Reach

김지수 (Jisu KIM)

통계이론세미나 - 위상구조의 통계적 추정, 2023 가을학기

For this lecture note, for $A \subset \mathbb{R}^d$, we use the notation $A^r \coloneqq \left\{x \in \mathbb{R}^d : d(x, A) < r\right\}$ for the r-offset of A.

Definition of reach

First introduced by Federer [4], the reach is a regularity parameter defined as follows. Given a closed subset $A \subset \mathbb{R}^d$, the medial axis of A, denoted by $\mathrm{Med}(A)$, is the subset of $\mathbb{R}^d$ composed of the points that have at least two nearest neighbors on A. Namely, denoting by $d_A(x) = d(x, A) = \inf_{q \in A} \|q - x\|$ the distance function to A,

$$\mathrm{Med}(A) = \left\{x \in \mathbb{R}^d | \exists q_1 \neq q_2 \in A, ||q_1 - x|| = ||q_2 - x|| = d(x, A)\right\}. \tag{1}$$

The reach of A is then defined as the minimal distance from A to $Med(A)$. See Figure

Definition ([4, 4.1 Definition])**.** The reach of a closed subset $A \subset \mathbb{R}^d$ is defined as

$$\tau_A = \inf_{q \in A} d\left(q, \mathrm{Med}(A)\right) = \inf_{q \in A, x \in \mathrm{Med}(A)} ||q - x||. \tag{2}$$

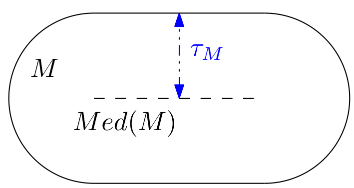

Figure 1: The medial axis of a set M is the set of points that have at least two nearest neighbors on the set M, and the reach is the distance between the set and its medial axis.

Definition. Some authors refer to τ_A^{-1} as the *condition number* [6, 7]. From the definition of the medial axis in (1), the projection $\pi_A(x) = \arg\min_{p \in A} \|p - x\|$ onto A is well defined outside $Med(A)$. The reach is the largest distance $r \geq 0$ such that π_A is well defined on the r-offset. Hence, the reach condition can be seen as a generalization of convexity, since a set $A \subset \mathbb{R}^d$ is convex if and only if $\tau_A = \infty$.

In the case of submanifolds, one can reformulate the definition of the reach in the following manner.

Theorem ([4, Theorem 4.18])**.** *For all submanifold $M \subset \mathbb{R}^d$,*

$$\tau_M = \inf_{q_1 \neq q_2 \in M} \frac{\|q_1 - q_2\|_2^2}{2d(q_2 - q_1, T_{q_1} M)}. \tag{3}$$

This formulation has the advantage of involving only points on M and its tangent spaces, while (2) uses the distance to the medial axis $Med(M)$, which is a global quantity.

The ratio appearing in (3) can be interpreted geometrically, as suggested in Figure 2. This ratio is the radius of an ambient ball, tangent to M at q_1 and passing through q_2.

