



Uniform Brackets, Containers, and Combinatorial Macbeath Regions

Kunal Dutta   

Department of Informatics, University of Warsaw, Poland

Arijit Ghosh  

Indian Statistical Institute, Kolkata, India

Shay Moran   

Technion – Israel Institute of Technology, Haifa, Israel

Google Research, Tel Aviv, Israel

Abstract

We study the connections between three seemingly different combinatorial structures – *uniform brackets* in statistics and probability theory, *containers* in online and distributed learning theory, and *combinatorial Macbeath regions*, or *Mnets* in discrete and computational geometry. We show that these three concepts are manifestations of a single combinatorial property that can be expressed under a unified framework along the lines of Vapnik-Chervonenkis type theory for uniform convergence. These new connections help us to bring tools from discrete and computational geometry to prove improved bounds for these objects. Our improved bounds help to get an optimal algorithm for distributed learning of halfspaces, an improved algorithm for the distributed convex set disjointness problem, and improved regret bounds for online algorithms against σ -smoothed adversary for a large class of semi-algebraic threshold functions.

2012 ACM Subject Classification Theory of computation

Keywords and phrases communication complexity, distributed learning, empirical process theory, online algorithms, discrete geometry, computational geometry

Digital Object Identifier 10.4230/LIPIcs.ITCS.2022.59

Related Version *Full Version*: <https://arxiv.org/abs/2111.10048>

Funding *Kunal Dutta*: Supported by the Polish NCN SONATA Grant no. 2019/35/D/ST6/04525. *Shay Moran*: Shay Moran is a Robert J. Shillman Fellow and his research is supported by the ISF, grant no. 1225/20, by an Azrieli Faculty Fellowship, and by BSF grant 2018385.

1 Introduction

A particularly pleasing situation in theoretical studies is when seemingly independent notions arising in disparate areas with different applications and techniques, turn out to have a common theoretical basis. In this article, we study a combinatorial notion whose manifestations appear in three different areas as distinct combinatorial objects – as *uniform brackets* in statistical learning and empirical process theory, *containers* in online and distributed learning theory, and *Combinatorial Macbeath regions*, or *Mnets* in discrete and computational geometry – and show that these are consequences of an underlying combinatorial property. The close connection between uniform brackets and containers has been known [14, 22]. We connect these notions with Mnets, which are discrete analogues of a classical theorem of Macbeath in convex geometry. This allows us to import tools from discrete and computational geometry to solve problems and improve bounds in each of these areas, in some cases proving optimal new bounds.

As we aim to keep this paper accessible to readers from all three communities, we begin with a brief introduction to the notions involved. Given a probability space $(\mathcal{X}, \Omega, \mu)$, together with a family \mathcal{H} of measurable sets in Ω and a parameter $\varepsilon \in (0, 1)$, an ε -*uniform bracket*, or



© Kunal Dutta, Arijit Ghosh, and Shay Moran;

licensed under Creative Commons License CC-BY 4.0

13th Innovations in Theoretical Computer Science Conference (ITCS 2022).

Editor: Mark Braverman; Article No. 59; pp. 59:1–59:10

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

ε -bracket for short, for \mathcal{H} is a family \mathcal{B} of measurable sets such that for every $H \in \mathcal{H}$, there exist sets $A, B \in \mathcal{B}$ with

$$A \subseteq H \subseteq B \text{ and } \mu(B \setminus A) \leq \varepsilon.$$

The ε -bracketing number $N_{[\cdot]}(\mathcal{H}, \mu, \varepsilon)$ of \mathcal{H} with respect to the measure μ , is the smallest possible size of an ε -bracket for \mathcal{H} . The logarithm of $N_{[\cdot]}(\mathcal{H}, \mu, \varepsilon)$ is referred to as the *bracketing entropy*.

