

# Simplified and Improved Bounds on the VC-Dimension for Elastic Distance Measures

**Frederik Brüning** Anne Driemel

University of Bonn

Ioannina, March 13, 2024

Range space  $R$  with ground set  $X$ :

Set  $R$  s.t. any  $r \in R$  is of the form  $r \cap X$

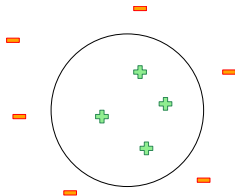
Example:

Balls with ground set  $X = \mathbb{R}^2$ :

$$R = \{b(c; r) \mid c \in \mathbb{R}^2, r \in \mathbb{R}_+\}$$

where

$$b(c; r) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq r\}$$



# VC-Dimension of a Range Space

Range space  $R$  with ground set  $X$ :

Set  $R$  s.t. any  $r \in R$  is of the form  $r \subseteq X$

Shattering:

$A \subseteq X$  is shattered by  $R$  ( $\cdot$ )

$\exists A^0 \subseteq A \exists r \in R$  s.t.  $A^0 = r \cap A$

(all subsets of  $A$  can be realized by ranges in  $R$ )

$VC\text{-dim}(R)$ :

Maximal size of a shattered subset  $A \subseteq X$

Example:

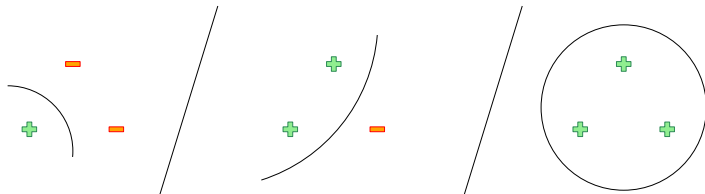
Balls with ground set  $X = \mathbb{R}^2$ :

$$R = \{b(c; r) \mid c \in \mathbb{R}^2, r \in \mathbb{R}_+\}$$

where

$$b(c; r) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq r\}$$

$VC\text{-dim}(R) = 3$



# VC-Dimension of a Range Space

Range space  $R$  with ground set  $X$ :

Set  $R$  s.t. any  $r \in R$  is of the form  $r = X$

Shattering:

$A \subseteq X$  is shattered by  $R$  ( $\cdot$ )

$\exists A^0 \subseteq A \exists r \in R$  s.t.  $A^0 = r \cap A$

(all subsets of  $A$  can be realized by ranges in  $R$ )

$VC\text{-dim}(R)$ :

Maximal size of a shattered subset  $A \subseteq X$

Example:

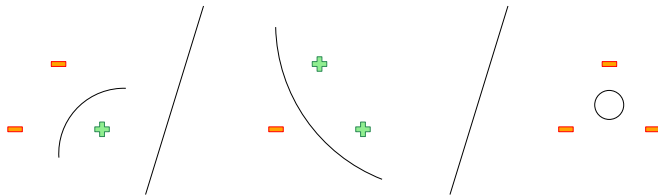
Balls with ground set  $X = \mathbb{R}^2$ :

$$R = \{b(c; r) \mid c \in \mathbb{R}^2, r \in \mathbb{R}_+\}$$

where

$$b(c; r) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq r\}$$

$VC\text{-dim}(R) = 3$



# VC-Dimension of a Range Space

Range space  $R$  with ground set  $X$ :

Set  $R$  s.t. any  $r \in R$  is of the form  $r \subseteq X$

Shattering:

$A \subseteq X$  is shattered by  $R$  ( $\cdot$ )

$\exists A^0 \subseteq A \exists r \in R$  s.t.  $A^0 = r \cap A$

(all subsets of  $A$  can be realized by ranges in  $R$ )

VC-dim( $R$ ):

Maximal size of a shattered subset  $A \subseteq X$

Example:

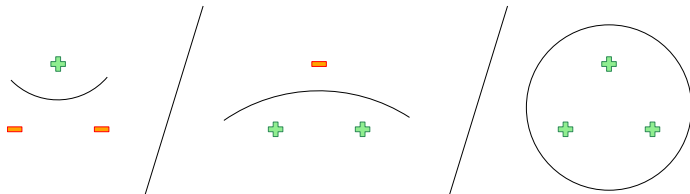
Balls with ground set  $X = \mathbb{R}^2$ :

$$R = \{b(c; r) \mid c \in \mathbb{R}^2, r \in \mathbb{R}_+\}$$

where

$$b(c; r) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq r\}$$

VC-dim( $R$ ) = 3



# VC-Dimension of a Range Space

Range space  $R$  with ground set  $X$ :

