

Simplicial Covering Dimension of Extremal Concept Classes

Ari Blondal 

McGill University, Montreal, Canada

Hamed Hatami 

McGill University, Montreal, Canada

Pooya Hatami 

Ohio State University, Columbus, OH, USA

Chavdar Lalov 

Ohio State University, Columbus, OH, USA

Sivan Tretiak 

Ohio State University, Columbus, OH, USA

Abstract

Dimension theory is a branch of topology concerned with defining and analyzing dimensions of geometric and topological spaces in purely topological terms. In this work, we adapt the classical notion of topological dimension (Lebesgue covering) to binary concept classes. The topological space naturally associated with a concept class is its space of realizable distributions. The loss function and the class itself induce a simplicial structure on this space, with respect to which we define a simplicial covering dimension.

We prove that for finite concept classes, this simplicial covering dimension exactly characterizes the list replicability number (equivalently, global stability) in PAC learning. This connection allows us to apply tools from classical dimension theory to compute the exact list replicability number of the broad family of extremal concept classes.

2012 ACM Subject Classification Computing methodologies → Supervised learning by classification; Theory of computation → Randomness, geometry and discrete structures; Mathematics of computing → Topology

Keywords and phrases PAC Learning, Extremal Concept Classes, Replicability, List Replicability, Topology, Geometry

Digital Object Identifier 10.4230/LIPIcs.ITCS.2026.22

Related Version *Full Version:* <https://arxiv.org/abs/2511.11819> [5]

Funding *Hamed Hatami:* Hamed Hatami is supported by an NSERC grant.

1 Introduction

In recent years, several intriguing advances in learning theory have been made using tools from topology. These developments point to a potential connection between learning theory and *topological dimension theory*, a classical branch of topology pioneered by Brouwer, Lebesgue, and others in the early twentieth century, who aimed to formalize a general notion of dimension that demonstrates $\mathbb{R}^m$ and $\mathbb{R}^n$ are homeomorphic only when $m = n$.

Motivated by this connection, we introduce a new notion of dimension for concept classes, inspired by the classical Lebesgue covering dimension and the simplicial structures that naturally arise in learning problems. We focus on binary classification, the task of learning an unknown function that maps the elements of a domain $\mathcal{X}$ to one of two possible labels, typically denoted by ± 1 . Formally, a binary classification problem is specified as a *binary concept class* $\mathcal{C}$, a set of functions from a domain $\mathcal{X}$ to the label set $\{\pm 1\}$. Throughout this work, we restrict attention to finite domains $\mathcal{X}$.



© Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak; licensed under Creative Commons License CC-BY 4.0

17th Innovations in Theoretical Computer Science Conference (ITCS 2026).

Editor: Shubhangi Saraf; Article No. 22; pp. 22:1–22:24



Leibniz International Proceedings in Informatics

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

We define the *simplicial covering dimension* $\text{SC}(\mathcal{C})$ of a concept class by studying the topology of its space of realizable distributions, endowed with the simplicial structure induced by $\mathcal{C}$ and the loss function. Our definition is partly inspired by the line of work [11, 30, 14, 10, 3, 12], which showed that *list-replicable learning* is intrinsically connected to the geometry and topology of the space of realizable distributions.

In the finite-domain setting, the simplicial covering dimension turns out (Theorem A) to coincide with the *list-replicability number* of the class:

$$\text{SC}(\mathcal{C}) = \text{LR}(\mathcal{C}) - 1, \quad (1)$$

where $\text{LR}(\mathcal{C})$ denotes the list-replicability number of $\mathcal{C}$. This well-studied parameter arose in recent years as part of a broader attempt to formalize the notion of replicability [7, 25, 11, 6, 23, 17, 18, 27, 15, 21, 22], which refers to the requirement that an algorithm produce consistent outcomes when repeated under similar conditions and with similar data. Our framework allows us to import classical tools from dimension theory into the analysis of replicable learning.

In particular, by analyzing SC , we determine the list-replicability number of *extremal concept classes* (also known as ample concept classes), a family that includes many of the most natural examples of concept classes. Our main theorem (Theorem B) establishes that every *extremal* concept class $\mathcal{E}$ over a finite domain $\mathcal{X}$ satisfies

$$\text{SC}(\mathcal{E}) = \begin{cases} \text{VC}(\mathcal{E}) - 1 & \text{if } \mathcal{E} = \{\pm 1\}^{\mathcal{X}} \\ \text{VC}(\mathcal{E}) & \text{otherwise} \end{cases}, \quad (2)$$

where $\text{VC}(\mathcal{E})$ denotes the VC dimension of $\mathcal{E}$.

This theorem resolves the list-replicability number of several well-studied classes, such as axis-parallel boxes and downward-closed classes [8, Section 3.2], whose values were previously unknown. Beyond the new cases, it also unifies the characterization of the list replicability number for several previously known cases, such as the binary cube [11], threshold functions [11], and more generally halfspaces [4].

1.1 Preliminaries and definitions

In this section, we provide precise definitions of the notions discussed in the introduction.

1.1.1 PAC learning, VC dimension, and list replicability

In probably approximately correct (PAC) learning, the learner is given parameters $\delta, \epsilon > 0$ and receives training data S consisting of $n = n(\mathcal{C}, \delta, \epsilon)$ independent labeled examples drawn from an unknown but fixed distribution μ over $\mathcal{X} \times \{\pm 1\}$. We work in the *realizable* setting: there exists some concept $c \in \mathcal{C}$ that correctly labels all examples in the support of μ . The learner's task is to use the training data S to output, with probability at least $1 - \delta$, a *hypothesis* $h : \mathcal{X} \rightarrow \{\pm 1\}$ whose *population loss*

$$\text{loss}_\mu(h) := \Pr_{(x,b) \sim \mu} [h(x) \neq b]$$

is at most ϵ .

Throughout this work, a *learning rule* refers to a (randomized) function $\mathcal{A}$ that maps any sample $S \in \bigcup_{n=0}^{\infty} (\mathcal{X} \times \{\pm 1\})^n$ to a hypothesis $\mathcal{A}(S) \in \{\pm 1\}^{\mathcal{X}}$. Since our primary focus is sample complexity rather than computational efficiency, we impose no computability constraints on $\mathcal{A}$.

VC dimension. A concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ shatters a set $S \subseteq \mathcal{X}$ if $\{c|_S : c \in \mathcal{C}\} = \{\pm 1\}^S$, where $c|_S$ denotes the restriction of c to S . The *Vapnik-Chervonenkis* (VC) dimension of $\mathcal{C}$ is defined as

$$\text{vc}(\mathcal{C}) := \sup\{|S| : S \subseteq \mathcal{X} \text{ is shattered by } \mathcal{C}\}.$$

The fundamental theorem of PAC learning states that the sample complexity of PAC learning is determined by the VC dimension. More precisely, the optimal sample size for PAC learning a binary class $\mathcal{C}$ with accuracy parameter ϵ and confidence parameter δ is $\Theta_{\epsilon, \delta}(\text{vc}(\mathcal{C}))$.

List replicability. List replicability, introduced in [11, 14], is an elegant reformulation of global stability [11, Definition 1].

► **Definition 1** (List replicability). *The list replicability number of a concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$, denoted $\text{LR}(\mathcal{C})$, is the smallest integer L such that the following holds. For every $\epsilon, \delta > 0$, there exists a sample size $n(\mathcal{C}, \epsilon, \delta)$ and a learning rule $\mathcal{A}$ such that for every realizable distribution μ on $\mathcal{X} \times \{\pm 1\}$, there exists a list of hypotheses $h_1, \dots, h_L \in \{\pm 1\}^{\mathcal{X}}$ satisfying*

- $\text{loss}_{\mu}(h_i) \leq \epsilon$ for all $i = 1, \dots, L$;
- $\Pr_{S \sim \mu^n} [\mathcal{A}(S) \notin \{h_1, \dots, h_L\}] \leq \delta$ where $n = n(\mathcal{C}, \epsilon, \delta)$.

1.1.2 Extremal classes

A class $\mathcal{C}$ over a domain $\mathcal{X}$ *strongly shatters* a set $S \subseteq \mathcal{X}$ if there exists a labeling $a \in \{\pm 1\}^{\mathcal{X} \setminus S}$ such that

$$\{c|_S : c \in \mathcal{C}, c|_{\mathcal{X} \setminus S} = a\} = \{\pm 1\}^S.$$

In other words, $\mathcal{C}$ realizes all $2^{|S|}$ labelings on S while fixing the labels outside S in some prescribed way. A class $\mathcal{C}$ is called *extremal* if every set shattered by $\mathcal{C}$ is also strongly shattered.

Many natural concept classes are either extremal or admit natural extremal extensions. We refer the reader to [8, Section 3.2] for a more comprehensive list of known examples; here, we present a few illustrative ones.

► **Example 2** (Sign patterns of convex sets [24]). Let $K \subset \mathbb{R}^n$ be a convex set. The class

$$\mathcal{C}(K) := \{\text{sign}(v) : v \in K, v_i \neq 0 \forall i \leq n\},$$

where $\text{sign}(v) \in \{\pm 1\}^n$ denotes the sign pattern of v , is extremal.

► **Example 3** (Homogeneous half-spaces). Let $P \subset \mathbb{R}^d$ be a finite set of points. Define the concept class of homogeneous half-spaces $\mathcal{H}_P$ over the domain P as follows: each homogeneous half-space in $\mathbb{R}^d$ whose defining hyperplane avoids P induces a labeling of P by $\{\pm 1\}$. Explicitly, the label of $p \in P$ is given by

$$p \mapsto \text{sign}(\langle v, p \rangle),$$

where the inequality $\langle v, x \rangle > 0$ defines the half-space.

The set of maps $p \mapsto \langle v, p \rangle$ (for all $v \in \mathbb{R}^d$) is convex, and therefore, by Example 2, the class $\mathcal{H}_P$ is extremal.

► **Example 4** (Axis-parallel boxes). Let $P \subset \mathbb{R}^d$ be a finite set of points, no two of which share a coordinate. Each axis-parallel box in $\mathbb{R}^d$ induces a labeling of P by ± 1 , depending on whether a point lies inside or outside the box. The corresponding concept class is extremal.

► **Example 5** (Median classes). A class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ is called *median* if it is closed under taking the majority of three concepts. More precisely, for any $c_1, c_2, c_3 \in \mathcal{C}$, the concept

$$x \mapsto \text{maj}(c_1(x), c_2(x), c_3(x))$$

also belongs to $\mathcal{C}$. All median classes are extremal [2, Proposition 2].

While not every class is extremal, a well-known open problem asks whether every class can be extended to an extremal one without significantly increasing its VC dimension.

► **Question 6** ([28, 9]). *Does there exist a function $t : \mathbb{N} \rightarrow \mathbb{N}$ such that, for every binary class $\mathcal{C}$ over a finite domain $\mathcal{X}$, there is an extremal class $\mathcal{E} \supseteq \mathcal{C}$ over $\mathcal{X}$ with*

$$\text{vc}(\mathcal{E}) \leq t(\text{vc}(\mathcal{C}))?$$

1.1.3 Topological dimension theory

For two families $\mathcal{A} = \{A_i\}_{i \in I}$ and $\mathcal{B} = \{B_j\}_{j \in J}$ of subsets of a set X , indexed by the same set I , we say that $\mathcal{B}$ is a *shrinkage* of $\mathcal{A}$ if $B_j \subseteq A_i$ for all $i \in I$.

More generally, a family $\mathcal{B} = \{B_j\}_{j \in J}$ of subsets of X is said to be a *refinement* of another family $\mathcal{A} = \{A_i\}_{i \in I}$, written $\mathcal{A} \prec \mathcal{B}$, if for every B_j , there exists A_i such that $B_j \subseteq A_i$. In particular, every shrinkage of $\mathcal{A}$ is a refinement of $\mathcal{A}$.

