Minimax Reach Estimator

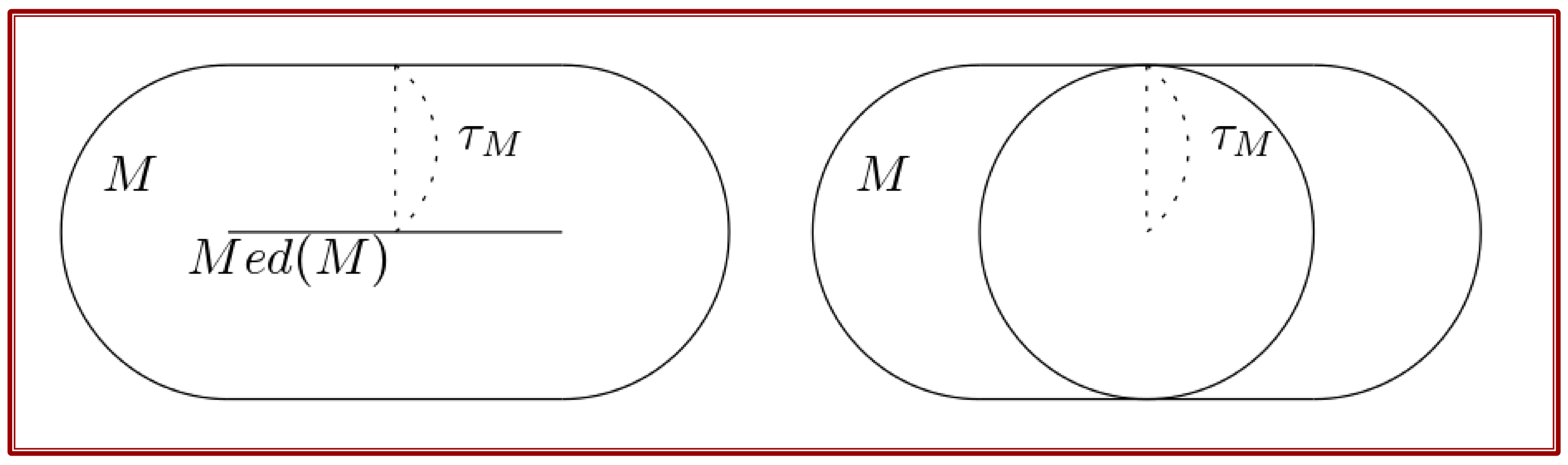
Eddie Aamari, Frédéric Chazal, Jisu Kim,
Bertrand Michel, Alessandro Rinaldo, Larry Wasserman

Abstract

Regularity parameters are crucial to derive approximation properties as well as in implementation. The reach has been proven to play a key role in computational geometry. It carries both local and global regularity information, and can be seen as a minimal scale parameter. The goal is to propose an estimator of the reach τ_M of a d-dimensional sub-manifold $M \subset \mathbb{R}^m$ given a random i.i.d. sample $\{X_1, \ldots, X_n\}$. We give minimax bounds for reach estimation over a class of manifolds satisfying natural geometric constraints.

Definition

Reach



- Medial axis of M, $Med(M)$, is the set of points in $\mathbb{R}^m$ that do not have unique nearest neighbors on M.
- The reach of M, denoted by τ_M, is the minimum distance from $Med(M)$ to M: $\tau_M = \inf_{x \in Med(M), y \in M} ||x - y||_2$.
- The reach τ_M gives maximal offset size of M on which the projection is well defined.
- The reach τ_M also gives the maximum radius of a ball that you can roll over M.

Minimax rate

- Suppose you observe iid data X and the set of distributions $\mathcal{P}$ and loss function $l(,)$ is fixed, and you are interested in parameter θ.
- Maximum risk of an estimator $\hat{\theta}$ is the risk that the estimator can make in the worst case, i.e.
$$\sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[l\left(\hat{\theta}(X), \theta(P)\right)\right].$$
- Minimax rate is the infimum of a maximum risk over all possible estimators, $\hat{\theta}$ i.e.
$$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[l\left(\hat{\theta}(X), \theta(P)\right)\right].$$
- In our case, loss function is $l(a, b) = \left|\frac{1}{a} - \frac{1}{b}\right|^p$, and the parameter of interest is the reach τ_P.

Regularity conditions

We consider $\mathcal{P}$ to be set of distributions with regular conditions:

- The reach τ_M of the support manifold M is lower bounded by $\tau_{\min}$.
- All the arc-length parametrized geodesic γ on the support manifold M has L-Lipsctiz 2nderivatives: for all $t, s \in \mathbb{R}$, $||\gamma''(t) - \gamma''(s)|| \le L|t - s|$.
- The probability distribution P has density f with respect to uniform measure on M satisfying $f_{min} \le f(x) \le f_{max}$.

Reach estimators

- The reach τ_M of M can be characterized as
$$\frac{1}{\tau_M} = \sup_{a, b \in M, a \ne b} \frac{d(b - a, T_a M)}{||b - a||_2^2}.$$

When tangent spaces are known:

- We consider the plugin estimator
$$\frac{1}{\hat{\tau}} = \sup_{1 \le i \ne j \le n} \frac{d(X_j - X_i, T_{X_i} M)}{||X_j - X_i||_2^2}.$$

When tangent spaces are unknown:

- $\hat{T}_i$ be an estimator of $T_{X_i} M$, then we consider the plugin estimator
$$\frac{1}{\hat{\tau}} = \sup_{1 \le i \ne j \le n} \frac{d(X_j - X_i, \hat{T}_i)}{||X_j - X_i||_2^2}.$$

Upper bound and Lower bound of Minimax Rate

Upper bound

- The maximum risk of $\hat{\tau}$ is $O\left(n^{-\frac{4p}{5d-1}}\right)$, i.e. there exists $C^{(1)}$ that depends only on $\tau_{\min}$, L, f_{min}, f_{max} that
$$\sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\left|\frac{1}{\hat{\tau}} - \frac{1}{\tau_M}\right|^p\right] \le C^{(1)} n^{-\frac{4p}{5d-1}}.$$
- The minimax rate is also upper bounded by $O\left(n^{-\frac{4p}{5d-1}}\right)$, i.e.
$$\inf_{\hat{\tau}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\left|\frac{1}{\hat{\tau}} - \frac{1}{\tau_M}\right|^p\right] \le C^{(1)} n^{-\frac{4p}{5d-1}}.$$

Lower bound

- The minimax rate is lower bounded by $O\left(n^{-\frac{p}{d}}\right)$, i.e. there exists $C^{(2)}$ that depends only on $\tau_{\min}$, L, f_{min}, f_{max} that
$$\inf_{\hat{\tau}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[\left|\frac{1}{\hat{\tau}} - \frac{1}{\tau_M}\right|^p\right] \ge C^{(2)} n^{-\frac{p}{d}}.$$

References

[1] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. Discrete Comput. Geom., 41(3):461–479, 2009.
[2] F. Chazal and A. Lieutier. The λ-medial axis. Graphical Models, 67(4):304–331, July 2005.
[3] H. Federer. Curvature measures. Trans. Amer. Math. Soc., 93, 1959.
[4] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. 39(1):419–441, 2008.
[5] http://www.stat.cmu.edu/topstat/
[6] https://team.inria.fr/geometrica/