Set  $R$  s.t. any  $r \in R$  is of the form  $r \subseteq X$

Shattering:

$A \subseteq X$  is shattered by  $R$  ( $\cdot$ )

$\exists A^\theta \subseteq A \exists r \in R$  s.t.  $A^\theta = r \cap A$

(all subsets of  $A$  can be realized by ranges in  $R$ )

VC-dim( $R$ ):

Maximal size of a shattered subset  $A \subseteq X$

Example:

Balls with ground set  $X = \mathbb{R}^2$ :

$$R = \{b(c; r) \mid c \in \mathbb{R}^2, r \in \mathbb{R}_+\}$$

where

$$b(c; r) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq r\}$$

VC-dim( $R$ ) = 3



# Motivation: What are VC-Dimension bounds used for?

## Sample bounds for computational tasks:

- $\epsilon$ -nets, relative-error  $(p_i)$ -approximations
- test error of classification model

## Applications:

- kernel density estimation
- coresets
- clustering
- object recognition
- ...

Range spaces:

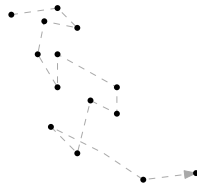
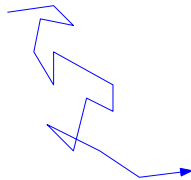
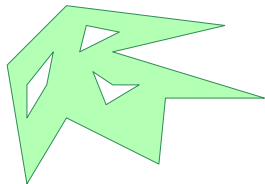
$$d = \text{distance measure on } (\mathbb{R}^d)^m$$
$$b(c; \cdot) = \{x \in (\mathbb{R}^d)^m \mid d(x; c) \leq \rho\}$$
$$R_{\rho; k} = \{b(c; \cdot) \mid c \in (\mathbb{R}^d)^k\}$$

Distance measures:

- Hausdorff
- (weak) Fréchet
- Dynamic Time Warping

Ground set  $((\mathbb{R}^d)^m)$ /Centers  $((\mathbb{R}^d)^k)$ :

- Polygonal regions
- continuous polygonal curves
- discrete polygonal curves





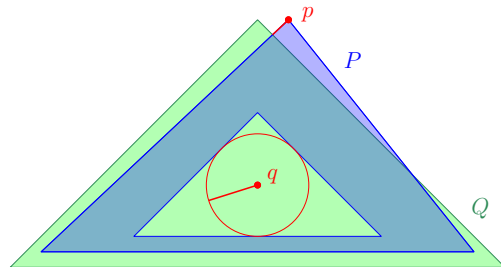
directed Hausdorff distance:

$$d_H(P; Q) = \sup_{p \in P} \inf_{q \in Q} \|p - q\|$$

Hausdorff distance:

$$d_H(P; Q) = \max\{d_H(P; Q); d_H(Q; P)\}$$

for  $P, Q \subset \mathbb{R}^d$



		new	old
discrete polygonal curves	DTW	$O(dk^2 \log(m))$	-
		$O(dkm \log(k))$	
	Hausdorff	$O(dk \log(km))$	$O(dk \log(dkm))$
Fréchet	$O(dk \log(km))^{( )}$		
continuous polygonal curves	Hausdorff	$O(dk \log(km))$	$O(d^2 k^2 \log(dkm))$
	Fréchet	$O(dk \log(km))^{( )}$	
	weak Fréchet	$O(dk \log(km))^{( )}$	$O(d^2 k \log(dkm))$
<b>polygons <math>\mathbb{R}^2</math></b>	<b>Hausdorff</b>	<b><math>O(k \log(km))</math></b>	-

**Table:** Overview of VC-dimension bounds. Results marked with ( ) were independently obtained by Cheng and Huang [2024]. The old results were obtained by Driemel, Nusser, Phillips and Psarros [2021].