The *order* $\text{ord}(\mathcal{A})$ of a family $\mathcal{A} = \{A_i\}_{i \in I}$ of sets is defined as one less than the maximum number of sets in $\mathcal{A}$ that have a nonempty intersection. Now suppose $\mathcal{A}$ is a *finite open cover* of a topological space X . The *Lebesgue covering number* of $\mathcal{A}$ is defined¹ as

$$L(\mathcal{A}) := \min\{\text{ord}(\mathcal{B}) : \mathcal{B} \text{ is an open cover of } X \text{ and a shrinkage of } \mathcal{A}\}.$$

► **Definition 7** (Topological dimension). *The Lebesgue covering dimension (or simply the topological dimension) of X is defined by*

$$\dim(X) := \sup\{L(\mathcal{A}) : \mathcal{A} \text{ is a finite open cover of } X\}. \quad (3)$$

► **Remark 8.** In this paper, we will always work with topological spaces of the form $X \subseteq \mathbb{R}^n$ endowed with the subspace topology inherited from $\mathbb{R}^n$. Thus, a set $S \subseteq X$ is open if and only if $S = U \cap X$ for some open set $U \subseteq \mathbb{R}^n$. For such spaces, it is often convenient to allow covers of X whose elements may extend outside X . In this case, we identify $\mathcal{A} = \{A_i\}_{i \in I}$ with $\{A_i \cap X\}_{i \in I}$, and call $\mathcal{A}$ an open cover of X if $\{A_i \cap X\}_{i \in I}$ is an open cover of X . We write $\mathcal{A} \prec \mathcal{B}$ to mean $\{A_i \cap X\}_{i \in I} \prec \{B_j \cap X\}_{j \in J}$. However, we still calculate $\text{ord}(\mathcal{A})$ as one less than the maximum number of sets containing any one point $x \in X$, and define $L(\mathcal{A})$ accordingly.

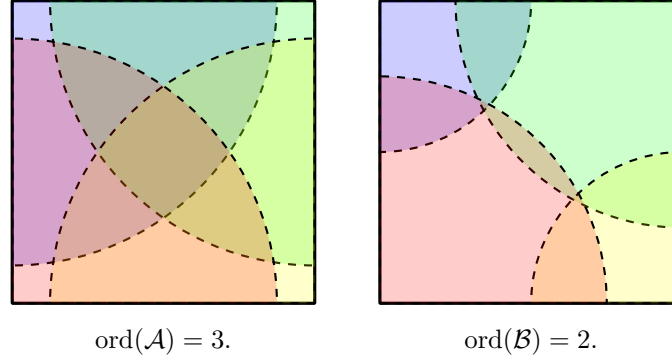
The notion of Lebesgue covering dimension of Definition 7 was formally introduced by Eduard Čech in 1933, building on a theorem of Lebesgue that serves as one of the main tools in this paper.²

► **Theorem 9** (Lebesgue covering theorem). *Suppose a d -dimensional cube $[-1, 1]^d \subset \mathbb{R}^d$ is covered by a finite family $\mathcal{A}$ of open sets, none of which contains points of opposite faces of the cube. Then $\text{ord}(\mathcal{A}) \geq d$.*

See Figure 1 for an illustration of the Lebesgue covering dimension of a square in $\mathbb{R}^2$.

¹ Traditionally, $L(\mathcal{A})$ is defined by taking the minimum of $\text{ord}(\mathcal{B})$ over all finite open covers $\mathcal{B}$ that refine $\mathcal{A}$. However, as shown in [13, Proposition 1.1.7], it suffices to consider only shrinkages of $\mathcal{A}$.

² The Lebesgue covering theorem is usually stated for closed sets, as in Theorem 36. The open-set formulation presented here is derived from the closed version in Section 3.



■ **Figure 1** The illustrated open cover $\mathcal{A}$ of the square $[-1, 1]^2$ has order 3, but it admits a shrinkage $\mathcal{B}$ of order 2, implying $L(\mathcal{A}) \leq 2$. Moreover, since no set in $\mathcal{A}$ contains points from opposite faces, the Lebesgue covering theorem shows $L(\mathcal{A}) \geq 2$.

We also rely on the following standard facts about topological dimension:

- **Theorem 10** ([29, Theorems 50.2 and 50.6]). *Topological dimension satisfies:*
- (i) *Every compact subspace of $\mathbb{R}^d$ has topological dimension at most d .*
 - (ii) *Let $X = Y \cup Z$, where Y and Z are closed subspaces of X having finite topological dimension. Then $\dim(X) = \max \{\dim(Y), \dim(Z)\}$.*

From Theorem 9, the cube $[-1, 1]^d$ has topological dimension at least d , while Theorem 10 (i) provides the upper bound of d . Hence, its topological dimension is exactly d . Consequently, any set homeomorphic to $[-1, 1]^d$, such as a d -simplex, or the unit ball in $\mathbb{R}^d$, also has topological dimension d .

For further reading on dimension theory, see the classical texts [20, 16].

1.1.4 Simplicial covering dimension of binary concept classes

In this section, we describe how the notion of topological dimension can be naturally adapted to the space of realizable distributions associated with a concept class. Our starting point is to represent every realizable distribution as a point in $\mathbb{R}^{\mathcal{X}}$. Every vector μ on the ℓ_1 -unit sphere in $\mathbb{R}^{\mathcal{X}}$ naturally encodes a distribution over $\mathcal{X} \times \{\pm 1\}$: for each $x \in \text{supp}(\mu)$, the distribution assigns mass $|\mu(x)|$ to the labeled example $(x, \text{sign}(\mu(x)))$. This correspondence identifies the realizable distributions of a concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ with the subset

$$\Delta_{\mathcal{C}} := \{\mu \in \mathbb{R}^{\mathcal{X}} : \|\mu\|_1 = 1 \text{ and } \exists c \in \mathcal{C} \text{ with } c(x) = \text{sign}(\mu(x)) \forall x \in \text{supp}(\mu)\} \subseteq \mathbb{R}^{\mathcal{X}}. \quad (4)$$

Note, however, that the topology of $\mathbb{R}^{\mathcal{X}}$ does not capture the connection between $\Delta_{\mathcal{C}}$ and the binary classification task, since the relevant notion of proximity in learning is not the Euclidean distance but rather the population loss, which only measures discrepancies on points where the given ± 1 labels differ. For instance, a given hypothesis h incurs zero population loss on all the realizable distributions in

$$B_h := \{\mu \in \mathbb{R}^{\mathcal{X}} : \|\mu\|_1 = 1 \text{ and } \text{loss}_{\mu}(h) = 0\}, \quad (5)$$

even though these distributions may be far apart geometrically in $\mathbb{R}^{\mathcal{X}}$. We refer to B_h as the *zero-loss set* around h .

Each B_h is a $(|\mathcal{X}| - 1)$ -simplex in $\mathbb{R}^{\mathcal{X}}$, and $\Delta_{\mathcal{C}} = \bigcup_{c \in \mathcal{C}} B_c$. Hence, unless $\mathcal{C}$ is empty, $\Delta_{\mathcal{C}}$, equipped with the Euclidean topology of $\mathbb{R}^{\mathcal{X}}$, has the same topological dimension as each simplex B_c , namely $|\mathcal{X}| - 1$. However, from the perspective of population loss, it is natural to regard all distributions within a zero-loss set B_c as being at “distance” zero from one another, since they all induce the same labeling of $\mathcal{X}$. Consequently, each B_c should be viewed as a 0-dimensional set in this context.

In other words, a suitable notion of topological dimension for $\Delta_{\mathcal{C}}$ should consider only those covers $\mathcal{A}$ in (3) that respect the equivalence relation induced by each zero-loss set B_h . We formalize this idea in the following definition.

► **Definition 11** (Simplicial covering dimension of a binary concept class). *Let $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ be a binary concept class over a finite domain $\mathcal{X}$, and let $\mathcal{B} = \{B_h\}_{h \in \{\pm 1\}^{\mathcal{X}}}$ be the cover of $\Delta_{\mathcal{C}} \subseteq \mathbb{R}^{\mathcal{X}}$ defined above. The simplicial covering dimension of $\mathcal{C}$ is*

$$sc(\mathcal{C}) := \sup \{L(\mathcal{A}) : \mathcal{A} \text{ is a finite open cover of } \Delta_{\mathcal{C}} \text{ satisfying } \mathcal{A} \prec \mathcal{B}\}.$$

The definition of the simplicial covering dimension for concept classes can be naturally interpreted through the lens of simplicial complexes.

Consider a finite geometric simplicial complex Γ in $\mathbb{R}^n$ (Definition 20). By definition, Γ is a collection of simplices whose union forms the polyhedron $\|\Gamma\| = \bigcup_{\sigma \in \Gamma} \sigma \subseteq \mathbb{R}^n$. It is natural to study the topological dimension of $\|\Gamma\|$ relative to this simplicial cover. More generally, for simplicial complexes $\Gamma \subseteq \Gamma'$, the faces $\sigma \in \Gamma'$ need not lie entirely in $\|\Gamma\|$, but by our convention Remark 8 they nevertheless define a cover of $\|\Gamma\|$.

► **Definition 12.** *Let $\Gamma \subseteq \Gamma'$ be geometric simplicial complexes in $\mathbb{R}^n$. The relative simplicial covering dimension of Γ with respect to Γ' is defined by $sc_{\Gamma'}(\Gamma) := \sup_{\mathcal{A}} L(\mathcal{A})$, where the supremum is taken over all finite open covers $\mathcal{A}$ of $\|\Gamma\|$ satisfying $\mathcal{A} \prec \Gamma'$.*

The quantity $sc_{\Gamma'}(\Gamma)$ only depends³ on the abstract simplicial structures (Definition 21) of Γ and Γ' , which allows us to define:

► **Definition 13.** *Let $K \subseteq K'$ be finite abstract simplicial complexes. The relative simplicial covering dimension of K with respect to the faces of K' is defined by*

$$sc_{K'}(K) := sc_{\Gamma'}(\Gamma),$$

where Γ' is any geometric realization of K' and $\Gamma \subseteq \Gamma'$ is its restriction to K .

Understanding the quantities $sc_K(K)$, and more generally $sc_{K'}(K)$, appears to be an intriguing problem in its own right, independent of the learning-theoretic motivations of this work.

► **Question 14.** *Is there a simple combinatorial characterization of $sc_K(K)$, and more generally $sc_{K'}(K)$, where $K' \supseteq K$ are finite simplicial complexes?*

Finally, the simplicial covering dimension of concept classes introduced in Definition 11 arises as a special case of Definition 13. Indeed, since $\Delta_{\mathcal{C}} = \bigcup_{c \in \mathcal{C}} B_c$, the set $\Delta_{\mathcal{C}}$ can be viewed as the polyhedron of the geometric simplicial complex Γ whose facets are the simplices B_c for $c \in \mathcal{C}$. Letting Γ' denote the simplicial complex whose facets are B_h for all $h \in \{\pm 1\}^{\mathcal{X}}$, we recover Definition 11 from Definition 13.

³ See Proposition 22.

1.2 Main contributions

Our first theorem establishes that, for binary concept classes over finite domains, the simplicial covering dimension exactly characterizes the list replicability number.

► **Theorem A.** *Let $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ be a concept class over a finite domain $\mathcal{X}$. Then*

$$LR(\mathcal{C}) = SC(\mathcal{C}) + 1.$$

Combining Theorem A with the known value of $LR(\{\pm 1\}^{\mathcal{X}})$ from [11] yields the following.

► **Theorem 15.** *For any finite domain $\mathcal{X}$, we have*

$$SC(\{\pm 1\}^{\mathcal{X}}) = |\mathcal{X}| - 1 \text{ and } LR(\{\pm 1\}^{\mathcal{X}}) = |\mathcal{X}|.$$

Consequently, every binary class $\mathcal{C}$ over a finite domain satisfies

$$SC(\mathcal{C}) \geq VC(\mathcal{C}) - 1 \text{ and } LR(\mathcal{C}) \geq VC(\mathcal{C}).$$

Proof. The equality $LR(\{\pm 1\}^{\mathcal{X}}) = |\mathcal{X}|$ is due to [11]. Combined with Theorem A, we obtain $SC(\mathcal{C}) = |\mathcal{X}| - 1$. Now, let $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ be a class over a finite domain $\mathcal{X}$, and let $S \subseteq \mathcal{X}$ be a finite set. Trivially, we have $LR(\mathcal{C}) \geq LR(\mathcal{C}|_S)$, and on the other hand, by considering any finite shattered set $S \subseteq \mathcal{X}$, we have

$$SC(\mathcal{C}) + 1 = LR(\mathcal{C}) \geq LR(\mathcal{C}|_S) = LR(\{\pm 1\}^S) = |S|. \quad \blacktriangleleft$$

► **Remark 16.** In [11], the lower bound $LR(\{\pm 1\}^{\mathcal{X}}) \geq |\mathcal{X}|$ is proved using the Poincaré-Miranda theorem and the upper bound $LR(\{\pm 1\}^{\mathcal{X}}) \leq |\mathcal{X}|$ is proved by designing an explicit algorithm. In our framework, the upper bound is immediate, since $SC(\mathcal{C})$ is by definition upper bounded by the topological dimension of $\Delta_{\mathcal{C}}$, and $\Delta_{\{\pm 1\}^{\mathcal{X}}}$ is the empty cross-polytope in $\mathbb{R}^{\mathcal{X}}$, which has topological dimension $|\mathcal{X}| - 1$. Moreover, as shown in Lemma 40, the lower bound can alternatively be recovered from the Lebesgue covering theorem.