For a set system (X, \mathcal{F}) , where X is finite and $\mathcal{F} \subseteq 2^X$, a family of subsets \mathcal{B} of X is an ε -bracket if for all $F \in \mathcal{F}$ there exist B^+ and B^- in \mathcal{B} such that

$$B^- \subseteq F \subseteq B^+ \text{ and } |B^+ \setminus B^-| \leq \varepsilon|X|.$$

The significance of the bracketing number in empirical process theory stems from the fact that bounds on $N_{[\cdot]}(\mathcal{H}, \mu, \varepsilon)$ can be used to obtain simpler and more robust versions of uniform convergence and the law of large numbers for the corresponding families of events. In particular, the proof of uniform convergence using ε -brackets follows directly from standard concentration inequalities together with a union bound, and does not require the *symmetrization* trick of Vapnik and Chervonenkis [35]. Thus uniform convergence for families of bounded bracketing number, holds even when the point sample X is generated using non-*i.i.d.* processes. Recently, ε -brackets were used by Haghtalab, Roughgarden and Shetty [22] for the smoothed analysis of online and differentially private learning algorithms. For a more comprehensive introduction to these topics, we refer the reader to [1, 34].

Containers were recently introduced by Braverman, Kol, Moran and Saxena [14] to study the communication complexity of distributed learning problems. The choice of the term containers was inspired by the related notion of *containers for independent sets* in hypergraphs [10, 32]. Given a set system (X, \mathcal{F}) consisting of a ground set X and a family of subsets $\mathcal{F} \subseteq 2^X$, together with a parameter $\varepsilon \in (0, 1)$, an ε -container \mathcal{C} is a collection of subsets of X such that for every set $F \in \mathcal{F}$, there exists a member $C \in \mathcal{C}$ such that $F \subset C$ and $|C \setminus F| \leq \varepsilon n$. A set system of points and halfspaces in \mathbb{R}^d has a set X of points in \mathbb{R}^d and the collection \mathcal{F} as all possible subsets of X which can be generated via intersection with a halfspace in \mathbb{R}^d . Braverman, Kol, Moran and Saxena [14] proved a new dual version of the classical *Carathéodory's theorem* for points and halfspaces in \mathbb{R}^d , and used it to show that systems of points and halfspaces in \mathbb{R}^d have ε -containers of size $O((d/\varepsilon)^d)$. This allowed them to design improved protocols for bounding the communication complexity of learning problems such as distributed learning of halfspaces and distributed linear programming.

A classical theorem of Macbeath [26] in convex geometry states that for any $\varepsilon \in (0, 1)$, every convex body in \mathbb{R}^d of unit volume contains a collection of subsets, each of volume $\Omega(\varepsilon)$, such that any halfspace intersecting at least an ε -volume of the body must contain at least one of the subsets from the collection. Since its introduction Macbeath regions have been an important object of study in convex geometry [11, 12]. More recently, Macbeath regions were used for proving data structure lower bounds [15, 8], and convex body approximation problems in computational geometry [6, 7, 5]. *Mnets* were proposed as combinatorial analogues of Macbeath's theorem by Mustafa and Ray [28], who showed their existence for many geometrically defined classes of set systems. Later their result was generalized to hold for *semi-algebraic* set systems with bounded *shallow cell complexity* by Dutta, Ghosh, Jartoux and Mustafa [19]. A set system (X, \mathcal{S}) is said to have a λ -heavy ε -Mnet, if there exists a collection \mathcal{M} of subsets of X such that for any set $S \in \mathcal{S}$ with at least $\varepsilon|X|$ elements, there exists a member M of \mathcal{M} which is contained in S , and has at least $\lambda|S|$ elements. Mnets can be used to prove the existence of optimal-sized ε -nets for almost all studied classes of geometric set systems [19].

We will now present an outline of our results and their applications, together some previous works. For the explicit statements of our results and their proofs please refer to the full version of our paper [20].

2 Related work and outline of our results

Our contribution may be thought of as having two components – a conceptual component and a technical one. Conceptually, our main contribution is to find the connection between three combinatorial concepts – ε -brackets, ε -containers, and Mnets. Roughly, we show that the existence of any one of these structures in a set system implies the existence of the other two in the system or in its complement. To quantify these connections, we introduce the notion of Property \mathbb{M} , which essentially represents the existence of Mnets of bounded size in a set system. The definition of Property \mathbb{M} and the explicit connections are presented in the full version of the paper [20].

Our technical contribution is to exploit these connections to prove several new results improving existing bounds as well as finding new applications for Mnets, ε -containers and ε -brackets. These include improved bounds on the size of ε -brackets and ε -containers with optimal dependence on the ambient dimension and showing the existence of λ -heavy ε -Mnets for arbitrary λ . We proceed to give several applications of our improved bounds, such as in distributed learning of halfspaces and distributed linear programming, and the smoothed analysis of online and differentially private learning. We also extend the results of [28, 19], who showed the existence of Λ -heavy Mnets for a fixed $\Lambda \in (0, 1/2)$, to show the existence of λ -heavy Mnets for any given $\lambda \in (0, 1)$. These results follow from the new connections between brackets, containers and Mnets we have developed in this paper.

