The VC Dimension of Metric Balls Under Fréchet and Hausdorff Distances

Anne Driemel
University of Bonn, Germany
driemel@cs.uni-bonn.de

Jeff M. Phillips
University of Utah, Salt Lake City, USA
jeffp@cs.utah.edu

Ioannis Psarros
National & Kapodistrian University of Athens, Greece
ipsarros@di.uoa.gr

Abstract

The Vapnik-Chervonenkis dimension provides a notion of complexity for systems of sets. If the VC dimension is small, then knowing this can drastically simplify fundamental computational tasks such as classification, range counting, and density estimation through the use of sampling bounds. We analyze set systems where the ground set X is a set of polygonal curves in $\mathbb{R}^d$ and the sets $\mathcal{R}$ are metric balls defined by curve similarity metrics, such as the Fréchet distance and the Hausdorff distance, as well as their discrete counterparts. We derive upper and lower bounds on the VC dimension that imply useful sampling bounds in the setting that the number of curves is large, but the complexity of the individual curves is small. Our upper bounds are either near-quadratic or near-linear in the complexity of the curves that define the ranges and they are logarithmic in the complexity of the curves that define the ground set.

2012 ACM Subject Classification Theory of computation → Randomness, geometry and discrete structures; Theory of computation → Computational geometry

Keywords and phrases VC dimension, Fréchet distance, Hausdorff distance

Digital Object Identifier 10.4230/LIPIcs.SoCG.2019.28

Related Version A full version of the paper is available at `https://arxiv.org/abs/1903.03211`.

Funding *Anne Driemel*: Anne Driemel thanks the Hausdorff Center for Mathematics for their generous support and the Netherlands Organization for Scientific Research (NWO) for support under Veni Grant 10019853.
Jeff M. Phillips: Jeff Phillips thanks his support from NSF CCF-1350888, ACI-1443046, CNS-1514520, CNS-1564287, and IIS-1816149. Part of the work was completed while visiting the Simons Institute for Theory of Computing.
Ioannis Psarros: This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme ≪ Human Resources Development, Education and Lifelong Learning ≫ in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

Acknowledgements We thank Peyman Afshani for useful discussions on the topic of this paper. We also thank the organizers of the 2016 NII Shonan Meeting "Theory and Applications of Geometric Optimization" where this research was initiated.

1 Introduction

A *range space* $(X, \mathcal{R})$ (also called *set system*) is defined by a ground set X and a set of ranges $\mathcal{R}$, where each $r \in \mathcal{R}$ is a subset of X. A data structure for range searching answers queries for the subset of the input data that lies inside the query range. In range counting, we are interested only in the size of this subset. In our setting, a range is a metric ball defined by a curve and a radius. The ball contains all curves that lie within this radius from the center under a specific distance function (e.g., Fréchet or Hausdorff distance).

A crucial descriptor of any range space is its VC-dimension [33, 31, 30] and related shattering dimension, which we define formally below. These notions quantify how complex a range space is, and have played fundamental roles in machine learning [34, 6], data structures [12], and geometry [24, 10]. For instance, specific bounds on these complexity parameters are critical for tasks as diverse as neural networks [6, 27], art-gallery problems [32, 21, 28], and kernel density estimation [26]. The Fréchet distance is a popular distance measure for curves. The Fréchet distance is very similar to the Hausdorff distance for sets, which is defined as the minimal maximum distance of a pair of points, one from each set, under all possible mappings between the two sets. The difference between the two distance measures is that the Fréchet distance requires the mapping to adhere to the ordering of the points along the curve. Both distance measures allow flexible associations between parts of the input elements which sets them apart from classical ℓ_p distances and makes them so suitable for trajectory data under varying speeds. In particular, the last five years have seen a surge of interest into data structures for trajectory processing under the Fréchet distance, manifested in a series of publications [14, 23, 15, 2, 35, 8, 19, 11, 18, 7, 20].

Our contribution in this paper is a comprehensive analysis of the Vapnik-Chervonenkis dimension of the corresponding range spaces. The resulting VC dimension bounds, while being interesting in their own right, have a plethora of applications through the implied sampling bounds. We detail a range of implications of our bounds in Section 10.

2 Definitions

In this section, we formally define the distances between curves as well as VC-dimension and range spaces, so we can state our main results. This basic set up will be enough to prove the main results for discrete distance. Then in Section 6 we provide more advanced geometric definitions and properties about VC dimension with our proofs for the continuous distances.

2.1 Distance measures

In the following, we define the Hausdorff distance, the discrete and the continuous Fréchet distance, and the Weak Fréchet distance. We denote by $\|\cdot\|$ the Euclidean norm $\|\cdot\|_2$.

▶ **Definition 1** (Directed Hausdorff distance.)**.** *Let X, Y be two subsets of some metric space (M, d). The directed Hausdorff distance from X to Y is:*

$$\mathrm{d}_{\overrightarrow{H}}(X, Y) = \sup_{u \in X} \inf_{v \in Y} \mathrm{d}(u, v).$$

▶ **Definition 2** (Hausdorff distance.)**.** *Let X, Y be two subsets of some metric space (M, d). The Hausdorff distance between X and Y is:*

$$\mathrm{d}_H(X, Y) = \max\{d_{\overrightarrow{H}}(X, Y), d_{\overrightarrow{H}}(Y, X)\}.$$

▶ **Definition 3.** *Given polygonal curves V and U with vertices $v_1, \dots, v_{m_1}$ and $u_1, \dots, u_{m_2}$ respectively, a traversal $T = (i_1, j_1), \dots, (i_t, j_t)$ is a sequence of pairs of indices referring to a pairing of vertices from the two curves such that:*
1. $i_1, j_1 = 1$, $i_t = m_1$, $j_t = m_2$.
2. $\forall (i_k, j_k) \in T : i_{k+1} - i_k \in \{0, 1\}$ *and* $j_{k+1} - j_k \in \{0, 1\}$.
3. $\forall (i_k, j_k) \in T : (i_{k+1} - i_k) + (j_{k+1} - j_k) \geq 1$.

▶ **Definition 4** (Discrete Fréchet distance)**.** *Given polygonal curves V and U with vertices $v_1, \dots, v_{m_1}$ and $u_1, \dots, u_{m_2}$ respectively, we define the Discrete Fréchet Distance between V and U as the following function:*

$$\mathrm{d}_{dF}(V, U) = \min_{T \in \mathcal{T}} \max_{(i_k, j_k) \in T} \|v_{i_k} - u_{j_k}\|,$$

where $\mathcal{T}$ denotes the set of all possible traversals for V and U.

Any polygonal curve V with vertices $v_1, \dots, v_{m_1}$ and edges $\overline{v_1 v_2}, \dots, \overline{v_{m_1 - 1} v_{m_1}}$ has a uniform parametrization that allows us to view it as a parametrized curve $v : [0, 1] \mapsto \mathbb{R}^2$.

▶ **Definition 5** (Fréchet distance)**.** *Given two parametrized curves $u, v : [0, 1] \mapsto \mathbb{R}^2$, their Weak Fréchet distance is defined as follows:*

$$\mathrm{d}_F(u, v) = \min_{\substack{f : [0,1] \mapsto [0,1] \\ g : [0,1] \mapsto [0,1]}} \max_{\alpha \in [0,1]} \|v(f(\alpha)) - u(g(\alpha))\|,$$

where f ranges over all continuous and monotone bijections with $f(0) = 0$ and $f(1) = 1$. The Weak Fréchet distance d_{wF} is defined as above, except that f and g range over all continuous functions (not exclusively bijections) with $f(0) = 0$ and $f(1) = 1$ and $g(0) = 0$ and $g(1) = 1$.