The difficulty of computing the list replicability number of a concept class is in part due to the markedly different techniques required for upper and lower bounds. Upper bounds typically use explicit learning algorithms engineered to output from a small list, whereas lower bounds rely on ad hoc topological arguments.

In light of Theorem A, a natural alternative is to work directly with the simplicial covering dimension of the class. Using this approach, we determine the exact list replicability number for all extremal classes.

► **Theorem B (Main theorem).** *If $\mathcal{E} \neq \{\pm 1\}^{\mathcal{X}}$ is an extremal class over a finite domain $\mathcal{X}$, then*

$$SC(\mathcal{E}) = VC(\mathcal{E}) \text{ and } LR(\mathcal{E}) = VC(\mathcal{E}) + 1.$$

Note that the lower bound $SC(\mathcal{E}) \geq VC(\mathcal{E})$ in Theorem B improves by one upon the general lower bound $SC(\mathcal{C}) \geq VC(\mathcal{C}) - 1$ from Theorem 15 that holds for all finite-domain concept classes.

Proof overview of Theorem B. The proof analyzes the simplicial structure of $\Delta_{\mathcal{E}}$ and its relation to a cubical complex $\Gamma_{\mathcal{E}}$ defined by the strongly shattered sets of $\mathcal{E}$. The key property of extremal classes is that $\Gamma_{\mathcal{E}}$ is contractible [8]. It was shown in [12, D.5] that this contractibility implies a deformation retraction of $\Delta_{\mathcal{E}}$ onto $\Gamma_{\mathcal{E}}$.

Our proof relies on a different retraction (given in the full version of the paper [5]) which is tailored to our notion of dimension. This allows us to reduce the problem of computing $\text{SC}(\mathcal{E})$ to analyzing $\Gamma_{\mathcal{E}}$, where classical tools from dimension theory apply. The simplicial covering dimension of $\Gamma_{\mathcal{E}}$ is equal to the size of the largest strongly shattered subset of $\mathcal{X}$, and matches the dimension required for the upper bound. To obtain the matching lower bound, we apply Lebesgue's covering theorem [20, Theorem IV 2] to the maximal solid cube within $\Gamma_{\mathcal{E}}$.

The difference between the cases $\mathcal{E} = \{\pm 1\}^{\mathcal{X}}$ and $\mathcal{E} \neq \{\pm 1\}^{\mathcal{X}}$ is essentially explained by the following observation. If $S \subsetneq \mathcal{X}$ is strongly shattered by $\mathcal{E}$, then $\Delta_{\mathcal{E}}$ contains a full copy of the solid $|S|$ -dimensional cube, as $\mu \in \Delta_{\mathcal{E}}$ can allocate an arbitrary portion of its probability mass on coordinates outside S . In contrast, when $S = \mathcal{X}$, the normalization condition $\|\mu\|_1 = 1$ forces the entire mass to lie on S , so only the boundary of the $|S|$ -dimensional cube is realized. ◀

The extremal class of homogeneous half-spaces, discussed in Example 3, naturally connects to the notion of *sign-rank*, a key geometric measure of complexity in learning theory. The *sign-rank* of a class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$, denoted $\text{sign-rank}(\mathcal{C})$, is the minimum d such that $\mathcal{C}$ can be realized as points and homogeneous half-spaces in $\mathbb{R}^d$. Equivalently, it is the minimum rank of a real $\mathcal{C} \times \mathcal{X}$ matrix whose sign pattern encodes $\mathcal{C}$.

Since for any $P \subseteq \mathbb{R}^d$, the corresponding class of homogeneous half-spaces $\mathcal{H}_P$ is extremal and satisfies $\text{VC}(\mathcal{H}_P) \leq d$, Theorems A and B imply the following bounds as a corollary.

► **Corollary 17.** *Every binary concept class $\mathcal{C}$ over a finite domain satisfies*

$$\text{SC}(\mathcal{C}) \leq \text{sign-rank}(\mathcal{C}) - 1 \quad \text{and equivalently} \quad \text{LR}(\mathcal{C}) \leq \text{sign-rank}(\mathcal{C}).$$

The inequality $\text{LR}(\mathcal{C}) \leq \text{sign-rank}(\mathcal{C})$ was conjectured by Chase et al. [11], who verified it in the case $\text{sign-rank}(\mathcal{C}) = 2$. The full conjecture was later resolved in [4] using an algorithmic argument based on averaging multiple runs of a large-margin classifier and rounding to an ϵ -net. In contrast, Theorem B provides a purely topological proof. Furthermore, rather than the geometry of points and half-spaces, it only relies on the fact that $\mathcal{H}_P$ is extremal.

1.3 Concluding remarks

Arguably, the most intriguing open question regarding simplicial covering dimension (equivalently, list replicability [11, Section 3.4]) is whether it can be bounded by a function of the VC dimension.

► **Question 18.** *Is there a function $t : \mathbb{N} \rightarrow \mathbb{N}$ such that, for every concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ over a finite domain $\mathcal{X}$, $\text{SC}(\mathcal{C}) \leq t(\text{VC}(\mathcal{C}))$?*

By Theorem B, a positive answer to Question 6 would immediately imply a positive answer to Question 18.

The dual class $\mathcal{C}^*$ of $\mathcal{C}$ is obtained by swapping the roles of hypotheses $c \in \mathcal{C}$ and points $x \in \mathcal{X}$, or equivalently, by transposing the matrix $(c(x))_{c \in \mathcal{C}, x \in \mathcal{X}}$. It follows from [10, Theorem D] and Theorem A that

$$\text{SC}(\mathcal{C}) \geq \left\lfloor \frac{\text{VC}(\mathcal{C}^*)}{2} \right\rfloor.$$

For the cube concept class $\mathcal{Q} = \{\pm 1\}^{\mathcal{X}}$, we have

$$\text{VC}(\mathcal{Q}) = |\mathcal{X}| \quad \text{and} \quad \text{VC}(\mathcal{Q}^*) = \lfloor \log_2(|\mathcal{X}|) \rfloor.$$

Therefore, its dual class $\mathcal{Q}^*$ exhibits an exponentially large gap between the VC dimension and the simplicial covering dimension. Consequently, if the function $t(\cdot)$ of Question 18 exists, it must grow at least exponentially.

It is straightforward to verify that every class $\mathcal{C}$ satisfies $\text{vc}(\mathcal{C}^*) \leq 2^{\text{vc}(\mathcal{C})}$. Furthermore, if $\mathcal{E}$ is an extremal class, then it was shown in [9] that a stronger upper bound $\text{vc}(\mathcal{E}^*) \leq 2 \text{vc}(\mathcal{E}) + 1$ holds. It would be interesting to investigate whether an analogous relationship exists between $\text{sc}(\mathcal{C})$ and $\text{sc}(\mathcal{C}^*)$.

► **Question 19.** *What is the relation between $\text{sc}(\mathcal{C})$ and $\text{sc}(\mathcal{C}^*)$, where $\mathcal{C}^*$ denotes the dual class? For example, is it true that $\text{sc}(\mathcal{C}^*) \leq 2^{\text{sc}(\mathcal{C})}$?*

Finally, let us discuss the relationship between simplicial covering dimension and two other learning-theoretic complexity measures: Littlestone dimension and spherical dimension.

Littlestone dimension. Littlestone dimension is a refinement of VC dimension that determines the optimal number of mistakes in *online learning*. In particular, VC dimension is a lower bound on Littlestone dimension. The celebrated result of [1, Theorem 23], together with the relation between global stability and list replicability, implies that every class with Littlestone dimension d satisfies $\text{sc}(\mathcal{C}) \leq 2^{2^{O(d)}}$.

On the other hand, $\text{sc}(\mathcal{C})$ does not provide any upper bound on the Littlestone dimension: for $t > 2$, the class $\mathcal{T}_t$ of threshold functions over the domain $[t - 1]$ has Littlestone dimension $\lfloor \log_2(t) \rfloor$, yet it is an extremal case, and therefore, by Theorem B satisfies $\text{sc}(\mathcal{C}) = 1$.

Spherical dimension. The spherical dimension $\text{SD}(\mathcal{C})$ of a class $\mathcal{C}$, introduced in [12], is the largest integer d such that there exists a continuous antipodal map $f : \mathbf{S}^d \rightarrow \Delta_{\mathcal{C}}$. In the notation of Matoušek [26], this quantity is the $\mathbb{Z}_2$ -coindex of $\Delta_{\mathcal{C}}$. Spherical dimension provides a lower bound for simplicial covering dimension:

$$\text{sc}(\mathcal{C}) \geq \left\lfloor \frac{\text{SD}(\mathcal{C})}{2} \right\rfloor + 1.$$

It is open whether the list replicability number and equivalently the simplicial covering dimension can be bounded above by a function of the spherical dimension. A positive resolution of Question 18 would imply such a bound, since the spherical dimension admits tight lower bounds in terms of both VC and dual VC dimensions [12].

1.4 Paper organization

For the reader's convenience, in Section 2, we briefly review basic concepts from topological combinatorics, including simplicial and cubical complexes, abstract complexes, and their subdivisions. Furthermore, we discuss the simplicial structure of $\Delta_{\mathcal{C}}$ and the cubical complex $\Gamma_{\mathcal{C}}$ defined by the strongly shattered sets. In Section 3, we prove a few elementary facts about the simplicial covering dimension that we will use in the proofs of our main theorems. Section 4 contains the proof of Theorem A, and Section 5 contains the proof of our main result, Theorem B.

2 Background

Notation. For a positive integer $n \in \mathbb{N}$, we denote $[n] := \{1, \dots, n\}$. Since $\mathcal{X}$ is finite, all norms on $\mathbb{R}^{\mathcal{X}}$ define the same topology. In this paper, we use the ℓ_1 norm as it is more natural when working with distributions. Given $x \in X \subseteq \mathbb{R}^d$ and a radius $\epsilon > 0$, we denote the ℓ_1 -open ball of radius ϵ around x by

$$B_{\ell_1}(\epsilon, x) := \{y \in \mathbb{R}^d : \|x - y\|_1 < \epsilon\}.$$

Given $A \subseteq \mathbb{R}^d$ and $\epsilon > 0$, define

$$A^{(\epsilon)} := \{y \in \mathbb{R}^d : \|x - y\|_1 < \epsilon \text{ for some } x \in A\}.$$

For $X, Y \subseteq \mathbb{R}^d$, we also define

$$d_{\ell_1}(X, Y) := \inf_{x \in X, y \in Y} \|x - y\|_1$$

to be the minimal ℓ_1 distance between points of X and Y .

We will denote the abstract simplicial and cubical complexes with capital English letters, such as K, L, Q . We use capital Greek letters such as Γ and Δ to denote geometric simplicial and cubical complexes.

2.1 Simplicial complexes

A simplex σ is the convex hull of a *finite* affinely independent set A in $\mathbb{R}^d$. The elements of A are called the vertices of σ , and the dimension of σ is defined by $\dim(\sigma) := |A| - 1$. Following the convention of [26], we assume that the empty set is a simplex of dimension -1 .

The convex hull of any subset of the vertices of σ is called a face of σ ; note that every face is itself a simplex, and according to our convention, the empty set is a face of every simplex.

► **Definition 20** (Geometric Simplicial Complex). *A nonempty collection Δ of simplices in $\mathbb{R}^d$ is called a (geometric) simplicial complex if the following two conditions hold:*

1. *Every face of a simplex $\sigma \in \Delta$ also belongs to Δ ;*
2. *The intersection of any two simplices $\sigma_1, \sigma_2 \in \Delta$ is a face of both σ_1 and σ_2 .*

The dimension of Δ is the maximum of the dimensions of simplices in Δ .

Every geometric simplicial complex Δ gives rise to a topological space on $\|\Delta\| = \bigcup_{\sigma \in \Delta} \sigma$, called its *polyhedron* or *underlying space*. Giving each simplex its natural topology as a subspace of $\mathbb{R}^d$, one then topologizes $\|\Delta\|$ by declaring a subset $C \subseteq \|\Delta\|$ to be closed if and only if $C \cap \sigma$ is closed in σ , for each $\sigma \in \Delta$. In general, this topology does not coincide with the subspace topology inherited from $\mathbb{R}^d$; however, for *finite* simplicial complexes, the two topologies agree, and we may simply regard $\|\Delta\|$ as a subset of $\mathbb{R}^d$.