Our general bounds are in terms of shallow cell complexity and Property \mathbb{M} and so can seem somewhat abstract. Therefore we are deferring the conceptual connections in their full generality to the full version. For a set system (X, \mathcal{R}) , its *projection* on to a subset $Y \subset X$ of the ground set is the system $(Y, \mathcal{R}_{|Y})$, where $\mathcal{R}_{|Y} := \{R \cap Y \mid R \in \mathcal{R}\}$. The *VC dimension* of (X, \mathcal{R}) is the size of the largest subset $Y \subset X$, such that $\mathcal{R}_{|Y} \equiv 2^Y$, i.e. the entire power set of Y is expressible as a collection of intersections with members of the family \mathcal{R} . In this section, we will present a more simplified version of the structural results, in terms of the VC dimension, and give applications of these results to online and distributed learning.

2.1 Bounds for Semi-algebraic Set Systems

Let $X \subseteq \mathbb{R}^d$ be a domain with $|X| = n$. A subset $Y \subseteq X$ is called *semi-algebraic subset* if it is the intersection of X with a *semi-algebraic region*. Recall that a semi-algebraic region is a subset $R \subseteq \mathbb{R}^d$ which is in the algebra generated by polynomial equalities and inequalities. Notice that every semi-algebraic region has natural measures of complexity according to the degrees of the polynomials and the number of boolean operations which are used to define the region. In the statements of the results presented below we will assume that these complexities are constant in the sense that we will not specify how the stated bounds depend on these measures. Note that in the full version we provide a more detailed account [20]. Semi-algebraic sets of constant complexity include objects like halfspaces, rectangles, triangles, k -polytopes (where k is constant), etc. A set system (X, \mathcal{R}) , with $X \subseteq \mathbb{R}^d$, is said to be *induced* by semi-algebraic regions in \mathbb{R}^d if all the subsets of X in \mathcal{R} are semi-algebraic subsets of X .

We say a set system (X, \mathcal{R}) has *shallow-cell complexity* $\psi(\cdot, \cdot)$ if for any finite subset Y of X , the number of subsets of Y of size at most ℓ in $\mathcal{R}_{|Y}$ is at most $|Y| \cdot \psi(|Y|, \ell)$.

For semi-algebraic set systems of bounded shallow cell complexity, our bounds for Mnets, containers and brackets can be stated more explicitly, as given below.

Mnets of arbitrary heaviness

The Mnet construction of Mustafa and Ray [28] as well as those obtained in [19] are λ -heavy where $\lambda \leq 1/2$. In fact in the case of the Mnets obtained in [19] λ is given by the multilevel polynomial partitioning theorem, and depends inversely polynomially on the ambient dimension, the maximum degree of the polynomial family, and the number of allowed Boolean operations. A natural question that arises is, can the heaviness of the constructed Mnets be improved beyond $1/2$ or even be made arbitrarily close to 1? A priori, this does not seem possible using the previous techniques, as these rely on an application of the pigeonhole principle to choose a region from an integral number of regions, all of which are enclosed by a range. Thus, in the best case, there are 2 regions inside a range and one is chosen, which gives $\lambda = 1/2$.

Our first result is that for semi-algebraic systems, Mnets of arbitrarily small heaviness can be boosted to get Mnets of any desired heaviness λ . This extends and generalizes the results of [28] and [19], whose techniques, as we observed earlier, cannot give Mnets of heaviness more than $1/2$.

► **Theorem 1** (Informal statement: Mnet for semi-algebraic set system). *Let $X \subset \mathbb{R}^d$, and (X, \mathcal{R}) be a set system induced by semi-algebraic regions in \mathbb{R}^d of constant complexity with VC dimension d_0 . Then there exists λ -heavy η -Mnets \mathcal{M} of (X, \mathcal{R}) of size at most*

$$|\mathcal{M}| \leq \left(\frac{2}{1-\lambda}\right)^{c_1 d_0} \times \left(\frac{c_2}{\eta}\right)^{2d_0},$$

where c_1 depends only on d and the complexity of the semi-algebraic regions, and c_2 is an absolute constant. (For a more precise result in terms of the shallow cell complexity please refer to the full version of the paper [20])

Containers

Generalizing the results of Braverman et al. [14] showing the existence of containers for points and halfspaces, we show that containers can be obtained for semi-algebraic set systems.