2.2 Range spaces

Each range space can be defined as a pair of sets $(X, \mathcal{R})$, where X is the *ground set* and $\mathcal{R}$ is the *range set*. Let $(X, \mathcal{R})$ be a range space. For $Y \subseteq X$, we denote:

$$\mathcal{R}_{|Y} = \{R \cap Y \mid R \in \mathcal{R}\}.$$

If $\mathcal{R}_{|Y}$ contains all subsets of Y, then Y is *shattered* by $\mathcal{R}$.

▶ **Definition 6** (Vapnik-Chernovenkis dimension)**.** *The Vapnik-Chernovenkis dimension [30, 31, 33] (VC dimension) of $(X, \mathcal{R})$ is the maximum cardinality of a shattered subset of X.*

▶ **Definition 7** (Shattering dimension)**.** *The shattering dimension of $(X, \mathcal{R})$ is the smallest δ such that, for all m,*

$$\max_{\substack{B \subset X \\ |B| = m}} |\mathcal{R}_{|B}| = O(m^{\delta}).$$

It is well-known [6, 24] that for a range space $(X, \mathcal{R})$ with VC-dimension ν and shattering dimension δ that $\nu \leq O(\delta \log \delta)$ and $\delta = O(\nu)$. So bounding the shattering dimension and bounding the VC-dimension are asymptotically equivalent within a log factor.

▶ **Definition 8** (Dual range space)**.** *Given a range space $(X, \mathcal{R})$, for any $p \in X$, we define*

$$\mathcal{R}_p = \{R \mid R \in \mathcal{R}, p \in R\}.$$

The dual range space of $(X, \mathcal{R})$ is the range space $(\mathcal{R}, \{\mathcal{R}_p \mid p \in X\})$.

It is a well-known fact that if a range space has VC dimension ν, then the dual range space has VC dimension $\leq 2^{\nu+1}$ (see e.g. [24]).

Many ways are known to bound the VC dimension of geometric range spaces. For instance when the ground set is $\mathbb{R}^d$ and the ranges are defined by inclusion in halfspaces, then the range space and its dual range space are isomorphic and both have VC-dimension and shattering dimension d. When the ranges are defined by inclusion in balls, then the VC-dimension and shattering dimension is $d+1$, and the dual range spaces have bounds of d [24]. It is also for instance known [9] that the composition ranges formed as the k-fold union or intersection of ranges from a range space with bounded VC-dimension ν induces a range space with VC-dimension $O(\nu k \log k)$, and this was recently shown that this is tight for even some simple range spaces such as those defined by halfspaces [13]. More such results are deferred to Section 6.

2.3 Range spaces induced by distance measures

Let (M, d) be a pseudometric space. We define the *ball* of radius r and center p, under the distance measure d, as the following set:

$$b_{\mathrm{d}}(p, r) = \{x \in M \mid \mathrm{d}(x, p) \leq r\},$$

where $p \in M$. The doubling dimension of a metric space (M, d), denoted as $\mathrm{ddim}(M, \mathrm{d})$, is the smallest integer t such that any ball can be covered by at most 2^t balls of half the radius.

In this paper, we study the VC dimension of range spaces $(X, \mathcal{R})$ induced by pseudometric spaces[1] (M, d) by setting $X = M$ and

$$\mathcal{R} = \{b_{\mathrm{d}}(p, r) \mid r \in \mathbb{R}, r > 0, p \in M\}.$$

It is a reasonable question to ask whether the doubling dimension of a metric space influences the VC dimension of the induced range space. In general, a bounded doubling dimension does not imply a bounded VC dimension of the induced range space and vice versa. Recently, Huang et al. [25] showed that if we allow a small $(1+\varepsilon)$-distortion of the distance function d, the shattering dimension can be upper bounded by $O(\varepsilon^{-O(\mathrm{ddim}(M,\mathrm{d}))})$. It is conceivable that the doubling dimension of the metric space of the Discrete Fréchet distance and Hausdorff distance is bounded, as long as the underlying metric has bounded doubling dimension. However, for the continuous Fréchet distance, the doubling dimension is known to be unbounded [16]. Moreover, we will see that much better bounds can be obtained by a careful study of the specific distance measure.

Specifically, we study an *unbalanced* version of the above range space, in that we distinguish between the complexity of objects of the ground set and the complexity of objects defining the ranges. To this end, we define, for any integers d and m, $\mathbb{X}_m^d := \left(\mathbb{R}^d\right)^m$ and we treat the elements of this set as ordered sets of points in $\mathbb{R}^d$ of size m. Formally, we study range spaces with ground set $\mathbb{X}_m^d$ and range set defined as

$$\mathcal{R}_{\mathrm{d},k} = \left\{b_{\mathrm{d}}(p, r) \cap \mathbb{X}_m^d \mid r \in \mathbb{R}, r > 0, p \in \mathbb{X}_k^d\right\}$$

under different variants of the Fréchet and the Hausdorff distance. We emphasize that the range space consists of ranges of all radii.

[1] While we may use the term *metric* or *pseudometric* to define the range, our methods do not assume any metric properties of the inducing distance measure.

3 Our Results

Table 1 shows an overview of our bounds. For metric balls defined on point sets (resp. point sequences) in $\mathbb{R}^d$ we show that the VC dimension is at most near-linear in dk, the complexity of the ball centers that define the ranges, and at most logarithmic in dm, the complexity of point sets of the ground set. Our lower bounds show that these bounds are almost tight in all parameters k, d, and m. For the Fréchet distance, where the ground set X are continuous polygonal curves in $\mathbb{R}^d$ we show an upper bound that is quadratic in k, quadratic in d, and logarithmic in m. The same bounds in k and m hold for the Hausdorff distance, where the ground set are sets of line segments in $\mathbb{R}^2$. We obtain slightly better bounds in k for the Weak Fréchet distance. Our lower bounds extend to the continuous case, but are only tight in the dependence on m – the complexity of the ground set.

Table 1 Our bounds on the VC dimension of range spaces of the form $(\mathbb{X}^d_m, \mathcal{R}_{\mathrm{d},k})$, for d being the distance measures in the table. In the first column we distinguish between $\mathbb{X}^d_m$ consisting of *discrete* point sequences vs. $\mathbb{X}^d_m$ consisting of *continuous* polygonal curves. The lower bounds hold for all distance measures in this table.

discrete	Hausdorff	$O(dk\log(dkm))$ (Theorems 9,10)	$\Omega(\max(dk\log k, \log dm))$ ($d \geq 4$, Theorem 23) $\Omega(\max(k, \log m))$ ($d \geq 2$, Theorem 22)
	Fréchet		
continuous	weak Fréchet	$O(d^2k\log(dkm))$ (Theorem 16)	
	Fréchet	$O(d^2k^2\log(dkm))$ (Theorem 18)	
	Hausdorff	$O(k^2\log(km))$ ($d = 2$, Theorem 21)	

While the VC dimension bounds for the discrete Hausdorff and Fréchet metric balls may seem like an easy implication of composition theorems for VC dimension [9, 13], we still find three things about these results remarkable:

1. First, for Fréchet variants, there are $\Theta(2^k 2^m)$ valid alignment paths in the free space diagram. And one may expect that these may materialize in the size of the composition theorem. Yet by a simple analysis of the shattering dimension, we show that they do not.
2. Second, the VC dimension only has logarithmic dependence on the size m of the curves in the ground set, rather than a polynomial dependence one would hope to obtain by simple application of composition theorems. This difference has important implications in analyzing real data sets where we can query with simple curves (small k), but may not have a small bound on the size of the curves in the data set (large m).
3. Third, for the continuous variants, the range spaces can indeed be decomposed into problems with ground sets defined on line segments. However, we do not know of a general d-dimensional bound on the VC-dimension of range space with a ground set of segments, and ranges defined by segments within a radius r of another segment. We are able to circumvent this challenge with circuit-based methods to bound the VC-dimension and careful predicate design for the Fréchet distance, but for Hausdorff distance are only able to prove a bound in $\mathbb{R}^2$.