The simplicial structure of a geometric simplicial complex is captured by a combinatorial object called an abstract simplicial complex.

► **Definition 21** (Abstract simplicial complex). *An abstract simplicial complex on a set V is a collection K of finite subsets of V such that $s \in K$ and $t \subseteq s$ imply $t \in K$. The elements of K are called simplices of K , and the elements of V are called the vertices of K . The dimension of K is defined by $\dim(K) := \max_{s \in K} |s| - 1$.*

Every geometric simplicial complex Δ determines an abstract simplicial complex K with vertex set $V = V(\Delta)$, where each $s \in K$ is the set of vertices of a simplex in Δ . In this case, we say that Δ is a *geometric realization* of K , and the corresponding topological space $\|\Delta\|$ is called a *polyhedron* of K .

Conversely, every *finite* abstract simplicial complex K admits a geometric realization in some Euclidean space $\mathbb{R}^d$. For instance, one may map the vertices of K to an affinely independent set of $|V|$ points in $\mathbb{R}^{|V|-1}$ and realize each simplex $s \in K$ as the convex hull of the corresponding points. This yields a geometric realization of K in $\mathbb{R}^{|V|-1}$.

The following simple proposition implies that all polyhedra of a finite abstract simplicial complex are homeomorphic, so the associated topological space is unique up to homeomorphism.

► **Proposition 22** ([26, Proposition 1.5.4]). *Let Δ and Δ' be geometric simplicial complexes, and let K and K' be their associated abstract simplicial complexes. Suppose $f: V(K) \rightarrow V(K')$ is a simplicial map; that is, it maps every face of K to a face of K' . Then there exists a continuous map $\rho: \|\Delta\| \rightarrow \|\Delta'\|$ such that:*

- *if f is injective, then ρ is injective;*
- *if f is an isomorphism, then ρ is a homeomorphism.*

Finally, we discuss the barycentric subdivision of a simplicial complex. Suppose K is an abstract simplicial complex and Δ is its geometric realization. Let $\sigma \in \Delta$ be a simplex and let $x \in \sigma$ be a point. Then the *barycentric coordinates* of x in σ are the unique solution to the equation $\sum_{v \in V(\sigma)} \alpha_v v = x$ and $\sum_{v \in V(\sigma)} \alpha_v = 1$. It is not hard to see there is a smallest simplex σ , called the *support* of x and denoted $\text{supp}_\Delta(x)$, that contains x . In that case, all barycentric coordinates satisfy $\alpha_v > 0$.

► **Definition 23** (Abstract subdivision). *The barycentric subdivision of an abstract simplicial complex K , denoted K_1 , is defined as the abstract simplicial complex whose vertices are nonempty simplices of K , i.e. $V(K_1) = K \setminus \{\emptyset\}$, and whose simplices are chains of non-empty simplices of K ordered by inclusion.*

The geometric realization of K_1 is natural.

► **Definition 24** (Geometric subdivision). *Let Δ be the geometric realization of K . For each nonempty simplex $\sigma \in \Delta$, define v_σ to be its barycenter, that is $v_\sigma = \frac{1}{|\sigma|} \sum_{v \in \sigma} v$. Having defined the coordinates of each vertex in the subdivision, we extend linearly for points in each simplex and obtain a geometric realization of K_1 , denoted Δ_1 .*

Note that $\|\Delta_1\|$ and $\|\Delta\|$ are homeomorphic.

2.2 Cubical complexes

Cubical complexes serve as a combinatorial and geometric framework analogous to simplicial complexes, but built from cubes instead of simplices.

► **Definition 25** (Elementary cubes). *An elementary interval in $\mathbb{R}$ is a closed interval of the form $[a, a]$ or $[a, a+1]$ for some $a \in \mathbb{Z}$. An elementary cube in $\mathbb{R}^d$ is a product $\kappa = I_1 \times \cdots \times I_d$, where each I_i is an elementary interval. The dimension of κ is the number of factors I_i of the form $[a, a+1]$.*

► **Definition 26** (Geometric cubical complex). *A geometric cubical complex Γ in $\mathbb{R}^d$ is a finite non-empty collection of elementary cubes such that:*

1. *If $\kappa \in \Gamma$ and τ is a face of κ (obtained by replacing one or more factors $[a, a+1]$ with $[a, a]$ or $[a+1, a+1]$), then $\tau \in \Gamma$;*
 2. *If $\kappa_1, \kappa_2 \in \Gamma$, then $\kappa_1 \cap \kappa_2$ is either empty or is a face of both κ_1 and κ_2 .*
- The polyhedron $\|\Gamma\|$ is the union of all cubes in Γ , equipped with the subspace topology from $\mathbb{R}^d$.*

Note that sometimes we allow the elementary geometric cubes in a geometric cubical complex to be scaled by some factor. Nevertheless, this does not change any of the topological properties of the cubical complex.

We proceed with the corresponding theory of abstract cubical complexes. The *standard abstract n -cube* is represented by the set $\{0, 1\}^n$. A *face* of the standard abstract n -cube is a product $q_1 \times \cdots \times q_n$, where each q_i is a non-empty subset of $\{0, 1\}$.

► **Definition 27** (Abstract cubical complex [19]). *An abstract cubical complex on a non-empty vertex set V is a collection Q of non-empty subsets of V , called cubes, satisfying:*

1. Q is a cover of V .
2. For any $q_1, q_2 \in Q$, $q_1 \cap q_2 \in Q$ or $q_1 \cap q_2 = \emptyset$.
3. For any $q \in Q$ there is a bijection $\phi_q : q \rightarrow \{0, 1\}^n$, for some n , satisfying: if $q_1 \subseteq q$, then $q_1 \in Q$ if and only if $\phi_q(q_1)$ is a face of $\{0, 1\}^n$.

The elements of V are called the vertices or 0-cubes of Q . In general, we say $q \in Q$ is an n -cube if $|q| = 2^n$.

As for simplicial complexes, each geometric cubical complex Γ has a corresponding abstract cubical complex Q : each abstract cube represents the set of vertices of its respective geometric cube. We again say Γ is a geometric realization of Q . To better understand the structure of an abstract cubical complex, it is often useful to study its barycentric subdivision, which is a *simplicial complex* defined as follows.

► **Definition 28** (Subdivision of a cubical complex [19]). *The barycentric subdivision of an abstract cubical complex Q , denoted Q_1 , is the abstract simplicial complex whose vertices are the cubes in Q and whose simplices are chains of cubes ordered by inclusion.*

2.3 Simplicial complex of a class

To describe the simplicial structure of the set of realizable distributions, we need a few notations.

A *partial concept* on a domain $\mathcal{X}$ is a labeling $h : \mathcal{X} \rightarrow \{\pm 1, *\}$, where $h(x) = *$ means that h is *undefined* at x . The *support* of h is $\text{supp}(h) := \{x \in \mathcal{X} : h(x) \neq *\}$. Denote by h_* the partial concept with empty support.

If $c \in \{\pm 1\}^{\mathcal{X}}$ and $h \in \{\pm 1, *\}^{\mathcal{X}}$ is a partial concept such that c and h agree on $\text{supp}(h)$, then c is called a *completion* of h . More generally, we say that a partial concept h_1 *extends* h_2 if $\text{supp}(h_2) \subseteq \text{supp}(h_1)$ and h_1 and h_2 agree on $\text{supp}(h_2)$. We denote this by $h_1 \geq h_2$.

Given a concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$, we say that a partial concept h is *realizable by $\mathcal{C}$* if there exists a concept $c \in \mathcal{C}$ that is a completion of h .

► **Definition 29** (Abstract simplicial complex of a class). *Given a concept class $\mathcal{C}$, let $\mathcal{C}^\bullet \subset \{\pm 1, *\}^{\mathcal{X}}$ be the set of all realizable partial concepts of $\mathcal{C}$. For each $h \in \mathcal{C}^\bullet$, define the set*

$$s_h := \{(x, h(x)) : x \in \text{supp}(h)\}.$$

The collection $\{s_h\}_{h \in \mathcal{C}^\bullet}$ forms an abstract simplicial complex, which we denote by $D_{\mathcal{C}}$.

In particular, note that $V(D_{\mathcal{C}}) \subseteq \mathcal{X} \times \{\pm 1\}$ and that the all- $*$ labeling h_* corresponds to the empty set. In addition, the maximal faces of $D_{\mathcal{C}}$ are exactly those defined by concepts in $\mathcal{C}$, i.e., s_c where $c \in \mathcal{C}$.

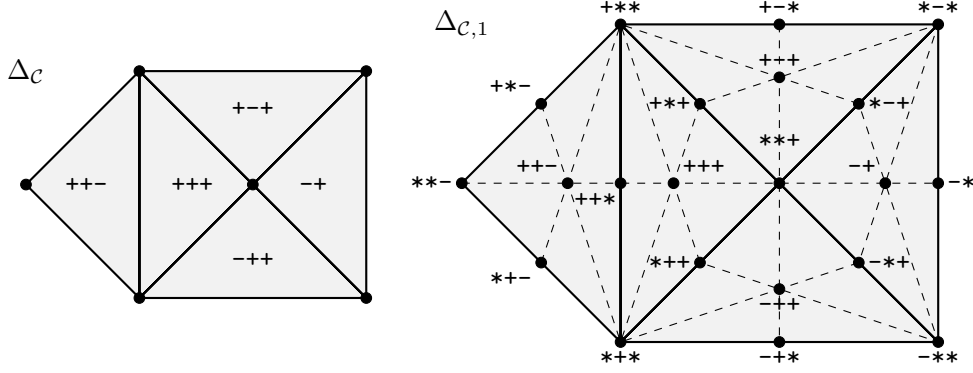
Since there is a bijective correspondence between simplices of $D_{\mathcal{C}}$ and partial concepts realizable by $\mathcal{C}$, we will often treat $h \in \mathcal{C}^\bullet$ directly as simplex of $D_{\mathcal{C}}$.

Notice $D_{\mathcal{C}}$ has a natural geometric realization in $\mathbb{R}^{\mathcal{X}}$. Indeed, let $\{\mathbf{e}_x\}_{x \in \mathcal{X}}$ be the *standard basis* of $\mathbb{R}^{\mathcal{X}}$.

► **Definition 30** (Geometric simplicial complex of a class). *For each realizable partial concept $h \in \mathcal{C}^\bullet$, define the geometric simplex*

$$\sigma_h := \text{conv}(\{\text{sign}(h(x))\mathbf{e}_x : x \in \text{supp}(h)\}),$$

which corresponds to the abstract simplex s_h in $D_{\mathcal{C}}$. The collection $\{\sigma_h : h \in \mathcal{C}^\bullet\}$ forms a geometric simplicial complex in $\mathbb{R}^{\mathcal{X}}$, denoted $\Delta_{\mathcal{C}}$.



■ **Figure 2** The simplicial complexes Δ_C and $\Delta_{C,1}$ for $\mathcal{C} = \{++-, +++, +--, -, -++\}$. Here, + and - are shorthand for +1 and -1, respectively.

For convenience, we will also use Δ_C (instead of $\|\Delta_C\|$) to denote $\bigcup_{h \in \mathcal{C}^\bullet} \sigma_h$, i.e., the underlying polyhedron of Δ_C . The meaning will be clear from context. Note that this is consistent with the definition of Δ_C in (4).

Similarly to the abstract setting, the maximal simplices σ_h are exactly those with $h \in \mathcal{C}$. Also note that for $h \in \mathcal{C}$, we have σ_h equals B_h , from (5), as subsets of $\mathbb{R}^X$. Nevertheless, we prefer to keep their notation separate in order to differentiate between the two.

Next, we consider the subdivision of D_C , denoted $D_{C,1}$. Recall that in a *simplicial subdivision*, simplices correspond to chains of *non-empty* simplices in the original complex (Definition 23). Here, we call a sequence of realizable partial concepts $h_1, \dots, h_k$ a *chain* if $h_1 \geq \dots \geq h_k$, i.e., each partial concept extends the next one in the sequence. The simplices of $D_{C,1}$ are in bijective correspondence with chains of realizable partial concepts in $\mathcal{C}^\bullet \setminus \{h_*\}$ where h_* is excluded as it corresponds to the empty simplex. In particular, $V(D_{C,1}) = \mathcal{C}^\bullet \setminus \{h_*\}$.