► **Theorem 2** (Informal statement: Containers for semi-algebraic set systems). *Let $X \subset \mathbb{R}^d$, and (X, \mathcal{R}) be a set system induced by semi-algebraic regions in \mathbb{R}^d of constant complexity with VC dimension d_0 . Then there exists an ε -container \mathcal{C} for (X, \mathcal{R}) of size at most*

$$|\mathcal{C}| \leq \left(\frac{2}{\varepsilon}\right)^{cd_0},$$

where c depends only on d and the complexity of the semi-algebraic regions. (For a more precise bound in terms of the shallow cell complexity see the full version of the paper [20])

While the bounds on containers for points and halfspaces in [14] can be shown to hold for semi-algebraic systems using operations like Veronese mappings and lifts, such operations can blow up the ambient dimensionality – which appears in the exponent in the bounds – by a polynomial factor. The general version of Theorem 2 (see the full version [20]) gives direct bounds on the size of the container family in terms of shallow cell complexity, which in

some case has a lower dimensionality, and therefore better bounds, than those of [14]. This is usually the case for things like ε -nets as shallow cell complexity captures the combinatorial complexity of set systems at a much finer scale than, say VC dimension [4, 36, 17, 29, 30].

More specifically, for set system of points and halfspaces in \mathbb{R}^d we obtain the following improved bound for containers for points and halfspaces.

► **Theorem 3** (Improved container bounds for points and halfspaces). *Let $X \subset \mathbb{R}^d$, and $\varepsilon \in (0, 1)$. Then there exists a collection of subset \mathcal{C} of X of size at most $(\frac{2}{\varepsilon})^{O(d)}$ such that for all halfspaces h of \mathbb{R}^d there exists $C_h \in \mathcal{C}$ such that*

$$X \cap h \subseteq C_h \text{ and } |C_h \setminus (X \cap h)| \leq \varepsilon |X|.$$

The above theorem removes the multiplicative factor of $d^{O(d)}$ which appears in the bounds of Braverman et al. [14], thus significantly improving the dependence on the ambient dimension, from superexponential to exponential. It is easier to see the improvement in Theorem 3 if we fix some $\varepsilon \in (0, 1)$ and make d tend to infinity. This dynamic plays a crucial role in getting the optimal communication complexity of distributed learning of halfspace problem, see Theorem 9.

Uniform brackets

Finally, using Theorem 1, by setting $\lambda = 1 - \varepsilon$ and $\eta = \varepsilon$, and Theorem 2 we get the following bounds on the size of ε -brackets.

► **Corollary 4** (Informal statement: Brackets for semi-algebraic set systems). *Let $X \subset \mathbb{R}^d$, and (X, \mathcal{R}) be a set system induced by semi-algebraic regions in \mathbb{R}^d of constant complexity with VC dimension d_0 . Then there exists an ε -bracket \mathcal{B} for (X, \mathcal{R}) of size at most*

$$|\mathcal{C}| \leq \left(\frac{2}{\varepsilon}\right)^{cd_0},$$

where c depends only on d and the complexity of the semi-algebraic regions. (For a more general and precise bound in terms of the shallow cell complexity see the full version [20])

It is a simple exercise to see that any $\varepsilon/2$ -container for points and halfspaces in \mathbb{R}^d is also an ε -bracket for the same set of points and halfspaces in \mathbb{R}^d . Therefore, we get the following result directly from Theorem 3.

► **Corollary 5** (Improved bracketing bounds for points and halfspaces). *Let $X \subset \mathbb{R}^d$, and $\varepsilon \in (0, 1)$. Then there exists a collection of subset \mathcal{B} of X of size at most $(\frac{2}{\varepsilon})^{O(d)}$ such that for all halfspaces h of \mathbb{R}^d , there exist sets B_h^- and B_h^+ in \mathcal{B} such that*

$$B_h^- \subseteq X \cap h \subseteq B_h^+ \text{ and } |B_h^+ \setminus B_h^-| \leq \varepsilon |X|.$$

The above theorem extends to a *distribution-free*¹ bound with respect to arbitrary probability measures. See, Braverman et al. [14] and Haghtalab et al. [22].