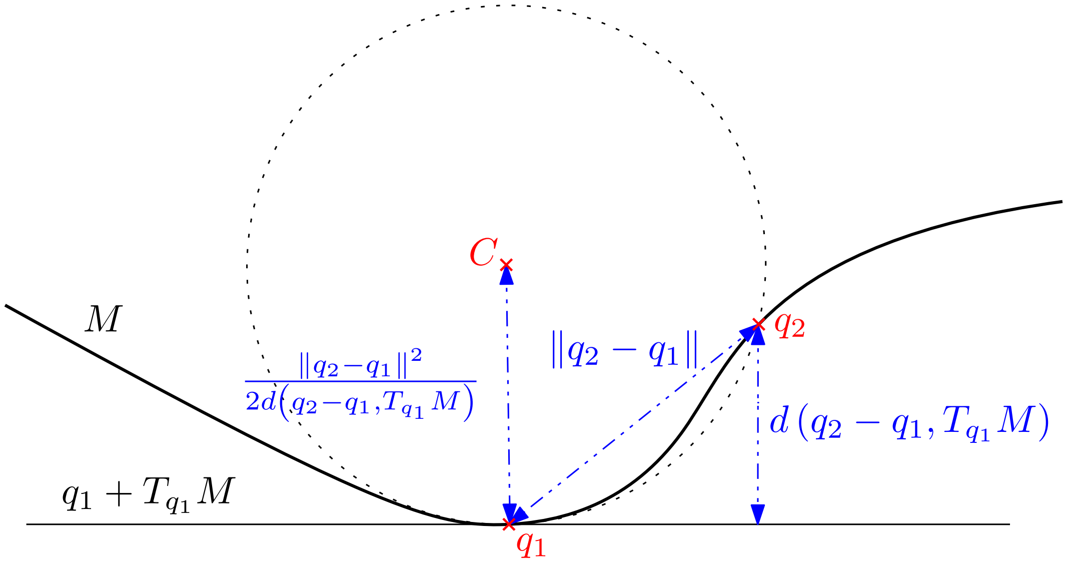

Figure 2: Geometric interpretation of quantities involved in (3).

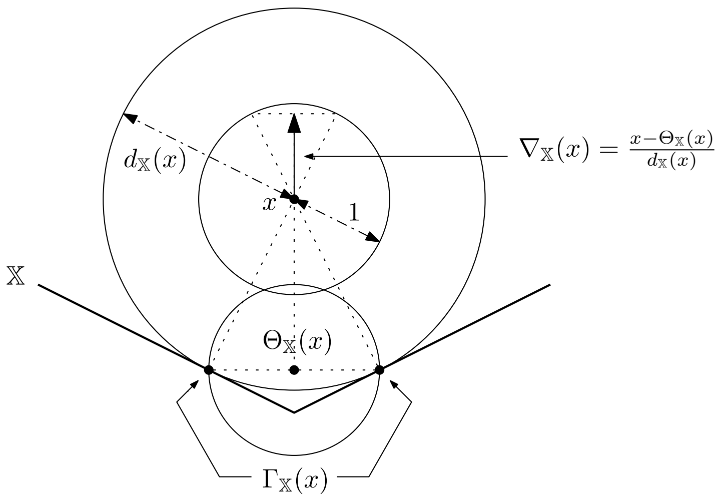

Figure 3: The graphical illustration for the generalized gradient $\nabla_A(x)$, from [3].

μ-reach and weak feature size

For a non-smooth target space, the reach of the space can be zero. In this case, we can deploy a more general notion of feature size, called μ-reach, introduced by [3]. For any point $x \in \mathbb{R}^d \setminus A$, let $\Gamma_A(x)$ be the set of points in A closest to x. Let $\Theta_A(x)$ be the center of the unique smallest closed ball enclosing $\Gamma_A(x)$. Then, for $x \in \mathbb{R}^d \setminus A$, the generalized gradient of the distance function d_A is defined as

$$\nabla_A(x) = \frac{x - \Theta_A(x)}{d_A(x)}, \tag{4}$$

and set $\nabla_A(x) = 0$ for $x \in A$. See Figure 3 for a graphical illustration.

Then, for $\mu \in (0, 1]$, the μ-medial axis of A is defined as

$$\mathrm{Med}_\mu(A) = \left\{x \in \mathbb{R}^d \setminus A : \|\nabla_A(x)\|_2 < \mu\right\},$$

[1] Finally, the μ-reach of A is defined as the minimal distance from A to $\mathrm{Med}_\mu(A)$. See Figure 4.

[1] In [1], the μ-medial axis is defined with $\|\nabla_A(x)\|_2 \leq \mu$.

Definition ([3])**.** The μ-reach of a closed subset $A \subset \mathbb{R}^d$ is defined as

$$\tau_A^\mu = \inf_{q \in A} d\left(q, \mathrm{Med}_\mu(A)\right) = \inf_{q \in A, x \in \mathrm{Med}_\mu(A)} ||q - x||. \tag{5}$$

Note that if $\mu = 1$, the corresponding μ-reach equals to the reach of A.

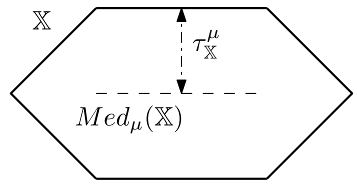

Figure 4: The μ-medial axis and the μ-reach.