4 Our Approach

Our methods use the fact that both the Fréchet distance and the Hausdorff distance are determined by one of a discrete set of events, where each event involves a constant number of simple geometric objects. For example, it is well known that the Hausdorff distance between two discrete sets of points is equal to the distance between two points from the two sets. The corresponding event happens as we consider a value $\delta > 0$ increasing from 0 and we record which points of one set are contained in which balls of radius δ centered at points from the other set. The same phenomenon is true for the discrete Fréchet distance between two point sequences. In particular, the so-called free-space matrix which can be used to decide whether the discrete Fréchet distance is smaller than a given value δ encodes exactly the information about which pairs of points have distance at most δ. The basic phenomenon remains true for the continuous versions of the two distance measures if we extend the set of simple geometric objects to include line segments and if we also consider triple intersections. Each type of event can be translated into a range space of which we can analyze the VC dimension. Together, the product of the range spaces encodes the information, which curves lie inside which metric balls, in the form of a set system. This representation allows us to prove bounds on the VC dimension of metric balls under these distance measures.

5 Basic Idea: Discrete Fréchet and Hausdorff

In this section we prove our upper bounds in the discrete setting. Let $\mathbb{X}^d_m = \left(\mathbb{R}^d\right)^m$; we treat the elements of this set as ordered sets of points in $\mathbb{R}^d$ of size m. The range spaces that we consider in this section are defined over the ground set $\mathbb{X}^d_m$ and the range set of balls under either the Hausdorff or the Discrete Fréchet distance. The proofs in the proceeding sections all follow the basic idea of the proof in the discrete setting.

▶ **Theorem 9.** *Let $(\mathbb{X}^d_m, \mathcal{R}_{dH,k})$ be the range space with $\mathcal{R}_{dH,k}$ the set of all balls under the Hausdorff distance centered at point sets in $\mathbb{X}^d_k$. The VC dimension is $O\left(dk\log(dkm)\right)$.*

Proof. Let $\{S_1, \ldots, S_t\} \subseteq \mathbb{X}^d_m$ and $S = \bigcup_i S_i$; we define S so that it ignores the ordering with each S_i and is a single set of size tm. Any intersection of a Hausdorff ball with $\{S_1, \ldots, S_t\}$ is uniquely defined by a set $\{B_1 \cap S, \ldots, B_k \cap S\}$, where $B_1, \ldots, B_k$ are balls in $\mathbb{R}^d$. To see that, notice that the discrete Hausdorff distance between two sets of points is uniquely defined by the distances between points of the two sets.

Consider the range space $(\mathbb{R}^d, \mathcal{B})$, where $\mathcal{B}$ is the set of balls in $\mathbb{R}^d$. We know that the shattering dimension is $d+1$ [24]. Hence,

$$\max_{S\subseteq\mathbb{R}^d, |S|=tm} |\mathcal{B}_{|S}| = O((tm)^{d+1}).$$

This implies that $|\{\{B_1 \cap S, \ldots, B_k \cap S\} \mid B_1, \ldots, B_k \text{ are balls in } \mathbb{R}^d\}| \leq O((tm)^{(d+1)k})$, and hence[2],

$$2^t \leq O\left((tm)^{(d+1)k}\right) \implies t = O\left(dk\log(dkm)\right). \qquad ◀$$

▶ **Theorem 10.** *Let $(\mathbb{X}^d_m, \mathcal{R}_{dF,k})$ be the range space with $\mathcal{R}_{dF,k}$ the set of all balls under the Discrete Fréchet distance centered at polygonal curves in $\mathbb{X}^d_k$. The VC dimension is $O\left(dk\log(dkm)\right)$.*

Proof. Let $\{S_1, \ldots, S_t\} \subseteq \mathbb{X}^d_m$ and $S = \bigcup_i S_i$. Any intersection of a Discrete Fréchet ball with $\{S_1, \ldots, S_t\}$ is uniquely defined by a sequence $B_1 \cap S, \ldots, B_k \cap S$, where $B_1, \ldots, B_k$ are balls in $\mathbb{R}^d$. The number of such sequences can be bounded by $O((tm)^{(d+1)k})$ as in the

[2] for $u > \sqrt{e}$ if $x/\ln(x) \leq u$ then $x \leq 2u\ln u$. Hence, if $tm/\log(tm) \leq dkm$, then $tm = O(dkm\log(dkm))$.

proof of Theorem 9. Enforcing that a sequence contains a valid alignment path only reduces the number of possible distinct sets formed by t curves, and it can be determined using these intersections and the two orderings of $B_1, \ldots, B_k$ and of vertices within some $S_j \in \mathbb{X}^d_m$. ◀

6 Preliminaries

In this section, we provide a more advanced set of geometric primitives and other technical known results about VC-dimension. We also derive some simple corollaries. We also provide some basic results about the distances which will couple with the geometric primitives in our proofs for continuous distance measures.

We again consider a ground set $\mathbb{X}^d_m = \left(\mathbb{R}^d\right)^m$ which we treat as a set of polygonal curves with points in $\mathbb{R}^d$ of size m. Given such a curve $s \in \mathbb{X}^d_m$, let $V(s)$ be its ordered set of vertices and $E(s)$ its ordered set of edges.

6.1 A simple model of computation

We consider a model of computation that will be useful for modeling primitive geometric sets, and in turn bounding the VC-dimension of an associated range space. These will be useful in that they allow the invocation of powerful and general tools to describe range spaces defined by distances between curves. We allow the following operations, which we call *simple operations*:

- the arithmetic operations $+, -, \times$, and $/$ on real numbers,
- jumps conditioned on $>, \geq, <, \leq, =$, and $\neq$ comparisons of real numbers, and
- output 0 or 1.

We say a function requires t simple operations if it can be computed with a circuit of depth t composed only of these simple operations. Notably, the lack of a square-root operator creates some challenges when dealing with geometric objects.