► **Corollary 6** (Improved bracketing number for halfspaces). *Let \mathcal{H} be a family of halfspaces in \mathbb{R}^d . For all $\varepsilon \in (0, 1)$, and probability measure μ over \mathbb{R}^d we have*

$$N_{[\cdot]}(\mathcal{H}, \mu, \varepsilon) \leq (2/\varepsilon)^{O(d)}.$$

¹ This bound on the ε -bracketing number for halfspaces is distribution-free in the sense that it does not depend on the probability measure μ .

Braverman et al. [14] and Haghtalab et al. [22] showed that distribution-free ε -bracketing number for halfspaces in \mathbb{R}^d is $(\frac{d}{\varepsilon})^{O(d)}$. Note that our result is an improvement over this bound by a factor of $d^{O(d)}$. More detailed calculations reveal the constant in the $O(d)$ -exponent to be less than 7.03 in our case.

Further, we give a lower bound showing that the upper bounds established above are best possible up to dimension-independent constants in the exponent

► **Theorem 7** (Lower bounds for ε -containers). *There exists $C_d > 0$ that depends only on d such that the following holds:*

- Given positive integers $d \geq 2$, n , and $\varepsilon \in (0, 1)$, there exists a set X of n points in \mathbb{R}^d such that any ε -container for the set system induced by the set X and halfspaces has size at least

$$C_d \cdot \frac{1}{\varepsilon^{\lfloor (d+1)/3 \rfloor}}.$$

- For all integers $d \geq 2$, $n \geq 0$, and $\varepsilon \in (0, 1)$, there exists a set Y of n points in \mathbb{R}^d such that any ε -container for the set system induced by the set Y and hyperplanes has size at least

$$C_d \cdot \frac{1}{\varepsilon^d}.$$

► **Remark 8.** Note that the set systems induced by halfspaces and hyperplanes in \mathbb{R}^d have VC dimension $d + 1$ and d respectively.

2.2 Applications

Our improved bounds have applications in several areas such as the smoothed analysis of online learning algorithms as well as in distributed learning algorithms, e.g. the disjointness of convex bodies and LP feasibility. Some of these applications are described below.

Distributed learning of halfspaces

Linear classifiers are objects of central importance in many machine learning algorithms, beginning from the original *perceptron* model of Rosenblatt [31] to modern algorithms like neural networks, kernel machines, etc. A basic problem in machine learning therefore, relates to the learning of linear classifiers, namely halfspaces. The distributed learning of halfspaces problem has received considerable attention [18, 21, 27, 9, 23, 24, 14]. Balcan et al. [9] and Daumé III et al. [23] proved an upper bound of $O(d \log^2 n)$ bits on the communication complexity of learning halfspaces over a domain of n points in \mathbb{R}^d , and Kane et al. [24] proved that any randomized protocol for the above problem will require $\Omega(d + \log n)$ bits of communication. Braverman et al. [14] gave an improved a deterministic protocol with communication complexity $O(d \log d \log n)$, and proved an almost matching lower bound of $\Omega(d \log(n/d))$.

Let U be a known set of n points in \mathbb{R}^d . In *distributed learning of halfspaces* problem, two players, Alice and Bob are given sets S_a and S_b where $S_a, S_b \subseteq U \times \{\pm 1\}$ respectively such that the sets $\{(x, +1) \in S_a \cup S_b : x \in U\}$ and $\{(x, -1) \in S_a \cup S_b : x \in U\}$ can be separated by a hyperplane in \mathbb{R}^d . The goal is for both the players, using to agree classifier $H : U \rightarrow \{\pm 1\}$, such that

- if $(x, +1) \in S_a \cup S_b$ then $H(x) = +1$, and
- if $(x, -1) \in S_a \cup S_b$ then $H(x) = -1$.

Using the communication protocol of Braverman et al. [14] for the problem together with Theorem 3, we get the following upper bound which tightly meets Braverman et al.'s lower bound of $\Omega(d \log(n/d))$ whenever $n \geq d^{1+\Omega(1)}$.

► **Theorem 9.** *Let U be a known n -sized subset of \mathbb{R}^d . Then, there exists a deterministic protocol for Learning Halfspaces over U with communication complexity $O(d \log n)$ bits.*

We note that in this context previous works typically assume that the number of domain points n is much larger than the euclidean-dimension, and often even that $n = \exp(d)$. (Consider e.g. the natural case when the domain $U = \{0, 1\}^d$ consists of all binary vectors in \mathbb{R}^d .) In such cases, the above upper bound completely resolves the communication complexity of distributed learning of halfspaces.