Since $D_{C,1}$ is a subdivision of D_C , it inherits a natural geometric realization. To a vertex (partial concept) $h \in \Delta_{C,1}$ we associate the uniform distribution over $\text{supp}(h)$. This extends linearly to all points in $\Delta_{C,1}$. For instance, if μ is a point in the simplex $h_1 \geq \dots \geq h_k$, then $\mu = \sum_{i=1}^k \alpha_i h_i$ where the barycentric coordinates satisfy $\sum_{i=1}^k \alpha_i = 1$ and each partial concept h_i represents the coordinates given by the uniform distribution over $\text{supp}(h_i)$. We denote by $\text{supp}_{\Delta_{C,1}}(\mu)$ the minimal simplex $h_1 \geq \dots \geq h_k$ containing μ , i.e., the simplex for which all barycentric coordinates $\alpha_i > 0$.

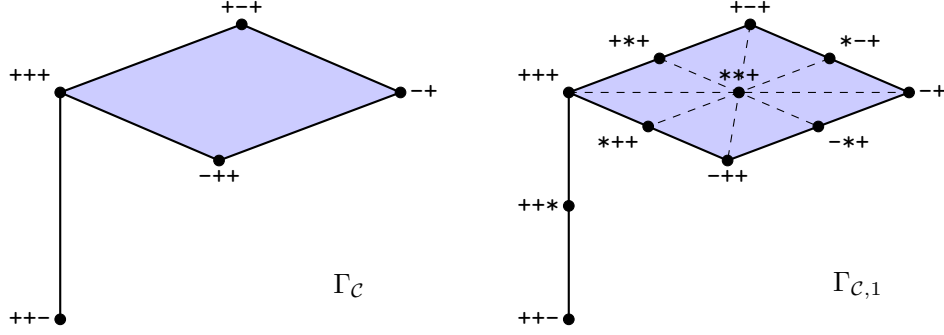
Finally, note Δ_C and $\Delta_{C,1}$ are equal as subsets of $\mathbb{R}^X$ and are, in particular, homeomorphic.

2.4 The cubical complex of a class

► **Definition 31** (Abstract cubical complex of a class). *Let $\mathcal{C}$ be a concept class. The abstract cubical complex $\mathcal{Q}_C$ of $\mathcal{C}$ consists of all subsets $q = \{c_1, \dots, c_k\}$ of $\mathcal{C}$ such that q strongly shatters some subset $S \subseteq \mathcal{X}$.*

It is not difficult to check that $\mathcal{Q}_C$ satisfies the properties in the definition of an abstract cubical complex (Definition 27). Note that the set of vertices (0-cubes) is $V(\mathcal{Q}_C) = \mathcal{C}$. In general, if $q = \{c_1, \dots, c_k\}$ strongly shatters a set $S \subseteq \mathcal{X}$, then q is an $|S|$ -cube. Finally, observe q can be represented by the partial concept h_q defined by

$$h_q(x) = \begin{cases} * & \text{if } x \in S \\ c_1(x) = \dots = c_k(x) & \text{otherwise} \end{cases}$$



■ **Figure 3** The cubical complex Γ_C and the simplicial complex $\Gamma_{C,1}$ for $\mathcal{C} = \{++-, +++, +--, --+, ---\}$.

as the set $\{c_1, \dots, c_k\}$ consists of all possible completions of h_q to a concept in $\{\pm 1\}^{\mathcal{X}}$. In fact, there is a bijective correspondence between cubes of $\mathcal{Q}_C$ and partial concepts $h \in \{\pm 1, *\}^{\mathcal{X}}$ whose every completion belongs to $\mathcal{C}$. For this reason, we sometimes treat partial concepts directly as cubes.

Next, we construct a natural geometric realization of $\mathcal{Q}_C$.

► **Definition 32** (Geometric cubical complex of a class). *For each abstract cube $q = \{c_1, \dots, c_k\}$ in $\mathcal{Q}_C$, define an elementary geometric cube κ_q by*

$$\kappa_q := \{y \in [-1, 1]^{\mathcal{X}} \mid y(x) = h_q(x) \text{ for } x \in \text{supp}(h_q)\}.$$

The collection $\{\kappa_q\}_{q \in \mathcal{Q}_C}$ is a geometric cubical complex denoted by Γ_C .

Note each concept (0-cube) $c \in \mathcal{C}$ corresponds to a point $y_c \in [-1, 1]^{\mathcal{X}}$ on the ℓ_∞ sphere, defined by $y_c = (\text{sign}(c(x)) : x \in X)$.

In general, observe that the dimension of each cube $\kappa_q \in \Gamma_C$ equals the cardinality of the subset $S \subseteq \mathcal{X}$ that is strongly shattered by q .

For convenience, we will use Γ_C (instead of $\|\Gamma_C\|$) to denote $\bigcup_{q \in \mathcal{Q}_C} \kappa_q$, i.e., the underlying polyhedron of Γ_C . The meaning will be clear from context.

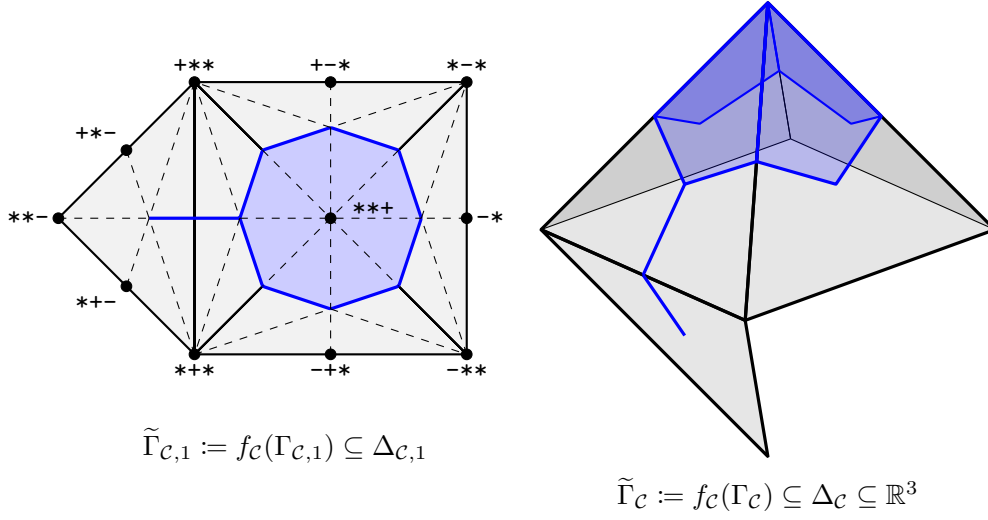
Next, we consider the barycentric subdivision of the abstract cubical complex of a concept class $\mathcal{C}$, denoted $\mathcal{Q}_{C,1}$. By Definition 28, $\mathcal{Q}_{C,1}$ is a simplicial complex whose vertices are the cubes of $\mathcal{Q}_C$ and whose simplices are chains of ascending cubes. Hence, there is a bijective correspondence between simplices of $\mathcal{Q}_{C,1}$ and chains of partial concepts $h_1 \geq \dots \geq h_k$ such that for any i , all completions of h_i are in $\mathcal{C}$.

Now $\mathcal{Q}_{C,1}$ also has a natural geometric realization. Indeed, map each vertex (partial concept) h to

$$y_h = \begin{cases} \text{sign}(h(x)) & \text{if } x \in \text{supp}(h) \\ 0 & \text{otherwise} \end{cases}$$

and extend linearly for each simplex $h_1 \geq \dots \geq h_k$ in $\mathcal{Q}_{C,1}$. We denote this geometric realization of $\mathcal{Q}_{C,1}$ by $\Gamma_{C,1}$. In addition, note $\Gamma_{C,1}$ is consistent with the geometric realization Γ_C in the sense that they are equal as subsets of $\mathbb{R}^{\mathcal{X}}$ and are, in particular, homeomorphic. Hence, one may study the geometry of Γ_C from both a simplicial and cubical perspective.

► **Lemma 33** ([12]). *Suppose $\mathcal{C} \neq \{\pm 1\}^{\mathcal{X}}$ is a concept class that is not the binary cube. Then $\mathcal{Q}_{C,1}$ is a full subcomplex of $D_{C,1}$, that is, if s is a simplex in $D_{C,1}$ and for all vertices $v \in s$, we have $v \in \mathcal{Q}_{C,1}$, then $s \in \mathcal{Q}_{C,1}$.*



■ **Figure 4** The embeddings $\Gamma_{C,1} \hookrightarrow \Delta_{C,1}$ and $\Gamma_C \hookrightarrow \Delta_C$ for the concept class $\mathcal{C} = \{++-, +++, +--, --, -++\}$.

Proof. Since $\mathcal{C} \neq \{\pm 1\}^{\mathcal{X}}$, the all- $*$ labelling h_* is not a cube. Hence, the vertex set $V(Q_{C,1}) \subseteq \mathcal{C}^\bullet \setminus \{h_*\} = V(D_{C,1})$. Each chain of cubes $h_1 \geq \dots \geq h_k$ (simplex of $Q_{C,1}$) is also a chain of realizable partial concepts (simplex of $D_{C,1}$). Thus, $Q_{C,1}$ is a subcomplex of $D_{C,1}$. Finally, if $h_1 \geq \dots \geq h_k$ is a chain of realizable partial concepts such that each h_i is a cube, then we immediately get a chain of cubes. We conclude $Q_{C,1}$ is a full subcomplex. ◀

The reason $Q_{C,1}$ does not embed into $D_{C,1}$ for $\mathcal{C} = \{\pm 1\}^{\mathcal{X}}$ is that h_* is not a vertex of $D_{C,1}$ by definition, while it is a vertex of $Q_{C,1}$ because all completions of h_* are in $\{\pm 1\}^{\mathcal{X}}$. Thus, so long as $\mathcal{C}$ is not the binary cube, there is a simplicial embedding $\Gamma_{C,1} \hookrightarrow \Delta_{C,1}$ which can also be thought of as an embedding $\Gamma_C \hookrightarrow \Delta_C$. This embedding can be realized as ℓ_1 normalization:

$$f_C: y \mapsto y / \|y\|_1. \quad (6)$$

For $\mathcal{C} \neq \{\pm 1\}^{\mathcal{X}}$, we shall denote

$$\tilde{\Gamma}_C := f_C(\Gamma_C) \subseteq \Delta_C, \text{ and } \tilde{\Gamma}_{C,1} := f_C(\Gamma_{C,1}) \subseteq \Delta_{C,1}. \quad (7)$$

3 Dimension theory toolkit

In this section, we prove a few elementary dimension theory facts, which we will invoke repeatedly in the proofs of our main theorems (Theorems A and B).

Recall that given $A \subseteq \mathbb{R}^n$ and $\epsilon > 0$, we defined

$$A^{(\epsilon)} := \{y \in \mathbb{R}^n : \|x - y\|_1 < \epsilon \text{ for some } x \in A\}.$$

As discussed in Remark 8, when working with topological spaces of the form $X \subseteq \mathbb{R}^n$, we write $\{A_i\}_{i \in I} \prec \{B_j\}_{j \in J}$ to mean $\{A_i \cap X\}_{i \in I} \prec \{B_j \cap X\}_{j \in J}$.

The following lemma gives a useful metric tool for constructing refinements.

► **Lemma 34.** *Let $\{A_i\}_{i \in I} \prec \{B_j\}_{j \in J}$ be covers of $X \subseteq \mathbb{R}^n$ such that*

- *$A_i \cap X$ are open for all $i \in I$;*
- *J is finite and $B_j \cap X$ are compact in X for all $j \in J$.*

There is some $\alpha > 0$ such that $\{A_i\}_{i \in I} \prec \{B_j^{(\alpha)}\}_{j \in J}$.

Proof. Since $\{B_j\}_{j \in J}$ is a refinement of the open cover $\{A_i\}_{i \in I}$, each B_j is contained in some open A_i . If B_j is empty, then $B_j^{(\alpha)}$ is also empty and is therefore contained in A_i for any $\alpha > 0$. Likewise, if $A_i \supseteq X$, then $B_j^{(\alpha)} \cap X \subseteq A_i$ for any $\alpha > 0$.