6.2 Geometric primitives

For any $p \in \mathbb{R}^d$ we denote by $B_r(p)$ the ball of radius r, centered at p. For any two points $s, t \in \mathbb{R}^d$, we denote by $\overline{st}$ the line segment from s to t. Whenever we store such a line segment, for technicalities within the lemma below, we store the coordinates of its endpoints s and t. For any two points $s, t \in \mathbb{R}^d$, we define the stadium centered at $\overline{st}$, $D_r(\overline{st}) = \left\{x \in \mathbb{R}^d \mid \exists p \in \overline{st}\ \|p - x\| \leq r\right\}$. For any two points $s, t \in \mathbb{R}^d$, we define a cylinder $C_r(\overline{st}) = \left\{x \in \mathbb{R}^d \mid \exists p \in \ell(\overline{st})\ \|p - x\| \leq r\right\}$, where $\ell(\overline{st})$ denotes the line supporting the edge $\overline{st}$. Finally, for any two points $s, t \in \mathbb{R}^d$, we define the capped cylinder centered at $\overline{st}$: $R_r(\overline{st}) = \{p + u \mid p \in \overline{st} \text{ and } u \in \mathbb{R}^d \text{ s.t. } \|u\| \leq r \text{ and } \langle t - s, u\rangle = 0\}$.

For each of these geometric sets, we can determine if a point $x \in \mathbb{R}^d$ is in the set with a constant number of operations under a simple model of computation.

▶ **Lemma 11.** *For a point $x \in \mathbb{R}^d$, and any set of the form $B_r(p)$, $D_r(\overline{st})$, $C_r(\overline{st})$, or $R_r(\overline{st})$, we can determine if x is in that set (returns 1, otherwise 0) using $O(d)$ simple operations.*

Proof. For ball $B_r(p)$ we can compute a distance $\|x - p\|^2$ in $O(d)$ time, and determine inclusion with a comparison to r^2. For cylinder $C_r(\overline{st})$ we can compute the closest point to x on this line as

$$\pi_{\overline{st}}(x) = t + \frac{(s - t)\langle (s - t), x\rangle}{\|s - t\|^2}.$$

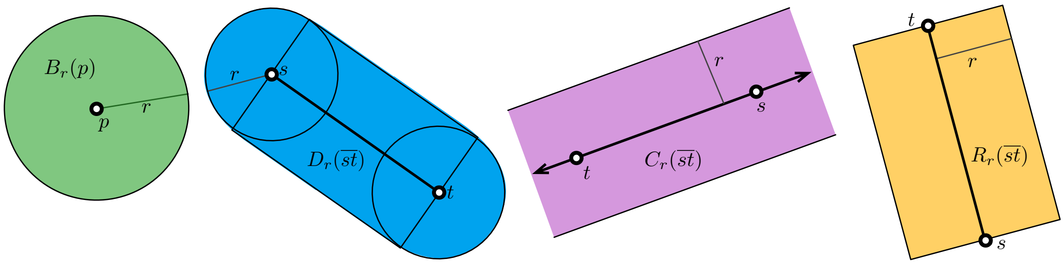

Figure 1 Illustration of basic shapes in $\mathbb{R}^2$, from left to right: a ball $B_r(p)$, a stadium $D_r(\overline{st})$, a cylinder $C_r(\overline{st})$, and a capped cylinder $R_r(\overline{st})$.

Then we can determine inclusion by comparing $\|\pi_{\overline{st}}(x) - x\|^2$ to r^2. For capped cylinder $R_r(\overline{st})$ we also need to compare $\|\pi_{\overline{st}}(x) - t\|^2$ and $\|\pi_{\overline{st}}(x) - s\|^2$ to see if either of these terms is greater than $\|s - t\|^2$. For stadium $D_r(\overline{st})$ we determine inclusion if any x is in any of $R_r(\overline{st})$, $B_r(s)$ or $B_r(t)$. ◀

6.3 Bounding the VC-Dimension

For range spaces defined on continuous curves, our proofs use a powerful theorem from Goldberg and Jerrum [22] as improved and restated by Anthony and Bartlett [6]. It allows one to easily bound the VC-dimension of geometric range spaces under our simple model of computation.

▶ **Theorem 12** (Theorem 8.4 [6]). *Suppose h is a function from $\mathbb{R}^d \times \mathbb{R}^n$ to $\{0, 1\}$ and let*

$$H = \{x \mapsto h(\alpha, x) : \ \alpha \in \mathbb{R}^d\}$$

be the class determined by h. Suppose that h can be computed by an algorithm that takes as input the pair $(\alpha, x) \in \mathbb{R}^d \times \mathbb{R}^n$ and returns $h(\alpha, x)$ after no more than t simple operations. Then, the VC dimension of H is $\leq 4d(t + 2)$.

An example implication can be seen for geometric sets via Lemma 11. Note that this implies any VC dimension upper bound proved in this approach applies to both the range space and its dual range space because the function h is unchanged and the ranges can still be described by $O(d)$ real coordinates.

▶ **Corollary 13.** *For range spaces defined on $\mathbb{R}^d$ with geometric sets $B_r(p)$, $D_r(\overline{st})$, $C_r(\overline{st})$, or $R_r(\overline{st})$ as ranges, the VC dimension is $O(d^2)$. The same $O(d^2)$ VC dimension bound holds for the corresponding dual range spaces, with ground sets as the geometric sets, and ranges defined by stabbing using points in $\mathbb{R}^d$.*

Note that these bounds are not always tight. Specifically, because the VC-dimension for ranges defined geometrically by balls $B_r(p)$ is $O(d)$ [24]. Moreover, the VC-dimension of range spaces defined by cylinders $C_r(\overline{st})$ is known to be $O(d)$ [4]. The ranges defined by capped cylinders $R_r(\overline{st})$ are the intersection of a cylinder and two halfspaces, each with VC-dimension $O(d)$ and hence by the composition theorem [9], this full range spaces also has VC-dimension $O(d)$. Finally, the stadium $D_r(\overline{st})$ is defined by the union of a capped cylinder $R_r(\overline{st})$ and two balls $B_r(s)$ and $B_r(t)$; hence again by the composition theorem [9], its VC-dimension is $O(d)$. However, it is not clear that these improved bounds hold for the dual range spaces, aside for the case of B_r. Moreover, when the ground set X of the range space $(X, \mathcal{R})$ is not $\mathbb{R}^d$, then we need to be cautious in using the k-fold composition theorem [9], which bounds the VC-dimension of complex range spaces derived as the logical intersection or union of simpler range spaces with bounded VC-dimension. In the case of a ground set $X = \mathbb{R}^d$, logical and geometric intersections are the same, but for other ground sets (like dual objects, or line segments $\mathbb{X}_2^d$) this is not necessarily the case. For instance, a line segment $e \in \mathbb{X}_2^d$ may intersect a ball B_r and also a halfspace H while not intersecting the intersection $B_r \cap H$.

6.4 Representation by predicates

In order to prove bounds on the VC dimension of range spaces defined on continuous curves, we establish sets of geometric predicates which are sufficient to determine if two curves have distance at most r to each other. Analyzing the range spaces associated with these predicates (over all possible radii r) allows us to compose them further and to establish VC dimension bounds for the range space induced by the corresponding distance measure. For the Fréchet and Weak Fréchet distance, the predicates mirror those used in range searching data structures [2, 1]. And for the Hausdorff distance on continuous curves, the predicates are derived from the Voronoi diagram [5]. The technical challenges for each case are similar, but require different analyses.

7 The Fréchet distance

We consider the range spaces $(\mathbb{X}_m^d, \mathcal{R}_{F_k})$ and $(\mathbb{X}_m^d, \mathcal{R}_{wF_k})$, where $\mathcal{R}_{F_k}$ (resp. $\mathcal{R}_{wF_k}$) denotes the set of all balls, centered at curves in $\mathbb{X}_k^d$, under the Fréchet distance (resp. weak Fréchet) distance.