Distributed convex set disjointness problem and LP feasibility

Kane et al. [24] introduced the *distributed convex set disjointness* problem in communication complexity, where, like in the case of distributed learning of halfspaces, there is a known n -sized domain $U \subset \mathbb{R}^d$ and two parties Alice and Bob are given as inputs S_a and S_b , with $S_a, S_b \subset U$, respectively. The goal is for both parties to decide if the convex hulls² $\text{conv}(S_a)$ and $\text{conv}(S_b)$ intersect or not. Note that the distributed convex set disjointness problem is equivalent to the fundamental problem of two-party *distributed Linear Programming (LP) feasibility*. For a more detailed discussion on this equivalence, see [14].

Vempala et al. [37] gave the first $O(d^3 \log^2 n)$ upper bound for the distributed convex set disjointness problem, and $\Omega(d \log n)$ and $\Omega(\log n)$ lower bounds for the deterministic and randomized settings respectively. Braverman et al. [14] gave an improved $O(d^2 \log d \log n)$ upper bound for the distributed convex set disjointness problem, and they also proved a randomized $\Omega(d \log(n/d))$ bits lower bound.

Using the Braverman et al. [14] communication protocol for the distributed convex set disjointness problem together with Corollary 5 we get the following result.

► **Theorem 10.** *Let U be an n -sized subset of \mathbb{R}^d . Then there exists a deterministic communication protocol for Convex Set Disjointness problem over U with communication complexity $O(d^2 \log n)$ bits.*

Theorem 10 gives a $\log d$ factor improvement over the bound of Braverman et al. [14].

Improved bracketing number and online algorithms

The *bracketing number* of a set system is a fundamental tool in statistics for proving uniform laws of large numbers for empirical processes [1]. More recently, Haghtalab, Roughgarden and Shetty [22] used bracketing numbers for smoothed analysis of online and differentially private learning algorithms.

Haghtalab, Roughgarden and Shetty [22], using the Braverman et al. [14] ε -container bound for points and halfspaces, showed that

$$N_{[\cdot]}(\mathcal{P}^{n,d}, \mu, \varepsilon) \leq \exp(c_1 n^d \ln(n^d/\varepsilon)), \text{ and } N_{[\cdot]}(\mathcal{Q}^{d,k}, \mu, \varepsilon) \leq \exp(c_2 nk \ln(nk/\varepsilon)). \quad (1)$$

where $\mathcal{P}^{n,d}$ denotes the class of d -degree polynomial threshold functions in \mathbb{R}^n and $\mathcal{Q}^{n,k}$ be the class of k -polytopes in \mathbb{R}^n , and c_1 and c_2 are absolute constants. Using Corollary 6, together with [22, Theorem 3.7], we can directly improve the distribution-free bounds for ε -bracketing numbers:

² For any subset S of \mathbb{R}^d , convex hull of S will be denoted by $\text{conv}(S)$.

► **Theorem 11.** *Let $(\mathbb{R}^n, \Omega, \mu)$ be a probability space. Then*

1. $N_{[]}(\mathcal{P}^{n,d}, \mu, \epsilon) \leq \exp(c_1 n^d \ln(1/\epsilon))$, where c_1 is an absolute constant.
2. $N_{[]}(\mathcal{Q}^{d,k}, \mu, \epsilon) \leq \exp(c_2 nk \ln(1/\epsilon))$, where c_2 is an absolute constant.

The notion of *regret* is a standard measure of the effectiveness of *online learning* algorithms. Online learnability is characterized by having finite Littlestone dimension [25, 13]. However, this can be a very restrictive condition, as there are instances of problems which have constant VC dimension, yet their Littlestone dimension is infinite [2, 3, 13, 16]. Recently, going beyond worst-case analysis, Haghtalab et al. [22] have introduced the *smoothed analysis* paradigm of Spielman-Teng [33] to the context of online learning algorithms. Using this paradigm, they designed online learning no-regret algorithms for several problems with infinite Littlestone dimension, even for the case of *adaptive adversaries*, provided the adversaries choose from a σ -smooth distribution. For an introduction to online regret minimization against an σ -smoothed adversary see [22].