Lower bound (Driemel, Nusser, Phillips and Psarros [2021]):

$(\max(dk \log(k); \log(dm)))$  for  $d \geq 4$  for polygonal curves.

# General Approach

## Approach:

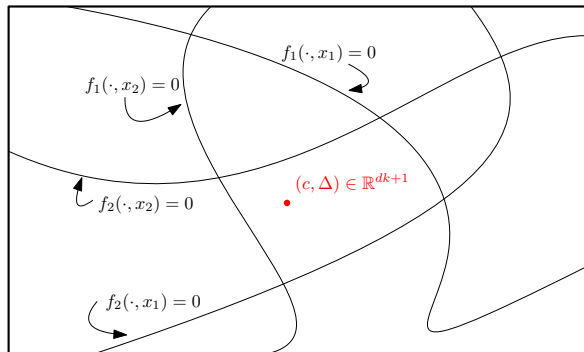
- Split query  $d_H(P; Q)$  into predicates
- Express predicates as combinations of sign values of polynomials
- Bound VC-dim based on number of cells in arrangement of zero sets of polynomials.

## Definition:

$F$ : Class of polynomials of constant degree from  $\mathbb{R}^{dk+1} \times \mathbb{R}^{dm}$  to  $\mathbb{R}$

$R$  is  $t$ -combination of  $\text{sgn}(F)$ :  $\exists$  boolean function  $g$ ,  $\exists f_1; \dots; f_t \in F$  s.t.  $\exists r \in R \exists y$  s.t.

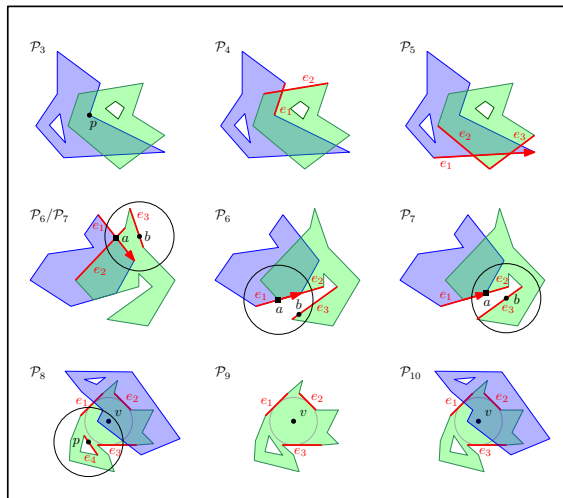
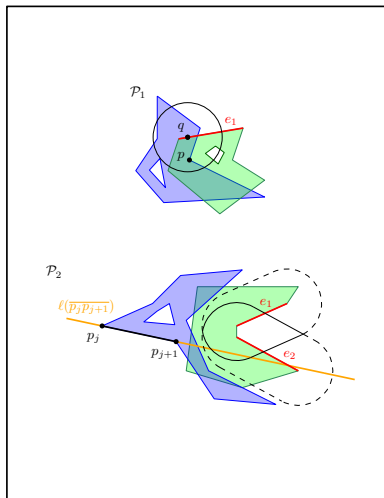
$$r = f(x) \wedge \exists y \left( g(\text{sgn}(f_1(y; x)); \dots; \text{sgn}(f_t(y; x))) = 1 \right)$$



## Theorem (Anthony and Bartlett 1999):

Suppose  $R$  is a  $t$ -combination of  $\text{sgn}(F)$ . Then  $\text{VC-dim}(R)$  is in  $O(dk \log(t))$ .

*Idea goes back to Goldberg and Jerrum [1993] and independently Ben-David and Lindenbaum [1993].*

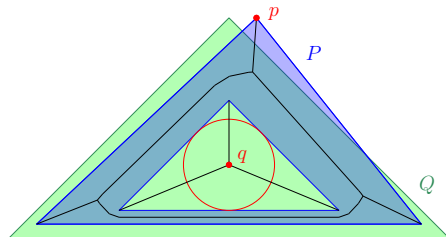


Predicates  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are by Driemel, Nusser, Phillips and Psarros [2021]

