Computing geometric feature sizes for algebraic manifolds

Sandra Di Rocco[1], Parker B. Edwards[2], David Eklund[3], Oliver Gäfvert[4], and Jonathan D. Hauenstein[2]

[1]*Department of Mathematics, KTH, 10044, Stockholm, Sweden*
[2]*Dept. of Applied and Computational Mathematics and Statistics, U. Notre Dame, Notre Dame, IN, USA*
[3]*RISE, Research Institutes of Sweden, Isafjordsgatan 22, 16440, Kista, Sweden*
[4]*Mathematical Institute, U. Oxford, United Kingdom*

Abstract

We introduce algorithms which use numerical algebraic geometric methods to compute lower bounds on the reach, local feature size, and the weak feature size of a real algebraic manifold starting with only the manifold's defining polynomial equations as input. In many cases our algorithms compute these quantities directly rather than a lower bound. One motivation for developing these algorithms is to combine them with methods which produce dense point samples of algebraic manifolds, then subsequently apply topological data analysis or computational geometry methods to the point samples either to extract information or to test those methods' performance. We consider computational results both from our feature size algorithms and from a proof-of-concept experiment that combines them with persistent homology computations. This second example compares the performance of two persistent homology approaches: the "standard" version with globally dense samples and an "adaptive" version with samples which are adaptively dense with respect to the local feature size.[1]

1 Introduction

The reach [9], local feature size [1], and weak feature size [4, 10] of a compact space $X \subseteq \mathbb{R}^n$ are geometric feature sizes that frequently arise in the context of topological data analysis (TDA) and computational geometry. In many applications the space X is not fully specified, e.g., when the input is a "dense" finite sample $P \subseteq \mathbb{R}^n$ which approximates X. It is only possible to estimate the feature sizes of X using P in this case.

In contrast, consider the case where X is defined as the set of solutions $\{x \in \mathbb{R}^n \mid f_1(x) = f_2(x) = \cdots = f_c(x) = 0\}$ to c polynomials $f_1, \ldots, f_c$ in n variables. The list of polynomials fully specifies X. These algebraic spaces are of intrinsic interest to algebraic geometers, but also arise in application domains like robotics and kinematics [11].

Recent years have witnessed growing interest in developing methods that combine numerical algebraic geometry algorithms (NAG, see e.g. [12]), which can compute numerical solutions to polynomial equations, with TDA and computational geometry approaches to analyze algebraic spaces. There are now NAG-based algorithms which perform a crucial first step, namely producing a provably dense point cloud sample P for an algebraic space X using the list of polynomials

[1]The full version of this paper has been submitted to a journal for peer review. A preprint is available at https://arxiv.org/abs/2209.01654.

defining X as input [7, 8], see Figure 1. Computing feature sizes of X is also necessary, however, to determine just how dense a sample one needs for subsequent geometry-based analysis.

In the full version of this paper, we provide NAG-based algorithms for computing lower bounds on the reach, local feature size, and weak feature size of an algebraic manifold using only its defining polynomial equations as input and examples from computations with these algorithms. Much of the contribution is theoretical algebraic geometry, but in this abstract and its associated talk we will particularly focus on the connection to TDA and computational geometry.

For example, our feature size computations provide a previously unavailable baseline to investigate the performance of methods which estimate feature sizes. As a proof of concept, we consider an "adaptive" subsampling method for persistent homology proposed by Dey et al. [6] that estimates local feature sizes. Their approach has recently been expanded to a general framework for adaptive subsampling by Cavanna and Sheehy [2, 3]. Using our algorithms, we find evidence that the method of Dey et al. performs comparably to a baseline analog of Chazal and Lieutier [5] which requires directly computed local feature sizes.

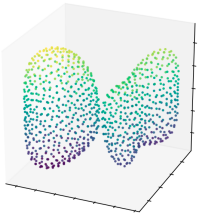

Figure 1: Dense sample from a quartic surface, the solutions to a polynomial equation of degree 4 in 3 variables, computed using its defining polynomial equation as input.