Otherwise, $X \setminus A_i$ is a nonempty closed set, $B_j \cap X$ is a nonempty compact set, and $B_j \cap (X \setminus A_i) = \emptyset$. Together these imply that the distance $\beta_j := d_{\ell_1}(X \setminus A_i, B_j \cap X)$ is positive, and so $B_j^{(\beta_j)} \cap X \subseteq A_i$. Since there are finitely many $j \in J$, taking $\alpha := \min_{j \in J} \beta_j$ completes the proof. ◀

We can pass from an open cover to a closed refinement without increasing order.

► **Lemma 35.** *Any open cover $\{U_i\}_{i \in I}$ of a compact set $X \subseteq \mathbb{R}^n$ has a closed refinement $\{F_i\}_{i \in I}$, and in particular, $\text{ord}(\{F_i\}_{i \in I}) \leq \text{ord}(\{U_i\}_{i \in I})$.*

Proof. Since $\{U_i\}_{i \in I}$ is a cover of X , for each $x \in X$ there exists an $i \in I$ such that $x \in U_i$. Because U_i is open and X is a subset of $\mathbb{R}^n$, there exists some open V_x such that

$$x \in V_x \subseteq \overline{V_x} \subseteq U_i.$$

By the compactness of X , the open cover $\{V_x\}_{x \in X}$ admits a finite subcover $\{V_{x_j}\}_{j \in J}$. Hence, $\{\overline{V_{x_j}}\}_{j \in J}$ is a finite closed cover. Using this cover, we define for each $i \in I$ a closed set

$$F_i := \bigcup \{ \overline{V_{x_j}} : j \in J, \overline{V_{x_j}} \subseteq U_i \}.$$

By construction, each $\overline{V_{x_j}}$ is contained in some F_i , and so $\{F_i\}_{i \in I}$ is a closed cover of X . Furthermore, for each distinct F_i and point $x \in X$ such that F_i contains x , there is a distinct U_i such that $x \in F_i \subseteq U_i$. We conclude that $\{F_i\}_{i \in I}$ refines $\{U_i\}_{i \in I}$ without increasing order, as claimed. ◀

Here we present the usual formulation of the Lebesgue covering theorem with closed sets, which can be used to derive the open version stated in Theorem 9.

► **Theorem 36** (Lebesgue covering theorem [20, Theorem IV 2.]). *Suppose a d -dimensional cube $[-1, 1]^d \subseteq \mathbb{R}^d$ is covered by a finite family $\mathcal{A}$ of closed sets, none of which contains points of opposite faces of the cube. Then $\text{ord}(\mathcal{A}) \geq d$.*

Proof of Theorem 9. Suppose that $\mathcal{A}$ is a finite family of open sets covering the cube $[-1, 1]^d$ such that no member of $\mathcal{A}$ contains points of opposite faces of the cube. By Lemma 35, there exists a closed refinement $\mathcal{F}$ of $\mathcal{A}$ that is also a cover, such that $\text{ord}(\mathcal{F}) \leq \text{ord}(\mathcal{A})$.

Since $\mathcal{F}$ refines $\mathcal{A}$, it also has no member set containing points of opposite faces of the cube. Applying Theorem 36 to the closed cover $\mathcal{F}$ yields that $\text{ord}(\mathcal{F}) \geq d$. We conclude that $\text{ord}(\mathcal{A}) \geq d$ as claimed. ◀

Finally, we record the following useful lemma, which shows that for any closed cover of order d of a compact space, there exists a sufficiently small radius $\alpha > 0$ such that every α -neighbourhood around any point meets at most $d + 1$ sets of the cover.

► **Lemma 37.** *If a finite closed cover $\mathcal{F}$ of a compact set $X \subseteq \mathbb{R}^n$ has order d , then there is some $\alpha > 0$ such that, for any $x \in X$, the ball $B_{\ell_1}(\alpha, x) \cap X$ intersects at most $d + 1$ sets of $\mathcal{F}$.*

Proof. For the sake of contradiction, assume that for every $\alpha > 0$ there is some $x \in X$ such that $B_{\ell_1}(\alpha, x) \cap X$ intersects at least $d + 2$ sets of $\mathcal{F}$. Then we may pick some sequence $\{x_k\}_{k=1}^\infty$ such that $B_{\ell_1}(\frac{1}{2^k}, x_k) \cap X$ intersects at least $d + 2$ sets of $\mathcal{F}$ for every $k \in \mathbb{N}$.

Since X is compact, $\{x_k\}_{k=1}^\infty$ has a subsequence $\{x_{k_i}\}_{i=1}^\infty$ which converges to some $x^* \in X$. Furthermore $\mathcal{F}$ is finite, so there are sets $F_1, \dots, F_{d+2} \in \mathcal{F}$, each of which intersects $B_{\ell_1}(\frac{1}{2^{k_i}}, x_{k_i}) \cap X$ for infinitely many $i \in \mathbb{N}$. Putting these observations together, we see that for any $j \in [d + 2]$, we have

$$d_{\ell_1}(F_j, x^*) \leq d_{\ell_1}(F_j, x_{k_i}) + d_{\ell_1}(x_{k_i}, x^*),$$

where the sum of the right-hand terms can be made arbitrarily small for an appropriate choice of $i \in \mathbb{N}$. Thus, each of the closed sets $F_1, \dots, F_{d+2}$ has x^* as a limit point, whereby x^* is contained in $d + 2$ sets of $\mathcal{F}$, which contradicts that $\mathcal{F}$ has order d . ◀

4 Theorem A: Simplicial covering dimension and replicability

In the introduction, when discussing the problem of learning a class $\mathcal{C}$, we assumed that the learner is allowed to output any hypothesis $h \in \{\pm 1\}^{\mathcal{X}}$ with small loss. However, in some settings, one restricts the learner to output hypotheses from some class $\mathcal{H} \supseteq \mathcal{C}$. For example, in *proper learning*, the output of the learner must be from the original class $\mathcal{C}$.

A more general setting. Let us denote by $\text{LR}_{\mathcal{H}}(\mathcal{C})$ the list replicability number of $\mathcal{C}$, with the extra requirement that the list in Definition 1 must consist of hypotheses in $\mathcal{H}$. With this notation, the unrestricted case is $\text{LR}(\mathcal{C}) = \text{LR}_{\{\pm 1\}^{\mathcal{X}}}(\mathcal{C})$. Similarly, define

$$\text{SC}_{\mathcal{H}}(\mathcal{C}) := \text{SC}_{\Delta_{\mathcal{H}}}(\Delta_{\mathcal{C}}),$$

so that $\text{SC}(\mathcal{C}) = \text{SC}_{\{\pm 1\}^{\mathcal{X}}}(\mathcal{C})$ in the unrestricted setting.

The following theorem, a generalization of Theorem A, establishes a precise quantitative connection between simplicial covering dimension and list replicability.

► **Theorem 38** (General form of Theorem A). *For any finite domain $\mathcal{X}$, concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$, and hypothesis class $\mathcal{H} \subseteq \{\pm 1\}^{\mathcal{X}}$ such that $\mathcal{C} \subseteq \mathcal{H}$, we have $\text{LR}_{\mathcal{H}}(\mathcal{C}) = \text{SC}_{\mathcal{H}}(\mathcal{C}) + 1$.*

Proof. For convenience, let $d := \text{SC}_{\mathcal{H}}(\mathcal{C})$ and let $L := \text{LR}_{\mathcal{H}}(\mathcal{C})$. Recall that by an abuse of notation, we use $\Delta_{\mathcal{C}}$ to denote both the simplicial complex and its polyhedron $\|\Delta_{\mathcal{C}}\|$. Since we are working with the topological space $\Delta_{\mathcal{C}} \subseteq \mathbb{R}^{\mathcal{X}}$, recall from Remark 8 that we write $\mathcal{A} \prec \mathcal{B}$ to mean $\{A_i \cap \Delta_{\mathcal{C}}\}_{i \in I} \prec \{B_j \cap \Delta_{\mathcal{C}}\}_{j \in J}$.

First, we will show that $d \leq L - 1$. This amounts to proving that if $\mathcal{U}$ is an open cover of $\Delta_{\mathcal{C}}$, such that $\mathcal{U} \prec \{\sigma\}_{\sigma \in \Delta_{\mathcal{H}}}$, then $\mathcal{U}$ has a refinement of order at most $L - 1$ that is an open cover of $\Delta_{\mathcal{C}}$. We begin by applying Lemma 34 to produce an $\alpha > 0$ such that $\mathcal{U} \prec \{\sigma^{(\alpha)}\}_{\sigma \in \Delta_{\mathcal{H}}}$. Fix $0 < \epsilon < \frac{\alpha}{4}$ and $0 < \delta < \frac{1}{2(L+1)}$. Since the list replicability number of $\mathcal{C}$ with respect to $\mathcal{H}$ is L , there exists an (ϵ, L) -list replicable learner $\mathcal{A}$ for $\mathcal{C}$ with outputs in $\mathcal{H}$ and with sample complexity $n := n(\epsilon, \delta)$. Using this learning rule, for each $h \in \mathcal{H}$, define an open set

$$V_h := \left\{ \mu \in \Delta_{\mathcal{C}} : \Pr_{S \sim \mu^n}[\mathcal{A}(S) = h] > \frac{1 - 2\delta}{L} \text{ and } \text{loss}_{\mu}(h) < 2\epsilon \right\}.$$

22:18 Simplicial Covering Dimension of Extremal Concept Classes

We will show that the family $\{V_h\}_{h \in \mathcal{H}}$ is an open cover of $\Delta_{\mathcal{C}}$, and is also indeed a refinement of $\mathcal{U}$ with order at most $L - 1$.

The sets V_h are open in $\Delta_{\mathcal{C}}$ by continuity of $\Pr_{S \sim \mu^n}[\mathcal{A}(S) = h]$ and $\text{loss}_{\mu}(h)$ in μ . To see why $\{V_h\}_{h \in \mathcal{H}}$ covers $\Delta_{\mathcal{C}}$, first recall that every $\mu \in \Delta_{\mathcal{C}}$ is a distribution realizable by $\mathcal{C}$, and so the list replicable learner $\mathcal{A}$ guarantees a list $\{h_1, \dots, h_L\} \subseteq \mathcal{H}$ such that

$$\Pr_{S \sim \mu^n}[\mathcal{A}(S) \in \{h_1, \dots, h_L\}] \geq 1 - \delta$$

and $\text{loss}_{\mu}(h_i) \leq \epsilon$ for each $i \in [L]$. It follows that there is some $h \in \{h_1, \dots, h_L\}$ such that

$$\Pr_{S \sim \mu^n}[\mathcal{A}(S) = h] \geq \frac{1 - \delta}{L} > \frac{1 - 2\delta}{L} \quad \text{and} \quad \text{loss}_{\mu}(h) \leq \epsilon < 2\epsilon,$$

whereby $\mu \in V_h$. Hence, $\{V_h\}_{h \in \mathcal{H}}$ covers $\Delta_{\mathcal{C}}$.

As for the order of $\{V_h\}_{h \in \mathcal{H}}$, suppose for contradiction that $h_1, \dots, h_{L+1} \in \mathcal{H}$ are distinct hypotheses such that $\mu \in \bigcap_{i \in [L+1]} V_{h_i}$ for some $\mu \in \Delta_{\mathcal{C}}$. It follows that the probability of outputting each of the distinct hypotheses $h_1, \dots, h_{L+1}$ is greater than $\frac{1-2\delta}{L}$. As these events are disjoint, we deduce that

$$\Pr_{S \sim \mu^n}[\mathcal{A}(S) \in \{h_1, \dots, h_{L+1}\}] > (L+1) \cdot \frac{1-2\delta}{L} > \frac{L+1}{L} \cdot \left(1 - \frac{1}{(L+1)}\right) = 1,$$

which gives the desired contradiction.

It remains to argue that

$$\mathcal{U} \prec \{\sigma^{(\alpha)}\}_{\sigma \in \Delta_{\mathcal{H}}} \prec \{V_h\}_{h \in \mathcal{H}}.$$

The first refinement is already verified by the choice of α , and the second refinement holds by construction since $\text{loss}_{\mu}(h) < 2\epsilon < \frac{\alpha}{2}$ and

$$\text{loss}_{\mu}(h) = \sum_{\substack{x \in \text{supp}(\mu) \\ h(x) \neq \text{sign}(\mu(x))}} |\mu(x)| = \frac{1}{2} \|\mu - |\mu|h\|_1 \geq \frac{1}{2} d_{\ell_1}(\mu, \sigma_h),$$

imply $\mu \in \sigma_h^{(\alpha)}$.