## Cases:

$d_H(P; Q)$  maximized at point  $p$  at the boundary of  $P$

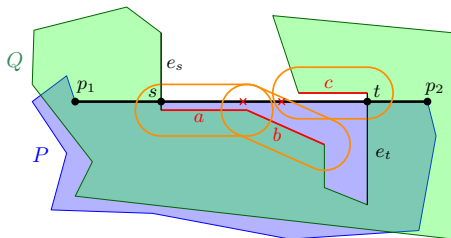
$d_H(Q; P)$  maximized at point  $q$  in the interior of  $Q$



## Predicates:

- $(B)$  (Boundary): True  $(\ )$   $d_H(@P; Q)$  .
- $(I)$  (Interior): True if  $d_H(P; Q)$  . False if  $d_H(P; Q) > d_H(@P; Q)$  and  $d_H(P; Q) >$  .

$d_H(P; Q)$   $(\ )$   $(B)$  and  $(I)$  true



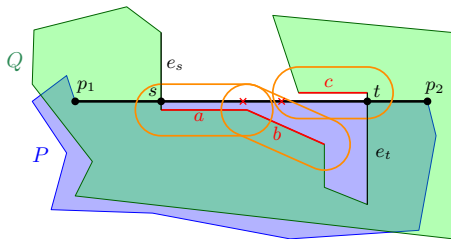
( $B$ ) (Boundary): True ( )  $d_H(@P; Q)$

$d_H(@P; Q)$  ( )  $d_H(e; Q)$  for every edge  $e$  of  $P$

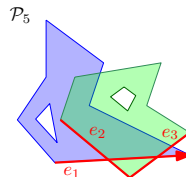
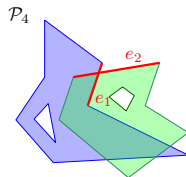
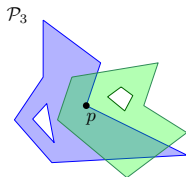
- Find for each  $e$  the part that is outside of  $Q$  (here  $\overline{st}$ )
- Find sequence of edges of  $Q$  such that these parts are included in their stadiums (here  $a; b; c$ )

# Predicate (B)

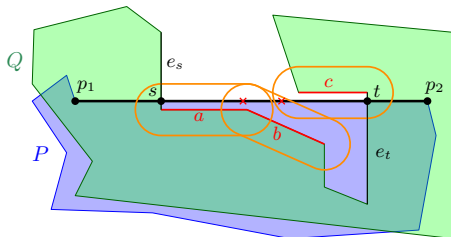
Find part of  $e$  that is outside of  $Q$  (here  $\overline{st}$ )



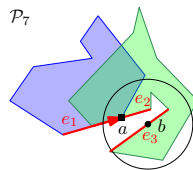
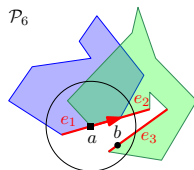
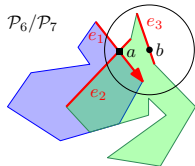
- $P_3: p \in Q?$
- $P_4: e_1 \setminus e_2 \notin ?$
- $P_5: e_1$  intersects  $e_2$  before  $e_3$ ?



Find stadiums that include  $s$  and  $t$  (here  $a$  and  $c$ )

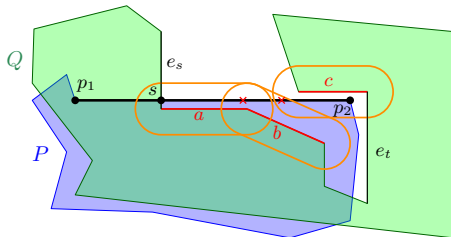


- $P_6 = P_7$ :  $9b$  on  $e_3$  with  $kb - ak$  ?
- $P_6$  :  $a$  is first point on  $e_1$  in  $e_1 \setminus e_2$
- $P_7$  :  $a$  is last point on  $e_1$  in  $e_1 \setminus e_2$

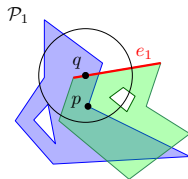




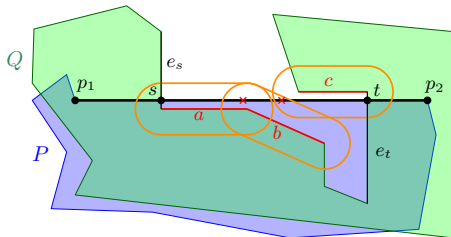
Find stadiums that include  $s$  and  $p_2$  (here  $a$  and  $c$ )



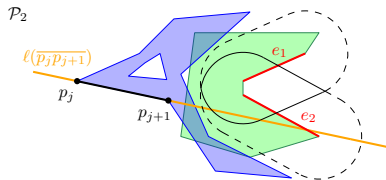
- $\mathcal{P}_1$ :  $\exists q$  on  $e_1$  with  $kp - qk$  ?