2 Feature sizes and numerical algebraic computations

We recall the definition of various feature sizes and consider a high-level overview of how to translate the weak feature size into polynomial conditions in the algebraic setting. Following the terminology of Chazal and Lieutier [4], let $X \subseteq \mathbb{R}^n$ be a non-empty and compact subspace. The distance-to-X function $d_X : \mathbb{R}^n \to \mathbb{R}$ is defined by $d_X(z) = \min_{x \in X} \|x - z\|$, where $\|\cdot\|$ is the standard norm. For any $z \in \mathbb{R}^n$, the set of closest points $\pi_X(z)$ is $\{x \in X \mid \|x - z\| = d_X(z)\}$. The *medial axis* of X, M_X, is the closure of the set of points in $\mathbb{R}^n$ with at least two closest points on X, i.e. $M_X = \overline{\{z \in \mathbb{R}^n \mid |\pi_X(z)| > 1\}}$.

Definition 1. The minimum distance from X to M_X, i.e. $\min_{x \in X} d_{M_X}(x)$, is the *reach* [9] of X. For a point $z \in \mathbb{R}^n$, the distance $d_{M_X}(z)$ is the *local feature size* [1] of X at z. A point $z \in M_X$ is a *critical point of* d_X or *geometric bottleneck* of X if $z \in \text{conv}(\pi_X(z))$. Let B_X be the set of geometric bottlenecks. The *weak feature size* [4, 10] of X is the minimum distance from X to B_X.

Suppose that $X \subseteq \mathbb{R}^n$ is algebraic, i.e. $X = \{x \in \mathbb{R}^n \mid f_1(x) = f_2(x) = \cdots = f_c(x) = 0\}$ for polynomial equations $f_1, \ldots, f_c$. Let $F = (f_1, \ldots, f_c) : \mathbb{R}^n \to \mathbb{R}^c$. To give some flavor of the algebra involved, we will consider computing the weak feature size. For the sake of clarity, we will suppress some technical assumptions regarding the algebraic structure of X.[2]

To perform algebraic computations, we require polynomial conditions which characterize when a point $z \in \mathbb{R}^n$ is a geometric bottleneck of X. If z is a geometric bottleneck, then $z \in \text{conv}(\pi_X(z))$.

[2]More precisely: the corresponding *complex* algebraic set $V(F) = \{x \in \mathbb{C}^n \mid F(x) = 0\}$ must be smooth and equidimensional of dimension $n - c$.

- There exists an affinely independent $\{x_1, x_2, \ldots, x_k\} \subseteq \pi_X(z)$ with $F(x_i) = 0$ for all i, and positive real numbers $t_1, t_2, \ldots, t_k$ with $z = \sum_{i=1}^{k} t_i x_i$. These are polynomial equalities with variables $t_1, \ldots, t_k$, the coordinates of z, and the coordinates of each x_i.
- With $x_1, \ldots, x_k$ as before, $\|x_1 - z\|^2 = \|x_i - z\|^2$ for $i = 2, \ldots, k$. These are polynomial equalities in the same variables.
- Since d_X^2 defined via the constrained optimization problem $d_X^2(z) = \min_{x \in X} \|x - z\|^2 = \min_{F(x)=0} \|x - z\|^2$, any point $x \in \pi_X(z)$ satisfies Lagrange multiplier conditions for the optimization problem. If $x_1, \ldots, x_k$ are as above, every point x_i fulfills these conditions. So there are multipliers $\lambda_1, \lambda_2, \ldots, \lambda_c$ where $x_i - z = \sum_{i=1}^{c} \lambda_i \nabla f_i(x)$.

In addition to the algorithms we discussed in the Introduction, we also show the following main result in the full version (Theorem 4.15).

Theorem 2. Almost all algebraic manifolds in a large class[3] without any defining polynomials of degree 2 have finitely many geometric bottlenecks.