The second half of the proof is to show that $L \leq d + 1$. To this end, fix $\epsilon > 0$. If the relative simplicial covering dimension of $\mathcal{C}$ with respect to $\mathcal{H}$ is d , then there exists a refinement $\mathcal{U}$ of $\{B_h^{(\epsilon/2)}\}_{h \in \mathcal{H}}$ such that $\mathcal{U}$ has order no more than d and is an open cover of $\Delta_{\mathcal{C}}$. Now pick any ordering on $\mathcal{H}$ and for $h \in \mathcal{H}$ let

$$V_h := \bigcup \left\{ U \in \mathcal{U} : h \text{ is the first hypothesis in } \mathcal{H} \text{ satisfying } U \subseteq B_h^{(\epsilon/2)} \right\}.$$

Note that by definition, $\mathcal{V} := \{V_h\}_{h \in \mathcal{H}}$ is a refinement of $\{B_h^{(\epsilon/2)}\}_{h \in \mathcal{H}}$ and has order no more than d . By Lemma 35, this open cover $\mathcal{V}$ has a closed refinement $\mathcal{F}$ (also indexed by $\mathcal{H}$) of order at most d . Lastly, applying Lemma 37 guarantees a $\beta > 0$ such that $\{\nu \in \Delta_{\mathcal{C}} : \|\nu - \mu\|_1 \leq \beta\}$ intersects at most $d + 1$ sets of $\mathcal{F}$ for any $\mu \in \Delta_{\mathcal{C}}$.

Now we are ready to give an $(\epsilon, d + 1)$ -list replicable learner for $\mathcal{C}$. For any distribution $\mu \in \Delta_{\mathcal{C}}$ and sample $S \sim \mu^k$ for some $k \in \mathbb{N}$, let $\hat{\mu}$ denote the empirical estimate of μ using S . Since the distributions in $\Delta_{\mathcal{C}}$ are defined on a finite domain, there exists, for any $\delta > 0$, some positive integer $n := n(\epsilon, \beta, \delta)$ such that

$$\Pr_{S \sim \mu^n} \left[\|\mu - \hat{\mu}\|_1 < \min \left(\frac{\epsilon}{2}, \beta \right) \right] \geq 1 - \delta$$

for all $\mu \in \Delta_{\mathcal{C}}$.

Given a replicability parameter $\delta > 0$, an unknown distribution μ , and a sample $S \sim \mu^n$, the learning rule is as follows:

1. Let $\hat{\mu}$ be the empirical estimate of μ using S .
2. Select an arbitrary closed set $F_h \in \mathcal{F}$ containing $\hat{\mu}$.
3. Output h .

First, let us check accuracy. We have that

$$\text{loss}_\mu(h) \leq \text{loss}_{\hat{\mu}}(h) + \|\mu - \hat{\mu}\|_1,$$

where $\text{loss}_{\hat{\mu}}(h) \leq \frac{\epsilon}{2}$ because $\hat{\mu} \in F_h \subseteq B_h^{(\epsilon/2)}$, and $\|\mu - \hat{\mu}\|_1 \leq \epsilon/2$ with probability at least $1 - \delta$. Second, we check list-replicability. With probability at least $1 - \delta$ we have $\|\mu - \hat{\mu}\|_1 < \beta$, in which case there are by choice of β at most $d + 1$ closed sets of $\mathcal{F}$ that could contain $\hat{\mu}$. ◀

5 Theorem B: Simplicial covering dimension of extremal classes

In this section, we classify the simplicial covering dimension $\text{sc}(\mathcal{E})$ of extremal classes. The argument is split between a lower bound given in Lemma 40 and a matching upper bound relying on Theorem 41. For the proof of Theorem 41, we refer the reader to the full version of the paper [5]. The classification of $\text{LR}(\mathcal{E})$ follows as a direct consequence of Theorem 38. Before presenting the proofs, we give a brief overview and discuss why the binary cube $\{\pm 1\}^{\mathcal{X}}$ must be treated as a separate case in both of our bounds.

5.1 Discussion and overview of the proof

The cubical complex $\Gamma_{\mathcal{E}}$ of a class $\mathcal{E}$ is composed of the union of all its cubes, where a cube of dimension k corresponds to a strongly shattered set of size k . By Theorem 10, $\Gamma_{\mathcal{E}}$ has topological dimension equal to that of any maximum cube, which in the case of extremal $\mathcal{E}$ is exactly $\text{vc}(\mathcal{E})$.

► **Proposition 39** (Dimension of the cubical complex). *For any concept class $\mathcal{C}$, the topological dimension of its cubical complex $\Gamma_{\mathcal{C}}$ is equal to the size of the largest strongly shattered set by $\mathcal{C}$. In particular, if $\mathcal{E}$ is an extremal class, then*

$$\dim(\Gamma_{\mathcal{E}}) = \text{vc}(\mathcal{E}).$$

In view of Proposition 39, the lower bound of Theorem B can be achieved by embedding $\Gamma_{\mathcal{E}}$ into $\Delta_{\mathcal{E}}$ and applying Lebesgue's covering theorem (Theorem 9). The upper bound follows from a retraction from $\Delta_{\mathcal{E},1}$ to the embedding of $\Gamma_{\mathcal{E},1}$, which we denoted by $\tilde{\Gamma}_{\mathcal{E},1}$ in (7). In particular, each vertex $h \in \Delta_{\mathcal{E},1}$ is mapped to a point in a simplex of $\tilde{\Gamma}_{\mathcal{E},1}$ supported on concepts extending h . The topological dimension of $\Gamma_{\mathcal{E}}$ is used to construct a suitable open cover of order $\text{vc}(\mathcal{E})$, which is pulled back to $\Delta_{\mathcal{E}}$ without increasing the order.

When $\mathcal{E} = \{\pm 1\}^{\mathcal{X}}$, however, the embedding given in Lemma 33 fails. In fact, $\Gamma_{\{\pm 1\}^{\mathcal{X}}}$ is the geometric cube of dimension $|\mathcal{X}|$, whereas $\Delta_{\{\pm 1\}^{\mathcal{X}}}$ is the cross-polytope of dimension $|\mathcal{X}| - 1$, so no one-to-one embedding exists. We can recover our argument by instead considering the boundary of $\Gamma_{\{\pm 1\}^{\mathcal{X}}}$, which is homeomorphic to $\Delta_{\{\pm 1\}^{\mathcal{X}}}$, and has dimension one less than the cube.

$$\dim(\partial\Gamma_{\{\pm 1\}^{\mathcal{X}}}) = \text{vc}(\{\pm 1\}^{\mathcal{X}}) - 1.$$

This modification leads to the off-by-one case in Theorem B.

5.2 Lower bound for simplicial covering dimension of extremal classes

► **Lemma 40.** *Let $\mathcal{E} \subseteq \{\pm 1\}^{\mathcal{X}}$ be an extremal class. Then*

$$sc(\mathcal{E}) \geq \begin{cases} vc(\mathcal{E}) - 1 & \text{if } \mathcal{E} = \{\pm 1\}^{\mathcal{X}} \\ vc(\mathcal{E}) & \text{otherwise.} \end{cases}$$

Proof. If $\mathcal{E}$ is not the binary cube $\{\pm 1\}^{\mathcal{X}}$, let Q be a largest cube within the cubical complex $Q_{\mathcal{E}}$. If $\mathcal{E} = \{\pm 1\}^{\mathcal{X}}$, then instead take Q to be an arbitrary $(vc(\mathcal{E}) - 1)$ -dimensional cube in $Q_{\mathcal{E}}$. Let $\Pi \subseteq \Gamma_{\mathcal{E}}$ be the geometric realization of Q . By Proposition 39, $\dim(\Pi) = vc(\mathcal{E})$ if $\mathcal{E} \neq \{\pm 1\}^{\mathcal{X}}$ and $\dim(\Pi) = vc(\mathcal{E}) - 1$ if $\mathcal{E} = \{\pm 1\}^{\mathcal{X}}$.

Since $f_{\mathcal{E}}$, defined in (6), is an injection, any cover $\mathcal{B}$ of $\Delta_{\mathcal{E}}$ induces a cover $g(\mathcal{B}) = f_{\mathcal{E}}^{-1}(\mathcal{B}) \cap \Pi$ of Π with $\text{ord}(\mathcal{B}) \geq \text{ord}(g(\mathcal{B}))$. Furthermore, any refinement of $\mathcal{B}$ induces a refinement of $g(\mathcal{B})$. We will construct a cover $\mathcal{A}$ of $\Delta_{\mathcal{E}}$ which is refined by the zero-loss sets $\{B_h\}_{h \in \{\pm 1\}^{\mathcal{X}}}$ and such that $g(\mathcal{A})$ fulfills the conditions for Theorem 9. Thus, any refinement of $\mathcal{A}$ must have order at least d .

Let $\alpha < 1/|\mathcal{X}|$. We define the open cover $\mathcal{A}$ of $\Delta_{\mathcal{E}}$: $\mathcal{A} = \{B_h^{(\alpha)} : h \in \{\pm 1\}^{\mathcal{X}}\}$.

It is clear that $\mathcal{A}$ is refined by the zero-loss sets $\{B_h\}_{h \in \{\pm 1\}^{\mathcal{X}}}$.

Let y lie on some face of Π , so that without loss of generality $y_x = 1$ for some $x \in \mathcal{X}$, and let h be any hypothesis in $\{\pm 1\}^{\mathcal{X}}$ such that $h(x) = -1$. Set $y' = f_{\mathcal{E}}(y)$, so that $y'_x \geq \frac{1}{|\mathcal{X}|}$. Since any point $z \in B_h$ has $z_x \leq 0$, y' is too far from B_h to lie in $B_h^{(\alpha)}$.

$$\min_{z \in B_h} \|y' - z\|_1 \geq \min_{z \in B_h} |y'_x - z_x| \geq |y'_x| \geq \frac{1}{|\mathcal{X}|} > \alpha.$$

Thus, if y and z are two points on opposite faces of Π , neither can lie in the same set $g(B_h^{(\alpha)})$. This means that $g(\mathcal{A})$ satisfies the conditions of Theorem 9, and we are done. ◀

This lower bound requires $\mathcal{E}$ to be extremal, as we require the largest shattered set to form a solid cube in the cubical complex.

5.3 Upper bound for simplicial covering dimension of extremal classes

The upper bound utilizes a “topological equivalence” between the simplicial complex of realizable distributions $\Delta_{\mathcal{E}}$ and the cubical complex of strongly shattered sets $\Gamma_{\mathcal{E}}$. This equivalence is made precise by the following definition. A *retraction* from a topological space X to a subspace A is a continuous mapping $r: X \rightarrow A$ such that the restriction of r to A is the identity, i.e. $r(a) = a$ for all $a \in A$.

The following theorem is related to [12, Theorem 9], and the proof uses similar ideas, particularly the inductive construction of retractions. However, both the theorem and its proof include new elements and techniques that place the result in the framework of simplicial covering dimension. For instance, in [12], the retraction is mainly concerned with preserving antipodality, while in our case, the retraction is designed to produce a suitable open cover of $\Delta_{\mathcal{E}}$, which requires subtle analysis.

► **Theorem 41.** *Let $\mathcal{E} \neq \{\pm 1\}^{\mathcal{X}}$ be an extremal class, and let $\epsilon_0 > 0$. Then there is a retraction f from $\Delta_{\mathcal{E}}$ to $\tilde{\Gamma}_{\mathcal{E}}$ and an open cover $\{U_g\}_{g \in \mathcal{E}}$ of $\tilde{\Gamma}_{\mathcal{E}}$ such that the collection $\{f^{-1}(U_g \cap \tilde{\Gamma}_{\mathcal{E}})\}_{g \in \mathcal{E}}$ is an open cover of $\Delta_{\mathcal{E}}$ of order $vc(\mathcal{E})$ and $f^{-1}(U_g \cap \tilde{\Gamma}_{\mathcal{E}}) \subseteq \sigma_g^{(\epsilon_0)}$ for all $g \in \mathcal{E}$, where σ_g is the maximal simplex in $\Delta_{\mathcal{E}}$ corresponding to the concept g .*

For the proof of Theorem 41, we direct the reader to the full version of the paper [5].

The theorem provides an upper bound on the simplicial covering dimension of $\mathcal{E}$ that is needed to complete the proof of our main result. Recall that $\text{LR}_{\mathcal{H}}(\mathcal{C})$ denotes the list replicability number of $\mathcal{C}$, with outputs in $\mathcal{H}$. Similarly, we have $\text{SC}_{\mathcal{H}}(\mathcal{C}) := \text{SC}_{\Delta_{\mathcal{H}}}(\Delta_{\mathcal{C}})$.