Using Theorem 11, together with [22, Theorem 3.3] we will get the following improved online algorithm whose regret against an adaptive σ -smoothed adversary on $\mathcal{P}^{n,d}$ and $\mathcal{Q}^{n,k}$ satisfies:

► **Theorem 12.** *There exists an online algorithm against an adaptive σ -smoothed adversary whose regret after T -steps is*

1. $O\left(\sqrt{T \cdot \text{VCdim}(\mathcal{P}^{n,d}) \log \frac{T}{\sigma}}\right)$ if the class of functions is $\mathcal{P}^{n,d}$.
2. $O\left(\sqrt{T \cdot \text{VCdim}(\mathcal{Q}^{n,k}) \log \frac{T}{\sigma}}\right)$ if the class of functions is $\mathcal{Q}^{n,d}$.

► **Remark 13.** Theorem 12 is an improvement over [22, Corollary 3.8] where the regret bounds were

$$O\left(\sqrt{T \cdot \text{VCdim}(\mathcal{P}^{n,d}) \left(\log \frac{T}{\sigma} + \log \text{VCdim}(\mathcal{P}^{n,d})\right)}\right)$$

and

$$O\left(\sqrt{T \cdot \text{VCdim}(\mathcal{P}^{n,d}) \left(\log \frac{T}{\sigma} + \log \text{VCdim}(\mathcal{P}^{n,d})\right)}\right)$$

for the class of functions $\mathcal{P}^{n,d}$ and $\mathcal{Q}^{n,d}$ respectively.

References

- 1 Terrence M. Adams and Andrew B. Nobel. Uniform Convergence of Vapnik–Chervonenkis Classes under Ergodic Sampling. *Ann. Probab.*, 38(4):1345–1367, 2010.
- 2 Noga Alon, David Haussler, and Emo Welzl. Partitioning and Geometric Embedding of Range Spaces of Finite Vapnik-Chervonenkis Dimension. In *Proceedings of the 3rd Annual Symposium on Computational Geometry, SoCG*, pages 331–340, 1987.
- 3 Noga Alon, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private PAC learning implies finite Littlestone dimension. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 852–860, 2019.
- 4 Boris Aronov, Esther Ezra, and Micha Sharir. Small-size epsilon-nets for axis-parallel rectangles and boxes. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC*, pages 639–648, 2009.
- 5 Rahul Arya, Sunil Arya, Guilherme Dias da Fonseca, and David M. Mount. Optimal Bound on the Combinatorial Complexity of Approximating Polytopes. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 786–805, 2020.