3 Computational examples

We conclude by discussing the results of some computations performed using our algorithms.

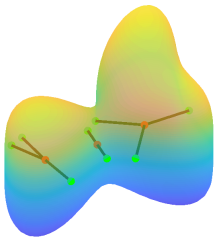

Figure 2: Geometric bottlenecks (red) with their closest points (green) for a quartic surface.

3.1 Quartic surface

Consider the quartic surface in $\mathbb{R}^3$ illustrated in Figure 2 and defined by

$$F = 4x^4 + 7y^4 + 3z^4 - 3 - 8x^3 + 2x^2y - 4x^2 - 8xy^2 - 5xy + 8x - 6y^3 + 8y^2 + 4y.$$

We computed the geometric 2-bottlenecks and geometric 3-bottlenecks (bottlenecks with 2 and 3 closest points) using our algorithm for finding a lower bound on the weak feature size. The lower bound we computed was approximately 0.354 and was attained at the geometric 2-bottleneck.

3.2 Adaptive subsamples and persistent homology

Consider the following greedy subsampling framework in Algorithm 1 of Dey et al. [6, Alg. 1].

Example 3 (Uniform subsample)**.** Let X be an algebraic manifold. We can compute $\omega < \mathrm{wfs}(X)$ via our NAG-based algorithm. For any $\lambda \in [0, 1]$, define $\omega_\lambda : \mathbb{R}^n \to \mathbb{R}$ to be the constant function given by $\omega_\lambda(z) = \lambda\omega$. The output of $\textsc{Subsample}(\hat{X}, \omega_\lambda)$ is a "uniform subsample" of $\hat{X}$.

[3] More precisely: general complete intersections.

Algorithm 1: SUBSAMPLE

Input : Sample $\hat{X} \subseteq \mathbb{R}^n$ and function $s : \hat{X} \to \mathbb{R}$
Output: A subsample of $\hat{X}$

Put $\hat{X}$ in a max priority queue sorted by s;
Set OUTPUT to $\emptyset$;
while *The queue is not empty* **do**
    Add the highest priority $\hat{x}_1$ in the queue to OUTPUT and remove it from the queue;
    Delete any point $\hat{x}_2$ from the queue where $\|\hat{x}_1 - \hat{x}_2\| \leq s(\hat{x}_1)$;
end
`Return(`OUTPUT`)`

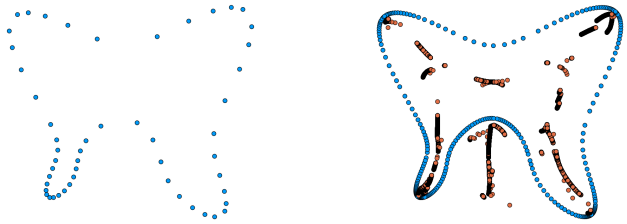

Figure 3: Adaptive samples of the butterfly curve, lfs (left) and "lean feature size" (right). More points are retained closer to the medial axis and "lean medial axis" (orange).

Example 4 (Local adaptive subsample)**.** With lfs the local feature size function of X, for any $\lambda \in [0, 1]$, define $\text{lfs}_\lambda : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by $\text{lfs}_\lambda(z) = \lambda \, \text{lfs}(z)$. The output of SUBSAMPLE$(\hat{X}, \text{lfs}_\lambda)$ is an "adaptive subsample" of $\hat{X}$ with respect to lfs. In practice, we compute $\widehat{\text{lfs}}(z) \leq \text{lfs}(z)$.