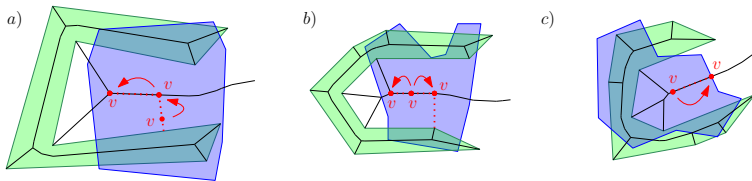
Find sequence of stadiums that include  $\overline{st}$  (here  $a; b; c$ )



- $D_{;2}(e_1; e_2)$ : intersection of stadiums around  $e_1; e_2$
- $P_2: \overline{(p_j p_{j+1})} \setminus D_{;2}(e_1; e_2) \notin ;?$

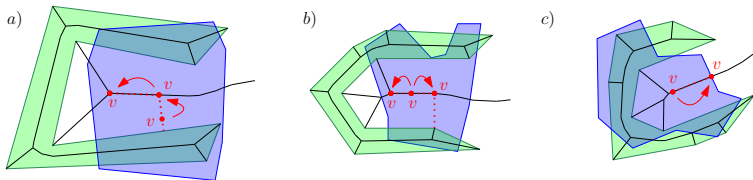


(I) (Distance realized in interior): We only check vertices of Voronoi diagram of edges of  $Q$ .



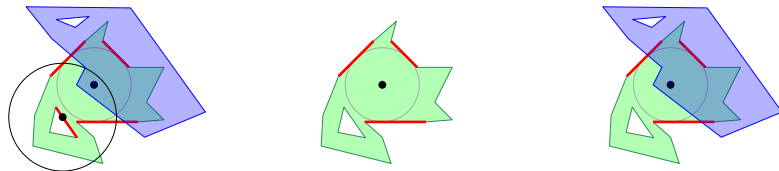
# Predicate (I)

(I) (Distance realized in interior): We only check vertices of Voronoi diagram of edges of  $Q$ .



We check distances of all Voronoi vertices to all edges of  $Q$ .

A Voronoi vertex is relevant if it is inside of  $P$  and outside of  $Q$ .



- $P_8$ :  $\exists p$  on  $e_4$  with  $kv - pk$  ?
- $P_9$ :  $v \in Q$ ?
- $P_{10}$ :  $v \in P$ ?

### Lemma

For any two polygonal regions  $P$  and  $Q$  (that may contain holes), given the truth values of all predicates of the type  $P_1; \dots; P_{10}$  one can determine whether  $d_H(P; Q)$  .

### Remaining step:

Express predicates  $P_1; \dots; P_{10}$  as combinations of sign values of polynomials.

		new	old
discrete polygonal curves	DTW	$O(dk^2 \log(m))$	-
		$O(dkm \log(k))$	
	Hausdorff	$O(dk \log(km))$	$O(dk \log(dkm))$
Fréchet	$O(dk \log(km))^{( )}$		
continuous polygonal curves	Hausdorff	$O(dk \log(km))$	$O(d^2 k^2 \log(dkm))$
	Fréchet	$O(dk \log(km))^{( )}$	
	weak Fréchet	$O(dk \log(km))^{( )}$	$O(d^2 k \log(dkm))$
polygons $\mathbb{R}^2$	<b>Hausdorff</b>	$O(k \log(km))$	-

**Table:** Overview of VC-dimension bounds. Results marked with ( ) were independently obtained by Cheng and Huang [2024]. The old results were obtained by Driemel, Nusser, Phillips and Psarros [2021]

Lower bound (Driemel, Nusser, Phillips and Psarros [2021]):

$(\max(dk \log(k); \log(dm)))$  for  $d \geq 4$  for polygonal curves.