- 6 Sunil Arya, Guilherme Dias da Fonseca, and David M. Mount. On the Combinatorial Complexity of Approximating Polytopes. *Discrete & Computational Geometry*, 58(4):849–870, 2017.
- 7 Sunil Arya, Guilherme Dias da Fonseca, and David M. Mount. Optimal Approximate Polytope Membership. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 270–288, 2017.
- 8 Sunil Arya, David M. Mount, and Jian Xia. Tight Lower Bounds for Halfspace Range Searching. *Discrete & Computational Geometry*, 47(4):711–730, 2012.
- 9 Maria-Florina Balcan, Avrim Blum, Shai Fine, and Yishay Mansour. Distributed Learning, Communication Complexity and Privacy. In *Proceedings of 25th Annual Conference on Learning Theory, COLT*, volume 23, pages 26.1–26.22, 2012.
- 10 József Balogh, Robert Morris, and Wojciech Samotij. Independent Sets in Hypergraphs. *Journal of the American Mathematical Society*, 28(3):669–709, 2015.
- 11 Imre Bárány. The Technique of M-Regions and Cap-Coverings: A Survey. *Rend. Circ. Mat. Palermo*, 65:21–38, 2000.
- 12 Imre Bárány. Extremal Problems for Convex Lattice Polytopes: A Survey. *Contemp. Math.*, 453:87–103, 2008.
- 13 Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic Online Learning. In *Proceedings of the 22nd Annual Conference on Learning Theory, COLT*, 2009.
- 14 Mark Braverman, Gillat Kol, Shay Moran, and Raghuvansh R. Saxena. Near Optimal Distributed Learning of Halfspaces with Two Parties. In *Proceedings of the 34th Annual Conference on Learning Theory, COLT*, volume 134, pages 724–758, 2021.
- 15 Hervé Brönnimann, Bernard Chazelle, and János Pach. How hard is half-space range searching. *Discrete & Computational Geometry*, 10:143–155, 1993.
- 16 Mark Bun, Roi Livni, and Shay Moran. An Equivalence Between Private Classification and Online Prediction. In *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 389–402, 2020.
- 17 Timothy M. Chan, Elyot Grant, Jochen Könemann, and Malcolm Sharpe. Weighted Capacitated, Priority, and Geometric Set Cover via Improved Quasi-Uniform Sampling. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1576–1585, 2012.
- 18 Edward Y. Chang, Kaihua Zhu, Hao Wang, Hongjie Bai, Jian Li, Zhihuan Qiu, and Hang Cui. Parallelizing Support Vector Machines on Distributed Computers. In *Proceedings of the 21st Annual Conference on Neural Information Processing Systems, NIPS*, pages 257–264, 2007.
- 19 Kunal Dutta, Arijit Ghosh, Bruno Jartoux, and Nabil H. Mustafa. Shallow Packings, Semialgebraic Set Systems, Macbeath Regions, and Polynomial Partitioning. *Discrete & Computational Geometry*, 61(4):756–777, 2019.
- 20 Kunal Dutta, Arijit Ghosh, and Shay Moran. Uniform Brackets, Containers, and Combinatorial Macbeath Regions. *CoRR*, abs/2111.10048, 2021. [arXiv:2111.10048](https://arxiv.org/abs/2111.10048).
- 21 Pedro A. Forero, Alfonso Cano, and Georgios B. Giannakis. Consensus-Based Distributed Support Vector Machines. *Journal of Machine Learning Research*, 11:1663–1707, 2010.
- 22 Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis of online and differentially private learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- 23 Hal Daumé III, Jeff M. Phillips, Avishek Saha, and Suresh Venkatasubramanian. Efficient Protocols for Distributed Classification and Optimization. In *Proceedings of 23rd International Conference on Algorithmic Learning Theory, ALT*, volume 7568, pages 154–168, 2012.
- 24 Daniel Kane, Roi Livni, Shay Moran, and Amir Yehudayoff. On Communication Complexity of Classification Problems. In *Proceedings of the 32nd Annual Conference on Learning Theory, COLT*, volume 99, pages 1903–1943, 2019.

- 25 Nick Littlestone. Learning Quickly when Irrelevant Attributes Abound: A New Linear-threshold Algorithm. *Machine Learning*, 2:285–318, 1988.
- 26 A. M. Macbeath. A Theorem on Non-Homogeneous Lattices. *Annals of Mathematics*, 56:269–293, 1952.
- 27 Ryan T. McDonald, Keith B. Hall, and Gideon Mann. Distributed Training Strategies for the Structured Perceptron. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, pages 456–464, 2010.
- 28 N. H. Mustafa and S. Ray. ϵ -Mnets: Hitting Geometric Set Systems with Subsets. *Discrete & Computational Geometry*, 57(3):625–640, 2017.
- 29 Nabil H. Mustafa, Kunal Dutta, and Arijit Ghosh. A Simple Proof of Optimal Epsilon-nets. *Combinatorica*, 38(5):1269–1277, 2018.
- 30 Nabil H. Mustafa and Kasturi R. Varadarajan. Epsilon-approximations and epsilon-nets. *CoRR*, abs/1702.03676, 2017. [arXiv:1702.03676](https://arxiv.org/abs/1702.03676).
- 31 F. Roosenblatt. A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, pages 65–386, 1958.
- 32 David Saxton and Andrew Thomason. Hypergraph Containers. *Inventiones Mathematicae*, 201(3):925–992, 2015.
- 33 Daniel A. Spielman and Shang-Hua Teng. Smoothed Analysis of Algorithms: Why the Simplex Algorithm usually takes Polynomial Time. *Journal of the ACM*, 51(3):385–463, 2004.
- 34 Ramon van Handel. The Universal Glivenko–Cantelli Property. *Probability Theory and Related Fields*, 155:911–934, 2013.
- 35 V. N. Vapnik and A. Y. Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.
- 36 Kasturi R. Varadarajan. Weighted Geometric Set Cover via Quasi-Uniform Sampling. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC*, pages 641–648, 2010.
- 37 Santosh S. Vempala, Ruosong Wang, and David P. Woodruff. The Communication Complexity of Optimization. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1733–1752, 2020.