An even weaker condition is the weak feature size.

Definition ([3])**.** Let $A \subset \mathbb{R}^d$ be a closed subset.

(a) The critical point of the distance function d_A is defined as the points x for which $\nabla_A(x) = 0$. Equivalently, a point x is a critical point if and only if it lies in the convex hull of $\Gamma_A(x)$. A real $c \geq 0$ is a critical value of d_A if there exists a critical point $x \in \mathbb{R}^d$ such that $d_A(x) = c$. A regular value of d_A is a value which is not critical.

(b) The weak feature size of A, denoted as $\mathrm{wfs}(A)$, is the infimum of the positive critical points of d_A. If d_A does not have critical values, then $\mathrm{wfs}(A) = \infty$.

Regularity conditions imposed by reach

Positive reach is the minimal regularity assumption in many geometrical and topological problems. The condition of the positive reach is required in homology inference, manifold reconstruction, volume estimation, manifold clustering, dimension estimation, diffusion maps, etc. We will see some of them later in the class.

One basic regularity condition imposed by reach is that the distance function d_A becomes somewhat Lipschitz.

Theorem. *Let $A \subset \mathbb{R}^d$ be a closed subset with its reach $\tau_A > 0$.*

1. *If $x \in \mathbb{R}^d \backslash Med(A)$ and $b \in A$, then*

$$\langle x - \pi_A(x), \pi_A(x) - b \rangle \geq -\frac{\|\pi_A(x) - b\|_2 \|x - \pi_A(x)\|_2}{2\tau_A}$$

2. *If $x, y \in \mathbb{R}^d$ with $d_A(x), d_A(y) \leq r < \tau_A$, then*

$$\|d_A(x) - d_A(y)\|_2 \leq \frac{\tau_A}{\tau_A - r} \|x - y\|_2 \, .$$

.

Now, consider regularity conditions imposed on the local structure of a manifold. First, (3) suggests that, at a differential level, the reach gives a lower bound on the radii of curvature of M. Equivalently, τ_M^{-1} is an upper bound on the curvature of M.

Proposition ([6, Proposition 6.1])**.** *Let $M \subset \mathbb{R}^d$ be a submanifold with its reach $\tau_M > 0$, and $\gamma_{p,v}$ an arc-length parametrized geodesic of M with $\gamma(0) = p \in M$ and $\gamma'(0) = v \in \partial B_{T_pM}(0, 1)$. Then for all t,*

$$\left\|\gamma_{p,v}''(t)\right\| \leq 1/\tau_M. \tag{6}$$

In analogy with function spaces, the class $\left\{M \subset \mathbb{R}^d : \tau_M \geq \tau_{min} > 0\right\}$ can be interpreted as the Hölder space $\mathcal{C}^2(1/\tau_{min})$.

Related regularity condition is that the tangent space also varies in Lipschitz manner.

Proposition ([6, Proposition 6.2])**.** *Let $M \subset \mathbb{R}^d$ be a submanifold with its reach $\tau_M > 0$, and let $p, q \in M$ be two points with its geodesic distance given as $d_M(p,q)$. Let ϕ be the angle between the tangent spaces T_p and T_q defined by $\cos(\phi) = \min_{u \in T_p} \min_{v \in T_q} |\langle u, v\rangle|$. Then*

$$\cos\phi \geq 1 - \frac{d_M(p,q)}{\tau_M}.$$

Another related regularity condition is that the geodesic distance is not too different from the Euclidean distance.

Proposition ([6, Proposition 6.2])**.** *Let $M \subset \mathbb{R}^d$ be a submanifold with its reach $\tau_M > 0$, and let $p, q \in M$ be two points with $\|p - q\|_2 \leq \frac{\tau_M}{2}$. Then the geodesic distance $d_M(p,q)$ is bounded as*

$$d_M(p,q) \leq \tau_M - \tau_M \sqrt{1 - \frac{2\|p-q\|_2}{\tau_M}}.$$

The reach also imposes global regularity condition as well. As illustrated in Figure 5, the reach condition τ_M prevents bottleneck structures where M is nearly self-intersecting.