Example 5 (Lean adaptive subsample)**.** Let $\hat{X} \subseteq \mathbb{R}^n$ be a point sample. In [6, Def. 3], Dey et al. define the $\frac{\pi}{5}$-*lean set* of $\hat{X}$, $L_{\frac{\pi}{5}}$, to be a subset of the set of midpoints $\{\frac{p+q}{2}\}_{p \neq q \in \hat{X}}$ that fulfills other geometric conditions. The distance function $d_{L_{\frac{\pi}{5}}}$ estimates the distance to a subset of M_X. For $\lambda \in [0, 1]$ define $\text{lnfs}_\lambda : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by $\text{lnfs}_\lambda(z) = \lambda d_{L_{\frac{\pi}{5}}}(z)$. The output of SUBSAMPLE$(\hat{X}, \text{lnfs}_\lambda)$ is an adaptive subsample of $\hat{X}$ with respect to the lean feature size.

We first computed dense samples of the "butterfly curve", $x^4 - x^2y^2 + y^4 - 4x^2 - 2y^2 - x - 4y + 1 = 0$, with density determined by homology inference theorems for persistent homology. We then computed the required feature sizes (e.g., Figure 3) and formed subsamples using the Examples above over a range of parameter values for λ. These subsamples were then used as input to compute degree 1 persistent homology, and scores computed using the resulting persistence diagrams were used to compare the three subsampling methods. The estimated lnfs-adaptive method performed comparably to the lfs-adaptive "baseline".

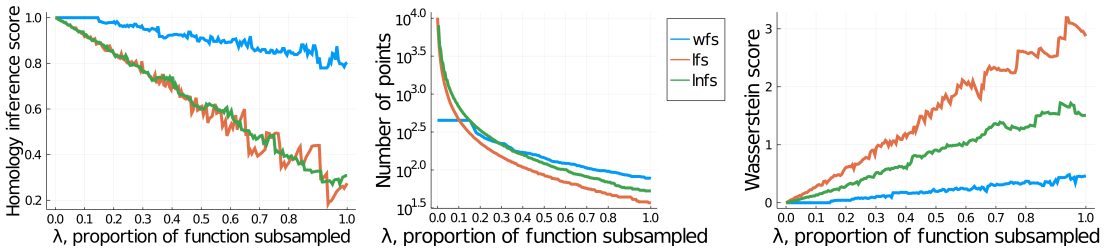

Figure 4: Computational results comparing the behavior of subsampling methods. Higher homology inference scores, lower numbers of points, and lower Wasserstein indicate better performance.

References

[1] N. Amenta and M. Bern. Surface reconstruction by voronoi filtering. *Discrete & Computational Geometry*, 22(4):481–504, 1999.

[2] N. J. Cavanna. *Methods in Homology Inference*. PhD thesis, University of Connecticut, 2019.

[3] N. J. Cavanna and D. R. Sheehy. Adaptive metrics for adaptive samples. *Algorithms*, 13(8):200, 2020.

[4] F. Chazal and A. Lieutier. Weak feature size and persistent homology: computing homology of solids in $\mathbb{R}^n$ from noisy data samples. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 255–262. ACM, 2005.

[5] F. Chazal and A. Lieutier. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Comput. Geom.*, 40(2):156–170, 2008.

[6] T. K. Dey, Z. Dong, and Y. Wang. Parameter-free topology inference and sparsification for data on manifolds. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2733–2747. SIAM, Philadelphia, PA, 2017.

[7] S. Di Rocco, D. Eklund, and O. Gäfvert. Sampling and homology via bottlenecks. *Mathematics of Computations, to appear*, 2022.

[8] E. Dufresne, P. Edwards, H. Harrington, and J. Hauenstein. Sampling real algebraic varieties for topological data analysis. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1531–1536. IEEE, 2019.

[9] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[10] K. Grove. Critical point theory for distance functions. In *Differential geometry: Riemannian geometry (Los Angeles, CA, 1990)*, volume 54 of *Proc. Sympos. Pure Math.*, pages 357–385. Amer. Math. Soc., Providence, RI, 1993.

[11] S. Martin, A. Thompson, E. A. Coutsias, and J.-P. Watson. Topology of cyclo-octane energy landscape. *The Journal of Chemical Physics*, 132(23):234115, 2010.

[12] A. J. Sommese and C. W. Wampler, II. *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.