7.1 Fréchet distance predicates

It is known that the Fréchet distance between two polygonal curves can be attained, either at a distance between their endpoints, at a distance between a vertex and a line supporting an edge, or at the common distance of two vertices with a line supporting an edge. The third type of event is sometimes called monotonicity event, since it happens when the Weak Fréchet distance is smaller than the Fréchet distance. In this sense, our representation of the ball of radius r under the Fréchet distance is based on the following predicates. Let $s \in \mathbb{X}_m^d$ with vertices $s_1, \ldots, s_m$ and $q \in \mathbb{X}_k^d$ with vertices $q_1, \ldots, q_k$.

- P_1 *(Endpoints (start))* This predicate returns true if and only if $\|s_1 - q_1\| \leq r$.
- P_2 *(Endpoints (end))* This predicate returns true if and only if $\|s_m - q_k\| \leq r$.
- P_3 *(Vertex-edge (horizontal))* Given an edge of s, $\overline{s_j s_{j+1}}$, and a vertex q_i of q, this predicate returns true iff there exist a point $p \in \overline{s_j s_{j+1}}$, such that $\|p - q_i\| \leq r$.
- P_4 *(Vertex-edge (vertical))* Given an edge of q, $\overline{q_i q_{i+1}}$, and a vertex s_j of s, this predicate returns true iff there exist a point $p \in \overline{q_i q_{i+1}}$, such that $\|p - s_j\| \leq r$.
- P_5 *(Monotonicity (horizontal))* Given two vertices of s, s_j and s_t with $j < t$ and an edge of q, $\overline{q_i q_{i+1}}$, this predicate returns true if there exist two points p_1 and p_2 on the *line* supporting the directed edge, such that p_1 appears before p_2 on this line, and such that $\|p_1 - s_j\| \leq r$ and $\|p_2 - s_t\| \leq r$.
- P_6 *(Monotonicity (vertical))* Given two vertices of q, q_i and q_t with $i < t$ and an directed edge of s, $\overline{s_j s_{j+1}}$, this predicate returns true if there exist two points p_1 and p_2 on the *line* supporting the directed edge, such that p_1 appears before p_2 on this line, and such that $\|p_1 - q_i\| \leq r$ and $\|p_2 - q_t\| \leq r$.

▶ **Lemma 14** (Lemma 9, [1]). *Given the truth values of all predicates* $(P1) - (P6)$ *of two curves* s *and* q *for a fixed value of* r*, one can determine if* $d_F(s, q) \leq r$*.*

Predicates $P_1 - P_4$ are sufficient for representing metric balls under the weak Fréchet distance. The proof can be found in the full version of the paper [17].

▶ **Lemma 15.** *Given the truth values of all predicates* $(P_1) - (P_4)$ *of two curves* s *and* q *for a fixed value of* r*, one can determine if* $\mathrm{d}_{wF}(s, q) \leq r$*.*

7.2 Fréchet distance VC dimension bounds

We first consider the range space $(\mathbb{X}^d_m, \mathcal{R}_{wF,k})$, where $\mathcal{R}_{wF,k}$ is the set of all balls under the Weak Fréchet distance centered at curves in $\mathbb{X}^d_k$. The main task is to translate the predicates $P_1 - P_4$ into simple range spaces, and then bound their associated VC dimensions. Consider any two polygonal curves $s \in \mathbb{X}^d_m$ and $q \in \mathbb{X}^d_k$. In order to encode the intersection of polygonal curves with metric balls, we will make use of the following sets:

- $P^r_1(q, s) = B_r(q_1) \cap V(s)$,
- $P^r_2(q, s) = B_r(q_k) \cap V(s)$,
- $P^r_3(q, s) = \{D_r(\overline{s_i s_{i+1}}) \cap V(q) \mid \overline{s_i s_{i+1}} \in E(s)\}$,
- $P^r_4(q, s) = \{D_r(\overline{q_i q_{i+1}}) \cap V(s) \mid \overline{q_i q_{i+1}} \in E(q)\}$.

The proof of the following theorem can be found in the full version of the paper [17].

▶ **Theorem 16.** *Let* $\mathcal{R}_{wF,k}$ *be the set of balls under the Weak Fréchet metric centered at polygonal curves in* $\mathbb{X}^d_k$*. The VC dimension of* $(\mathbb{X}^d_m, \mathcal{R}_{wF,k})$ *is* $O\left(d^2 k \log(dkm)\right)$*.*

We now consider the range space $(\mathbb{X}^d_m, \mathcal{R}_{F,k})$, where $\mathcal{R}_{F,k}$ denotes the set of all balls, centered at curves in $\mathbb{X}^d_k$, under the Fréchet distance. The approach is the same as with the Weak Fréchet distance, except we also need to bound VC dimension of range spaces associated with predicates P_5 and P_6 to encode monotonicity. While there exists geometric set constructions that are used in the context of range searching [2, 1] we can simply appeal to Theorem 12. We need to define a set to represent predicates P_5 and P_6. The appropriate ground set is over two points $q_j, q_t \in \mathbb{R}^d$, which for notational simplicity we reuse $\mathbb{X}^d_2$. Then the ranges $\mathcal{M}$ are defined by sets $M_r(\overline{st}) \in \mathcal{M}$, defined with respect to radii $r \geq 0$ and line segments $\overline{st}$. Specifically, $M_r(\overline{st}) \subset \mathbb{X}^d_2$ so any $\{q_1, q_2\} \in M_r(\overline{st})$ satisfies that

- $\|p_1 - q_1\| \leq r$ and $\|p_2 - q_2\| \leq r$;
- $p_1, p_2 \in \ell$ where $\overline{st}$ supports ℓ; and
- p_1 is less than p_2 along the line as $\langle p_1, t - s\rangle \leq \langle p_2, t - s\rangle$.

The predicate P_5 is satisfied if $s_j, s_t \in M_r(\overline{q_i q_{i+1}})$ and predicate P_6 is satisfied if $q_i, q_t \in M_r(\overline{s_j s_{j+1}})$.

▶ **Lemma 17.** *The VC dimension of the range space* $(\mathbb{X}^d_2, \mathcal{M})$*, and of the associated dual range space, is* $O(d^2)$*.*

Proof. We may assume that q_1 is the origin, since we can subtract q_1 from all vectors s, t, q_2 using $O(d)$ simple calculations, without changing the outcome. As with bounding the VC dimension of range spaces on $\mathbb{R}^d$ induced by sets $C_r(\overline{st})$, we can derive the closest points on ℓ as $\pi_{\overline{st}}(q_1)$ and $\pi_{\overline{st}}(q_2)$ using $O(d)$ simple operations.

If ℓ is not perpendicular to $q_2 - q_1$, then it intersects the bisector of q_1 and q_2 and we can compute this intersection point as follows:

$$b_\ell(q_1, q_2) := s - \frac{\langle q_2, s - q_2/2\rangle}{\langle t - s, q_2\rangle}(t - s), \tag{1}$$

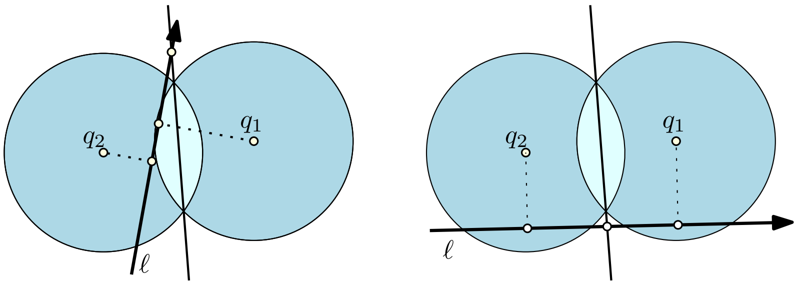

Figure 2 Illustration of predicate P_5 with line ℓ and the two disks centered at q_1 and q_2. In these examples, the projection of q_2 onto ℓ appears before the projection of q_1 onto ℓ along the direction of ℓ and the intersection of ℓ with the bisector lies outside of the lens formed by the two disks. On the left, the predicate is satisfied by setting $p_1 = p_2 = \pi_{\overline{st}}(q_1)$. On the right, the predicate evaluates to false.