► **Theorem 42** (General form of Theorem B). *Let $\mathcal{E} \subseteq \{\pm 1\}^{\mathcal{X}}$ be an extremal class. Then for every $\mathcal{E} \subseteq \mathcal{H} \subseteq \{\pm 1\}^{\mathcal{X}}$, we have*

$$\text{SC}_{\mathcal{H}}(\mathcal{E}) = \begin{cases} \text{VC}(\mathcal{E}) - 1 & \text{if } \mathcal{E} = \{\pm 1\}^{\mathcal{X}} \\ \text{VC}(\mathcal{E}) & \text{otherwise.} \end{cases}$$

Proof. Note that $\text{SC}(\mathcal{E}) \leq \text{SC}_{\mathcal{H}}(\mathcal{E}) \leq \text{SC}_{\mathcal{E}}(\mathcal{E})$ for any $\mathcal{H} \supseteq \mathcal{E}$. Hence, to determine $\text{SC}_{\mathcal{H}}(\mathcal{E})$, we bound $\text{SC}(\mathcal{E})$ from below and $\text{SC}_{\mathcal{E}}(\mathcal{E})$ from above. Immediately by Lemma 40, we have the lower bounds

$$\text{SC}(\mathcal{E}) \geq \begin{cases} \text{VC}(\mathcal{E}) - 1 & \text{if } \mathcal{E} = \{\pm 1\}^{\mathcal{X}} \\ \text{VC}(\mathcal{E}) & \text{otherwise.} \end{cases}$$

As for the upper bounds, consider first the case where $\mathcal{E} = \{\pm 1\}^{\mathcal{X}}$. Hence, $\text{SC}_{\mathcal{E}}(\mathcal{E}) = \text{SC}(\mathcal{E})$. Recall that

$$\text{SC}(\Delta_{\mathcal{E}}) = \sup \{ \text{SC}(\mathcal{A}) : \mathcal{A} \text{ is a finite open cover of } \Delta_{\mathcal{E}} \},$$

whereas $\text{SC}(\mathcal{E})$ is defined by taking the supremum over covers $\mathcal{A}$ that are refined by particular zero-loss sets. It follows that

$$\text{SC}(\mathcal{E}) \leq \text{SC}(\Delta_{\mathcal{E}}).$$

Since $\mathcal{E} = \{\pm 1\}^{\mathcal{X}}$, the space $\Delta_{\mathcal{E}}$ is the full cross polytope in $\mathbb{R}^{\mathcal{X}}$, which has topological dimension $|\mathcal{X}| - 1$. Furthermore, $\text{VC}(\mathcal{E}) = |\mathcal{X}|$ because the binary cube shatters its entire domain. Together, these imply that

$$\text{SC}(\mathcal{E}) \leq \text{SC}(\Delta_{\mathcal{E}}) = |\mathcal{X}| - 1 = \text{VC}(\mathcal{E}) - 1.$$

For the second case, let $\mathcal{E} \neq \{\pm 1\}^{\mathcal{X}}$. Let $\mathcal{A}$ be any finite open cover of $\Delta_{\mathcal{E}}$ such that $\mathcal{A} \prec \{\sigma \in \Delta_{\mathcal{E}}\}$. By Lemma 34 there exists an $\alpha > 0$ such that $\mathcal{A} \prec \{\sigma^{(\alpha)}\}_{\sigma \in \Delta_{\mathcal{E}}}$. This implies $\mathcal{A} \prec \{\sigma_g^{(\alpha)} : g \in \mathcal{E}\}$ where $\{\sigma_g : g \in \mathcal{E}\}$ are the maximal simplices of $\Delta_{\mathcal{E}}$. Pick $\epsilon_0 < \alpha$ and let f and $\{U_g\}_{g \in \mathcal{E}}$ be as in Theorem 41. Then $\{f^{-1}(U_g \cap \tilde{\Gamma}_{\mathcal{E}})\}_{g \in \mathcal{E}}$ is an open cover of $\Delta_{\mathcal{E}}$ of order $\text{VC}(\mathcal{E})$ that refines $\mathcal{A}$, whereby we conclude that $\text{SC}_{\mathcal{E}}(\mathcal{E}) \leq \text{VC}(\mathcal{E})$. ◀

As a consequence, we can exactly classify the proper ($\mathcal{H} = \mathcal{E}$) and improper list replicability number $\text{LR}_{\mathcal{H}}(\mathcal{E})$ of extremal concept classes:

► **Corollary 43.** *Let $\mathcal{E} \subseteq \{\pm 1\}^{\mathcal{X}}$ be an extremal class. Then for every $\mathcal{H} \supseteq \mathcal{E}$, we have*

$$\text{LR}_{\mathcal{H}}(\mathcal{E}) = \begin{cases} \text{VC}(\mathcal{E}) & \text{if } \mathcal{E} = \{\pm 1\}^{\mathcal{X}} \\ \text{VC}(\mathcal{E}) + 1 & \text{otherwise} \end{cases}.$$

Proof. The corollary follows directly by combining Theorem 42 and Theorem 38. ◀

References

- 1 Noga Alon, Mark Bun, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private and online learnability are equivalent. *J. ACM*, 69(4):Art. 28, 34, 2022. doi:10.1145/3526074.
- 2 Hans-Jürgen Bandelt, Victor Chepoi, Andreas Dress, and Jack Koolen. Combinatorics of lopsided sets. *European J. Combin.*, 27(5):669–689, 2006. doi:10.1016/j.ejc.2005.03.001.
- 3 Ari Blondal, Shan Gao, Hamed Hatami, and Pooya Hatami. Stability and list-replicability for agnostic learners, 2025. arXiv:2501.05333.
- 4 Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Borsuk-Ulam and replicable learning of large-margin halfspaces. *arXiv preprint arXiv:2503.15294*, 2025. doi:10.48550/arXiv.2503.15294.
- 5 Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Simplicial covering dimension of extremal concept classes. *arXiv preprint arXiv:2511.11819*, 2025. URL: <https://arxiv.org/abs/2511.11819>.
- 6 Mark Bun, Marco Gaboardi, Max Hopkins, Russell Impagliazzo, Rex Lei, Toniann Pitassi, Satchit Sivakumar, and Jessica Sorrell. Stability is stable: Connections between replicability, privacy, and adaptive generalization. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 520–527. ACM, 2023. doi:10.1145/3564246.3585246.
- 7 Mark Bun, Roi Livni, and Shay Moran. An equivalence between private classification and online prediction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402. IEEE, 2020. doi:10.1109/FOCS46700.2020.00044.
- 8 Jérémie Chalopin, Victor Chepoi, Shay Moran, and Manfred K Warmuth. Unlabeled sample compression schemes and corner peelings for ample and maximum classes. *Journal of Computer and System Sciences*, 127:1–28, 2022. doi:10.1016/j.jcss.2022.01.003.
- 9 Zachary Chase, Bogdan Chornomaz, Steve Hanneke, Shay Moran, and Amir Yehudayoff. Dual VC dimension obstructs sample compression by embeddings. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 923–946. PMLR, PMLR, 2024. URL: <https://proceedings.mlr.press/v247/chase24a.html>.
- 10 Zachary Chase, Bogdan Chornomaz, Shay Moran, and Amir Yehudayoff. Local Borsuk-Ulam, stability, and replicability. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, pages 1769–1780. Association for Computing Machinery, 2024. doi:10.1145/3618260.3649632.
- 11 Zachary Chase, Shay Moran, and Amir Yehudayoff. Stability and Replicability in Learning. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2430–2439. IEEE Computer Society, November 2023. doi:10.1109/FOCS57990.2023.00148.
- 12 Bogdan Chornomaz, Shay Moran, and Tom Waknine. Spherical dimension. In Nika Haghtalab and Ankur Moitra, editors, *The Thirty Eighth Annual Conference on Learning Theory, 30-4 July 2025, Lyon, France*, volume 291 of *Proceedings of Machine Learning Research*, pages 1259–1313. PMLR, 2025. URL: <https://proceedings.mlr.press/v291/chornomaz25a.html>.
- 13 Michel Coornaert. *Topological dimension and dynamical systems*. Universitext. Springer, Cham, 2015. Translated and revised from the 2005 French original.
- 14 Peter Dixon, A. Pavan, Jason Vander Woude, and N. V. Vinodchandran. List and certificate complexities in replicable learning. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, NeurIPS '23. Curran Associates Inc., 2023. URL: <http://papers.nips.cc/paper/2023/hash/61d0a96d4a73b626367310b3ad32579d-Abstract-Conference.html>.

- 15 Eric Eaton, Marcel Hushing, Michael Kearns, and Jessica Sorrell. Replicable reinforcement learning. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, NeurIPS '23. Curran Associates Inc., 2023. URL: <http://papers.nips.cc/paper/2023/hash/313829757739365201b5adb3a1cbd9bd-Abstract-Conference.html>.
- 16 Ryszard Engelking. *Dimension theory*, volume 19 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam-Oxford-New York; PWN—Polish Scientific Publishers, Warsaw, 1978. Translated from the Polish and revised by the author.
- 17 Hossein Esfandiari, Alkis Kalavasis, Amin Karbasi, Andreas Krause, Vahab Mirrokni, and Grigoris Velegkas. Replicable bandits. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL: <https://openreview.net/forum?id=gcD2UtCGMc2>.
- 18 Hossein Esfandiari, Amin Karbasi, Vahab Mirrokni, Grigoris Velegkas, and Felix Zhou. Replicable clustering. In *Advances in Neural Information Processing Systems*, volume 36, pages 39277–39320. Curran Associates, Inc., 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/7bc3fe234454107149fa9d44faacaa64-Paper-Conference.pdf.
- 19 Daniel S Farley. Finiteness and CAT(0) properties of diagram groups. *Topology*, 42(5):1065–1082, 2003.
- 20 Witold Hurewicz and Henry Wallman. *Dimension Theory*, volume vol. 4 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1941.
- 21 Alkis Kalavasis, Amin Karbasi, Kasper Green Larsen, Grigoris Velegkas, and Felix Zhou. Replicable learning of large-margin halfspaces. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, ICML'24. OpenReview.net, 2024. URL: <https://openreview.net/forum?id=CKCzfU9YKE>.
- 22 Alkis Kalavasis, Amin Karbasi, Shay Moran, and Grigoris Velegkas. Statistical indistinguishability of learning algorithms. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 15586–15622. PMLR, 2023. URL: <https://proceedings.mlr.press/v202/kalavasis23a.html>.
- 23 Amin Karbasi, Grigoris Velegkas, Lin Yang, and Felix Zhou. Replicability in reinforcement learning. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, volume 36, pages 74702–74735, 2023. URL: <http://papers.nips.cc/paper/2023/hash/ec4d2e436794d1bf55ca83f5ebb31887-Abstract-Conference.html>.
- 24 Jim Lawrence. Lopsided sets and orthant-intersection by convex sets. *Pacific J. Math.*, 104(1):155–173, 1983.
- 25 Maryanthe Malliaris and Shay Moran. The unstable formula theorem revisited via algorithms. *Ann. Pure Appl. Log.*, 176(10):103633, 2025. doi:10.1016/j.apal.2025.103633.
- 26 Jiří Matoušek. *Using the Borsuk-Ulam theorem*. Universitext. Springer-Verlag, Berlin, 2003. Lectures on topological methods in combinatorics and geometry, Written in cooperation with Anders Björner and Günter M. Ziegler.
- 27 Shay Moran, Hilla Scheffer, and Jonathan Shafer. The bayesian stability zoo. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, volume 36, pages 61725–61746, 2023. URL: <http://papers.nips.cc/paper/2023/hash/c2586b71fd150fb56952e253a9c551cc-Abstract-Conference.html>.
- 28 Shay Moran and Manfred K Warmuth. Labeled compression schemes for extremal classes. In *International Conference on Algorithmic Learning Theory*, pages 34–49. Springer, 2016. doi:10.1007/978-3-319-46379-7_3.

- 29 J.R. Munkres. *Topology*. Featured Titles for Topology. Prentice Hall, Incorporated, 2000.
- 30 Jason Vander Woude, Peter Dixon, Aduri Pavan, Jamie Radcliffe, and N. V. Vinodchandran. Replicability in learning: Geometric partitions and kkm-sperner lemma. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, volume 37, pages 78996–79028, 2024. URL: <http://papers.nips.cc/paper/2024/hash/8ff87c96935244b63503f542472462b3-Abstract-Conference.html>.