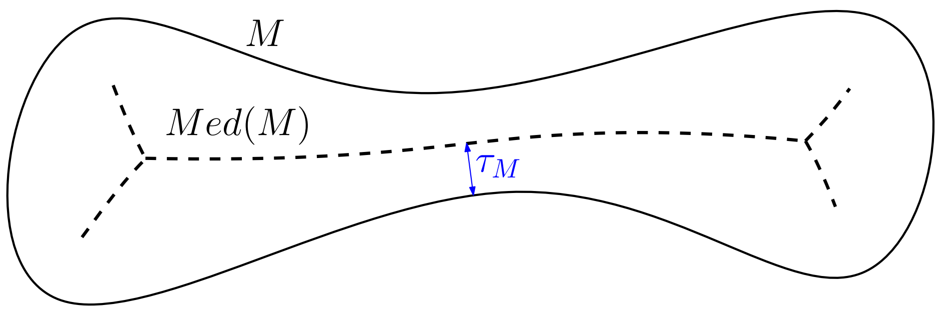

Figure 5: A narrow bottleneck structure yields a small reach τ_M.

One important condition in topology inference is when a set A is homotopy equivalent to its r-offset A^r. In fact, this homotopy equivalence holds under a weaker condition of μ-reach: if the set A has a positive μ-reach, then the offset A^r deformation retracts to A when the offset size is not large, and in particular, they are homotopy equivalent.

Theorem ([5, Theorem 12])**.** *Let $A \subset \mathbb{R}^d$ be a subset with positive μ-reach $\tau^\mu > 0$. For $r \leq \tau^\mu$, the r-offset A^r deformation retracts to A. In particular, A and A^r are homotopy equivalent.*

Geometry of reach

In (6), we have seen that the reach condition bounds the acceleration of a geodesic. In Figure 5, we have seen that the reach condition prevents bottleneck structures that the set is nearly self-intersecting. In fact, these are the two cases where the reach can arise. We will show that the reach is determined either by a bottleneck structure or an area of high curvature. We first provide the formal definition of the bottleneck structure.

Definition ([2, Definition 3.1, modified])**.** A pair of points (q_1, q_2) in M is said to be a bottleneck of M if there exists $z_0 \in Med(M)$ such that $q_1, q_2 \in \partial\mathcal{B}(z_0, r)$ and $\|q_1 - q_2\|_2 = 2r$ for some $r > 0$.

An immediate observation is that if $(q_1, q_2) \in M$ is a bottleneck, then $\frac{q_1+q_2}{2} \in Med(M)$ is a critical point of the distance function d_M. And now we turn to the definition of the minimal curvature radius, which describes the area of high curvature.

Definition ([1, Theorem 3.5, modified])**.** Let $M \subset \mathbb{R}^d$ be a compact submanifold, then the minimal curvature radius of M is defined as

$$R_l(M) := \inf_{p \in M, v \in \partial B_{T_pM}(0,1)} \left\|\gamma''_{p,v}(0)\right\|,$$

where $\gamma_{p,v}$ is an arc-length parametrized geodesic of M with $\gamma(0) = p$ and $\gamma'(0) = v$.

The bottleneck is a global structure of a manifold while the minimal curvature radius is a local structure, since the minimal curvature radius is determined from the differential structure of a manifold. Now, the reach is determined either by a bottleneck structure or an area of high curvature.

Theorem ([2, Theorem 3.4][1, Theorem 3.5]). *Let $M \subset \mathbb{R}^d$ be a compact submanifold with reach $\tau_M > 0$. At least one of the following two assertions holds:*

- *(Global Case) M has a bottleneck $(q_1, q_2) \in M^2$ with $\|q_1 - q_2\|_2 = 2\tau_M$.*
- *(Local case) There exists $q_0 \in M$ and an arc-length parametrized γ_0 such that $\gamma_0(0) = q_0$ and $\|\gamma_0''(0)\| = \frac{1}{\tau_M}$.*