This takes $O(d)$ simple operations. Now, predicate P_5 can be computed as follows (Predicate P_6 can be computed in the same way):

1: **if** $(\|\pi_{\overline{st}}(q_1) - q_1\|^2 > r^2)$ or $(\|\pi_{\overline{st}}(q_2) - q_2\|^2 > r^2)$ **then**
2:     **return** 0
3: **else if** $\langle \pi_{\overline{st}}(q_1), t - s\rangle \leq \langle \pi_{\overline{st}}(q_2), t - s\rangle$ **then**
4:     **return** 1
5: **else if** $\|\pi_{\overline{st}}(q_1) - q_2\|^2 \leq r^2$ **then**
6:     **return** 1
7: **else if** $\|\pi_{\overline{st}}(q_2) - q_1\|^2 \leq r^2$ **then**
8:     **return** 1
9: **else if** $\langle q_2 - q_1, t - s\rangle \neq 0$ **then**
10:     compute $b_\ell(q_1, q_2)$ using Eq. (1)
11:     **if** $\|b_\ell(q_1, q_2) - q_1\|^2 \leq r^2$ and $\|b_\ell(q_1, q_2) - q_2\|^2 \leq r^2$ **then**
12:         **return** 1
13:     **end if**
14: **else** **return** 0
15: **end if**

Lines 1-4 test if $p_1 = \pi_{\overline{st}}(q_1)$ and $p_2 = \pi_{\overline{st}}(q_2)$ satisfy the predicate. Lines 5-8 test if $p_1 = p_2 = \pi_{\overline{st}}(q_2)$ or $p_1 = p_2 = \pi_{\overline{st}}(q_1)$ satisfies the predicate. Then Line Line 9-12 tests if $p_1 = p_2 = b_\ell(q_1, q_2)$ satisfies the predicate. Otherwise, we conclude that the predicate is not satisfied for any choice of $p_1, p_2 \in \ell$. To see why this is correct, assume the test in line 3 evaluates to false. In this case, the predicate is satisfied only if ℓ intersects the lens formed by the intersection of the two balls centered at q_1, q_2. If the line intersects the bisector of q_1 and q_2 inside the lens, then we will find a satisfying assignment to p_1 and p_2. If the line intersects the bisector outside of the lens, then by convexity the intersection of the line with either ball is completely contained in one of the two halfspaces bounded by the bisector. Therefore we would find a satisfying assignment to p_1 and p_2 among the closest points on the line to q_1 or q_2, if there exists one. See Figure 2 for an example of the last case. ◀

We define sets to correspond with predicates P_5 and P_6:

- $P_5^r(q, s) = \{\{s_j, s_t\} \in V(s) \times V(s) \mid (s_j, s_t) \in M_r(\overline{q_i q_{i+1}})$ and $\overline{q_i q_{i+1}} \in E(q)\}$.
- $P_6^r(q, s) = \{\{q_i, q_t\} \in V(q) \times V(q) \mid (s_i, s_t) \in M_r(\overline{s_j s_{j+1}})$ and $\overline{s_j s_{j+1}} \in E(s)\}$.

▶ **Theorem 18.** *Let $\mathcal{R}_{F,k}$ be the set of all balls, under the Fréchet distance, centered at polygonal curves in $\mathbb{X}^d_k$. The VC dimension of $(\mathbb{X}^d_m, \mathcal{R}_{F,k})$ is $O\left(d^2k^2\log(dkm)\right)$.*

Proof. Due to Lemma 14, if $S \subset \mathbb{X}^d_m$ is a set of t polygonal curves and $q \in \mathbb{X}^d_k$, the set $\{s \in S \mid \mathrm{d}_F(s,q) \le r\}$ is uniquely defined by the sets

$$\bigcup_{s\in S} P_1^r(q,s), \bigcup_{s\in S} P_2^r(q,s), \bigcup_{s\in S} P_3^r(q,s), \bigcup_{s\in S} P_4^r(q,s), \bigcup_{s\in S} P_5^r(q,s), \bigcup_{s\in S} P_6^r(q,s).$$

As in the proof of Theorem 16, the number of all possible sets

$$\left(\bigcup_{r\ge0}\bigcup_{s\in S} P_1(q,s), \bigcup_{r\ge0}\bigcup_{s\in S} P_2(q,s), \bigcup_{r\ge0}\bigcup_{s\in S} P_3(q,s), \bigcup_{r\ge0}\bigcup_{s\in S} P_4(q,s)\right)$$

is bounded by $(tm)^{O(d^2k)}$.

By Lemma 17 we are able to bound the number of all possible sets $\bigcup_{r\ge0}\bigcup_{s\in S} P_5^r(q,s)$ as $(tm)^{O(d^2k^2)}$. The k^2 term arises because we consider $\Theta(k^2)$ pairs s_j, s_t for predicate P_5. And because this bound is proven using Theorem 12, then it applies to the dual range space, and we also bound the number of possible sets in $\bigcup_{r\ge0}\bigcup_{s\in S} P_6^r(s,q)$ as also $(tm)^{O(d^2k^2)}$. So ultimately,

$$2^t \le (tm)^{O(d^2k^2)} \implies t = O\left(d^2k^2\log(dkm)\right). \qquad ◀$$

8 Hausdorff distance

We consider the range space $(\mathbb{X}^d_m, \mathcal{R}^r_{H_k})$, where $\mathcal{R}^r_{H_k}$ denotes the set of all balls, of radius r centered at curves in $\mathbb{X}^d_k$, under the symmetric Hausdorff distance.[3] We also consider the same problems under both directed versions of the Hausdorff distance, and their induced range spaces $(\mathbb{X}^d_m, \mathcal{R}^r_{\overrightarrow{H}_k})$ and $(\mathbb{X}^d_m, \mathcal{R}^r_{\overleftarrow{H}_k})$. While some intermediate arguments hold in $\mathbb{R}^d$, we are only able to provide VC dimension bounds in $\mathbb{R}^2$. Proofs can be found in the full version of the paper [17].

