6	29,23, 1, 0, 0	30,30,30, 0, 0	30,30,20, 0, 0
S^7	30, 8, 1, 0, 0	30,30,29, 0, 0	29,30, 0, 0, 0
S^8	27, 2, 0, 0, 0	30,30, 9, 0, 0	30,30, 0, 0, 0
S^9	9, 0, 0, 0, 0	30,18, 2, 0, 0	30,30, 0, 0, 0

- ISOMAP face dataset (I-Head)
  - Three parameters : vertical and horizontal pose and lighting direction.

	Ours	MLE	MA	PN	LPCA	ISOMAP
I-Head	4	4.31	4.47	3.98	3	3

References

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[2] S.-W. Cheng, T.K. Dey and E.A. Ramos. Manifold reconstruction from point samples. *Proc. 16th Annu. ACM-SIAM SODA.*, 1018–1027, 2005.