In other words,

$$\tau_M = \min \{\mathrm{wfs}(M), R_l(M)\}.$$

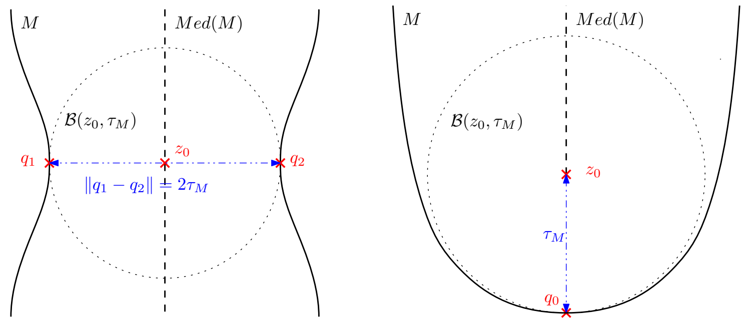

Figure 6: The reach can arise from either by a bottleneck structure (left), or an area of high curvature (right).

Reach Estimation and Minimax

Let $\mathcal{M}$ be the set of m-dimensional compact conneted submanifolds M of $\mathbb{R}^d$ with its reach lower bounded by $\tau_{\min} > 0$, and for some $k \geq 3$, there exists $L > 0$ such that for all $p \in M$ and $v \in \partial B_{T_p M}(0,1)$, $t \in (-\frac{1}{L}, \frac{1}{L})$, the k-th derivative of $\gamma_{p,v}$ is bounded as

$$\left\|\gamma_{p,v}^{(k)}(t)\right\|_2 \leq L,$$

where $\gamma_{p,v}$ is an arc-length parametrized geodesic of M with $\gamma(0) = p$ and $\gamma'(0) = v$. Let $\mathcal{M}_G := \{M \in \mathcal{M} : \mathrm{wfs}(M) \leq R_l(M)\}$ and $\mathcal{M}_L := \{M \in \mathcal{M} : R_l(M) < \mathrm{wfs}(M)\}$.

Let $0 < p_{\min} \leq p_{\max} < \infty$, and $\mathcal{P}_G$ be the set of distributions P such that its support $\mathrm{supp}(P) \in \mathcal{M}_G$ and it has a density p with respect to the volume measure on $\mathrm{supp}(P)$ with $0 < p_{\min} \leq p \leq p_{\max} < \infty$. Define similarly $\mathcal{P}_L$ with its support $\mathrm{supp}(P) \in \mathcal{M}_L$ and it has a density p with respect to the volume measure on $\mathrm{supp}(P)$ with $0 < p_{\min} \leq p \leq p_{\max} < \infty$.

Theorem ([1, Theorem 3.7, Theorem 6.6]). *For a distribution P, let* $\mathrm{reach}(P)$ *be the reach of its support* $\mathrm{supp}(P)$*, then*

$$C_1 \left(\frac{1}{n}\right)^{-\frac{k-2}{d}} \leq \inf_{\hat{\tau}} \sup_{P \in \mathcal{P}_L} \mathbb{E}_P \left[|\hat{\tau} - \mathrm{reach}(P)|\right] \leq C_2 \left(\frac{\log n}{n}\right)^{-\frac{k-2}{d}},$$

$$C_1 \left(\frac{1}{n}\right)^{-\frac{k}{d}} \leq \inf_{\hat{\tau}} \sup_{P \in \mathcal{P}_G} \mathbb{E}_P \left[|\hat{\tau} - \mathrm{reach}(P)|\right] \leq C_2 \left(\frac{\log n}{n}\right)^{-\frac{k}{d}}.$$

References

[1] Eddie Aamari, Clément Berenfeld, and Clément Levrard. Optimal reach estimation and metric learning. *Ann. Statist.*, 51(3):1086–1108, 2023.

[2] Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electron. J. Stat.*, 13(1):1359–1399, 2019.

[3] Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete & Computational Geometry*, 41(3):461–479, 2009.

[4] Herbert Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[5] Jisu Kim, Jaehyeok Shin, Frédéric Chazal, Alessandro Rinaldo, and Larry Wasserman. Homotopy Reconstruction via the Cech Complex and the Vietoris-Rips Complex. In Sergio Cabello and Danny Z. Chen, editors, *36th International Symposium on Computational Geometry (SoCG 2020)*, volume 164 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:19, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[6] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

[7] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012.