▶ **Theorem 19.** *Let $\overrightarrow{\mathcal{R}}_{H,k}$ be the set of all balls, under the directed Hausdorff distance* from *polygonal curves in $\mathbb{X}^2_k$. The VC dimension of $(\mathbb{X}^2_m, \overrightarrow{\mathcal{R}}_{H,k})$ is $O(k^2\log(km))$.*

▶ **Theorem 20.** *Let $\overleftarrow{\mathcal{R}}_{H,k}$ be the set of all balls, under the directed Hausdorff distance* to *polygonal curves in $\mathbb{X}^2_k$. The VC dimension of $(\mathbb{X}^2_m, \overleftarrow{\mathcal{R}}_{H,k})$ is $O(k\log(km))$.*

▶ **Theorem 21.** *Let $\mathcal{R}_{H,k}$ be the set of all balls, under the symmetric Hausdorff distance in $\mathbb{X}^2_k$. The VC dimension of $(\mathbb{X}^2_m, \mathcal{R}_{H,k})$ is $O(k^2\log(km))$.*

9 Lower bounds

Our lower bounds are constructed in the simplified setting that either $k = 1$ or $m = 1$, i.e., either the ground set or the curves defining the metric ball consist of one vertex only. In this case, all of our considered distance measures (except for one direction of the directed

[3] The proofs in this section are written for polygonal curves in $\mathbb{X}^d_m$ (resp. $\mathbb{X}^2_m$), but they readily extend to (not-necessarily connected) sets of line segments in $\mathbb{R}^d$ (resp. $\mathbb{R}^2$) of cardinality $m' = \frac{m-1}{2}$.

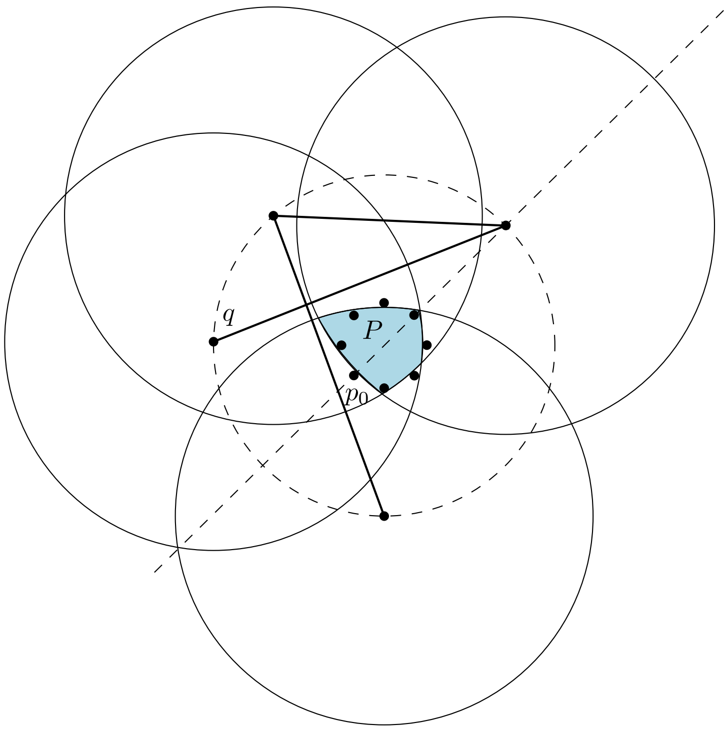

Figure 3 A curve q with metric ball of radius $R - \varepsilon$ containing a subset of P. The shaded area is the set of points that are contained inside the metric ball.

Hausdorff distance) are equal. The basic idea behind our lower bound construction is that, for $m = 1$, the ranges behave like convex polygons with k facets. In particular, the set of points contained inside the range centered at a curve q, is equal to the intersection of a set of equal-size Euclidean balls centered at the vertices of q. Figure 3 shows a sketch of the construction. For $k = 1$ and $m \geq 1$, we use [24, Lemma 5.18], which bounds the VC dimension of the dual range space as a function of the VC dimension of the primal space. Proofs of the lower bounds can be found in the full version of the paper [17].

▶ **Theorem 22.** *The VC-dimension of the range spaces* $(\mathbb{X}_m^2, \mathcal{R}_{dF,k})$, $(\mathbb{X}_m^2, \mathcal{R}_{dH,k})$, $(\mathbb{X}_m^2, \mathcal{R}_{F,k})$, $(\mathbb{X}_m^2, \mathcal{R}_{wF,k})$, *and* $(\mathbb{X}_m^2, \mathcal{R}_{H,k})$*is* $\Omega(\max(k, \log m))$.

▶ **Theorem 23.** *For* $d \geq 4$, *the VC-dimension of the range spaces* $(\mathbb{X}_m^d, \mathcal{R}_{dF,k})$, $(\mathbb{X}_m^d, \mathcal{R}_{dH,k})$, $(\mathbb{X}_m^d, \mathcal{R}_{wF,k})$, $(\mathbb{X}_m^d, \mathcal{R}_{F,k})$, *and* $(\mathbb{X}_m^d, \mathcal{R}_{H,k})$ *is* $\Omega(\max(dk \log k, \log dm))$.

10 Implications

In this section we demonstrate that bounds on the VC-dimension for the range space defined by metric balls on curves immediately implies various results about prediction and statistical generalization over the space of curves. In the following consider a range space $(X, \mathcal{R})$ with a ground set of X of curves of size $n = |X|$, where $\mathcal{R}$ are the ranges corresponding to metric balls for some distance measure we consider, and the VC-dimension is bounded by ν. This section discusses accuracy bounds that depend directly on the size $|X| = n$ and the VC-dimension ν. They will assume that X is a random sample of some much larger set X_{big} or an unknown continuous generating distribution μ. Under the randomness in this assumed sampling procedure, there is a probability of failure δ that often shows up in these bounds, but is minor since it shows up as $\log(1/\delta)$. These bounds take two closely-linked forms. First, given a limited set X from an unknown μ, then how accurate is a query or a prediction made using only X. Second, given the ability to draw samples (at a cost) from an unknown distribution μ, how many are required so the prediction on the samples set X has bounded prediction error. The theme of the following results, as implied by our above VC-dimension results, is that if these families of curves are only inspected with or queried with curves with a small number of segments (k is small), then the VC-dimension of the associated range space $\nu = O(k \log km)$ or $O(k^2 \log km)$ is small, and that such analyses generalize well. We show this in several concrete examples. More examples are detailed in the full version of the paper [17].

Approximate range counting on curves. Given a large set of curves X (of potentially very large complexity m), and a query curve q (with smaller complexity k) we would like to approximate the number of curves nearby q. For instance, we restrict X to historical queries at a certain time of day, and query with the planned route q, and would like to know the chance of finding a carpool. VC-dimension ν of the metric balls shows up directly in two analyses. First, if we assume $X \sim \mu$ where μ is a much larger unknown distribution (but the real one), then we can estimate the accuracy of the fraction of all curves in this range within additive error $O(\sqrt{(1/|X|)(\nu + \log(1/\delta))})$. On the other hand, if X is too large to conveniently query, we can sample a subset $S \subset X$ of size $O((1/\varepsilon^2)(\nu + \log(1/\delta)))$ and know that the estimate for the fraction of curves from S in that range is within additive ε error of the fraction from X. Such sampling techniques have a long history in traditional databases [29], and have more recently become important when providing online estimates during a long query processing time as incrementally increasing size subsets are considered [3]. Ours provides the first formal analysis of these results for queries over curves.

Density estimation of curves. A related task in generalization to new curves is density estimation. Consider a large set of curves X which represent a larger unknown distribution μ that models a distribution of curves; we want to understand how unusual a new curve q would be given we have not yet seen exactly the same curve before. One option is to use the distance to the (kth) nearest neighbor curve in X, or a bit more robust option is to choose a radius r, and count how many curves are within that radius (e.g., the approximate range counting results above). Alternatively, for $X \subset \mathbb{M}$, consider now a kernel density estimate $\textsc{kde}_X : \mathbb{M} \to \mathbb{R}$ defined by $\textsc{kde}_X(p) = \frac{1}{n}\sum_{p \in P} K(x,p)$ with kernel $K(x,p) = \exp(-\mathrm{d}(x,p)^2)$ (where d is some distance of choice among curves, e.g., d_F). The kernel is defined such that each superlevel set $K_x^\tau = \{p \in \mathbb{M} \mid K(x,p) \geq \tau\}$ corresponds with some range $R \in \mathcal{R}$ so that $R \cap X = K_x^\tau \cap X$. Then a random sample $S \subset X$ of size $O((1/\varepsilon^2)(\nu + \log \frac{1}{\delta}))$ satisfies that $\| \textsc{kde}_X - \textsc{kde}_s \|_\infty \leq \varepsilon$ [26]. Thus, again the VC-dimension ν of the metric balls directly influences this estimates accuracy, and for query curves with small complexity k, the bound is quite reasonable.

References

1 Peyman Afshani and Anne Driemel. On the complexity of range searching among curves. *CoRR*, arXiv:1707.04789v1, 2017. arXiv:1707.04789.

2 Peyman Afshani and Anne Driemel. On the complexity of range searching among curves. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*

2018, New Orleans, LA, USA, January 7-10, 2018, pages 898–917, 2018. doi:10.1137/1.9781611975031.58.

3 S. Agarwal, B. Mozafari, A. Panda, H. Milner, S. Madden, and I. Stoica. BlinkDB: queries with bounded errors and bounded response times on very large data. In *EuroSys*, 1993.

4 Yohji Akama, Kei Irie, Akitoshi Kawamura, and Yasutaka Uwano. VC Dimension of Principal Component Analysis. *Discrete & Computational Geometry*, 44:589–598, 2010.

5 Helmut Alt, Bernd Behrends, and Johannes Blömer. Approximate matching of polygonal shapes. *Annals of Mathematics and Artificial Intelligence*, 13(3):251–265, September 1995. doi:10.1007/BF01530830.

6 Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

7 Maria Astefanoaei, Paul Cesaretti, Panagiota Katsikouli, Mayank Goswami, and Rik Sarkar. Multi-resolution sketches and locality sensitive hashing for fast trajectory processing. In *International Conference on Advances in Geographic Information Systems (SIGSPATIAL 2018)*, volume 10, 2018.

8 Julian Baldus and Karl Bringmann. A Fast Implementation of Near Neighbors Queries for FréChet Distance (GIS Cup). In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'17, pages 99:1–99:4, 2017.

9 Anselm Blumer, A. Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the ACM*, 36:929–965, 1989.

10 Hervé Brönnimann and Michael T. Goodrich. Almost Optimal Set Covers in Finite VC-Dimension. *Discrete & Computational Geometry*, 1995.

11 Kevin Buchin, Yago Diez, Tom van Diggelen, and Wouter Meulemans. Efficient trajectory queries under the Fréchet distance (GIS Cup). In *Proc. 25th Int. Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 101:1–101:4, 2017.

12 Bernard Chazelle and Emo Welzl. Quasi-Optimal Range Searching in Spaces of Finite VC-Dimension. *Discrete and Computational Geometry*, 4:467–489, 1989.

13 Monika Csikos, Andrey Kupavskii, and Nabil H. Mustafa. Optimal Bounds on the VC-dimension. *arXiv:1807.07924*, 2018. arXiv:1807.07924.

14 Mark De Berg, Atlas F Cook, and Joachim Gudmundsson. Fast Fréchet queries. *Computational Geometry*, 46(6):747–755, 2013.

15 Mark de Berg and Ali D. Mehrabi. Straight-Path Queries in Trajectory Data. In *WALCOM: Algorithms and Computation - 9th Int. Workshop, WALCOM 2015, Dhaka, Bangladesh, February 26-28, 2015. Proceedings*, pages 101–112, 2015. doi:10.1007/978-3-319-15612-5_10.

16 Anne Driemel, Amer Krivošija, and Christian Sohler. Clustering time series under the Fréchet distance. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 766–785, 2016. doi:10.1137/1.9781611974331.ch55.

17 Anne Driemel, Jeff M. Phillips, and Ioannis Psarros. The VC dimension of metric balls under Fréchet and Hausdorff distances. *CoRR*, arXiv:1903.03211, 2019. arXiv:1903.03211.

18 Anne Driemel and Francesco Silvestri. Locally-sensitive hashing of curves. In *33st International Symposium on Computational Geometry, SoCG 2017*, pages 37:1–37:16, 2017.

19 Fabian Dütsch and Jan Vahrenhold. A Filter-and-Refinement- Algorithm for Range Queries Based on the Fréchet Distance (GIS Cup). In *Proc. 25th Int. Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, pages 100:1–100:4, 2017.

20 Ioannis Z. Emiris and Ioannis Psarros. Products of Euclidean Metrics and Applications to Proximity Questions among Curves. In *Proc. 34th Int. Symposium on Computational Geometry (SoCG)*, volume 99 of *LIPIcs*, pages 37:1–37:13, 2018.

21 Alexander Gilbers and Rolf Klein. A new upper bound for the VC-dimension of visibility regions. *Computational Geometry: Theory and Applications*, 74:61–74, 2014.

22 Paul W. Goldberg and Mark R. Jerrum. Bounding the Vapnik-Chervonenkis Dimension of Concept Classes Parameterized by Real Numbers. *Machine Learning*, 18:131–148, 1995.

23 Joachim Gudmundsson and Michiel Smid. Fast algorithms for approximate Fréchet matching queries in geometric trees. *Computational Geometry*, 48(6):479–494, 2015. doi:10.1016/j.comgeo.2015.02.003.

24 Sariel Har-Peled. *Geometric Approximation Algorithms*. American Mathematical Society, Boston, MA, USA, 2011.

25 Lingxiao Huang, Shaofeng Jiang, Jian Li, and Xuan Wu. Epsilon-Coresets for Clustering (with Outliers) in Doubling Metrics. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 814–825. IEEE, 2018.

26 Sarang Joshi, Raj Varma Kommaraju, Jeff M. Phillips, and Suresh Venkatasubramanian. Comparing Distributions and Shapes Using the Kernel Distance. In *ACM SoCG*, 2011.

27 Marek Karpinski and Angus Macintyre. Polynomial bounds for VC dimension of sigmoidal neural networks. In *STOC*, 1995.

28 Elmar Langetepe and Simone Lehmann. Exact VC-dimension for L1-visibility of points in simple polygons. *arXiv:1705.01723*, 2017. arXiv:1705.01723.

29 Frank Olken. *Random Sampling in Databases*. PhD thesis, University of California at Berkeley, 1993.

30 Norbert Sauer. On the Density of Families of Sets. *Journal of Combinatorial Theory Series A*, 13:145–147, 1972.

31 Saharon Shelah. A Combinatorial Problem; Stability and Order for Models and Theories in Infinitary Languages. *Pacific Journal of Mathematics*, 41(1), 1972.

32 Pavel Valtr. Guarding Galleries Where No Point Sees a Small Area. *Israel Journal of Mathematics*, 104:1–16, 1998.

33 Vladimir Vapnik and Alexey Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to their Probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.

34 Vladimir N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.

35 Martin Werner and Dev Oliver. ACM SIGSPATIAL GIS Cup 2017: Range queries under Fréchet distance. *SIGSPATIAL Special*, 10(1):24–27, June 2018. doi:10.1145/3231541.3231549.