# Sign-Rank, Index, and List Replicability: Connections and Separations

Ari Blondal* Hamed Hatami* Pooya Hatami † Chavdar Lalov† Sivan Tretiak†

April 2026

**Abstract**

In learning theory, the sign rank of a binary concept class captures the smallest dimension in which it can be represented by points and halfspaces. Despite tremendous interest, lower bounds on sign rank are notoriously difficult to come by. Two recent approaches to the problem establish lower bounds on sign rank by measures that are easier to analyze: the $\mathbb{Z}_2$-index and the list replicability number.

We order these measures, showing that the $\mathbb{Z}_2$-index is upper-bounded by a linear function of the list replicability number. As a main consequence, we obtain a strong separation between sign rank and $\mathbb{Z}_2$-index, thereby resolving a question of Frick, Hosseini, and Vasileuski.

This motivates a thorough study of list replicability, the stronger of the two lower-bounding measures. We establish upper bounds on the list replicability number by two combinatorial measures: height and minimum star number. We also prove a fundamental composition result, showing that the product of two concept classes has list replicability number bounded by the sum of the list replicability numbers of the two classes.

## Contents

---

*McGill University, `ari.blondal@mail.mcgill.ca`, `hamed.hatami@mcgill.ca`. Ari Blondal and Hamed Hatami are supported by NSERC grants.

†Ohio State University, {`hatami.2, lalov.1, tretiak.2`}`@osu.edu`

# 1 Introduction

In this paper, we study sign matrices and finite concept classes via three complementary notions of complexity: a geometric notion (sign-rank), a topological notion ($\mathbb{Z}_2$-index), and an algorithmic learning-theoretic notion (list replicability). We show that list replicability number lies between $\mathbb{Z}_2$-index and sign-rank, and use this to obtain a strong separation between $\mathbb{Z}_2$-index and sign-rank. We also show that several existing $\mathbb{Z}_2$-index upper bounds in fact hold at the stronger level for list replicability, and develop new upper bounds, lower bounds, and closure properties for list replicability itself.

**Sign-rank.** Sign-rank is a fundamental and well-studied parameter in learning theory which captures the smallest dimension in which a binary classification problem admits a linear representation. Formally, the *sign-rank* of a matrix $A$ with entries in $\{+1, -1\}$ is the minimum rank of a real matrix $B$ with $\operatorname{sign}(B_{i,j}) = A_{i,j}$ for all entries $i, j$.

Given a binary concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ over a *finite* domain $\mathcal{X}$, the *sign-rank* $\operatorname{signrk}(\mathcal{C})$ is the sign-rank of the $|\mathcal{C}| \times |\mathcal{X}|$ matrix $A$ defined by $A_{c,x} = c(x)$. Equivalently, it is the smallest $d$ for which there exist embeddings $\{u_c \in \mathbb{R}^d\}_{c\in\mathcal{C}}$ and $\{v_x \in \mathbb{R}^d\}_{x\in\mathcal{X}}$ satisfying

$$c(x) = \operatorname{sign}\langle u_c, v_x\rangle \quad \text{for all } c \in \mathcal{C},\ x \in \mathcal{X}. \tag{1}$$

Classical results on sign patterns of polynomials imply that most $N \times N$ sign matrices have sign-rank $\Omega(N)$ [AMY16, Lemma 22]. Nevertheless, proving even super-constant lower bounds for well-structured matrices that lack pseudorandom properties has remained elusive [HHP$^+$22]. A promising new approach was recently proposed by Frick, Hosseini, and Vasileuski [FHV26], who developed a topological framework for sign-rank lower bounds based on the $\mathbb{Z}_2$-index of a space associated with the sign matrix. We now describe this framework.

**The $\mathbb{Z}_2$-index of a concept class.** A distribution $\mu$ over $\mathcal{X} \times \{\pm 1\}$ is *realizable* by a concept $c \in \mathcal{C}$ if every pair $(x, b)$ in its support satisfies $b = c(x)$. Given $\mathcal{C}$, let $\mathcal{C}^{\pm} := \mathcal{C} \cup \{-c : c \in \mathcal{C}\}$ denote its *antipodal completion*, and let $\Delta_{\mathcal{C}^{\pm}}$ be the set of all distributions realizable by some $c \in \mathcal{C}^{\pm}$. Equipped with the total variation metric, $\Delta_{\mathcal{C}^{\pm}}$ becomes a topological space.

Recall that a $\mathbb{Z}_2$*-action* on a topological space $X$ is a continuous involution $\tau\colon X \to X$ (i.e., $\tau \circ \tau = \mathrm{id}$); it is *free* if $\tau$ has no fixed points. The two examples relevant to us are:

- The space $\Delta_{\mathcal{C}^{\pm}}$ with the label-negation action $\mu \mapsto -\mu$, where $(-\mu)(x, b) := \mu(x, -b)$.
- The unit sphere $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ with the antipodal action $x \mapsto -x$.

Let $X$ be a topological space with a free $\mathbb{Z}_2$-action $\tau$. A *continuous* map $\Phi : X \to \mathbb{S}^d$ is $\mathbb{Z}_2$*-equivariant* if it preserves the $\mathbb{Z}_2$-action: $\Phi(\tau(x)) = -\Phi(x)$ for all $x \in X$. The $\mathbb{Z}_2$*-index* of $X$ is

$$\operatorname{Ind}_{\mathbb{Z}_2}(X) \;:=\; \min\bigl\{d : \text{there exists a } \mathbb{Z}_2\text{-equivariant map } X \to \mathbb{S}^d\bigr\}.$$

In particular, $\operatorname{Ind}_{\mathbb{Z}_2}(\mathbb{S}^n) = n$; this is in fact equivalent to the Borsuk–Ulam theorem.

Frick et al. [FHV26] defined the $\mathbb{Z}_2$*-index of a concept class* $\mathcal{C}$ as $\operatorname{Ind}_{\mathbb{Z}_2}(\mathcal{C}) := \operatorname{Ind}_{\mathbb{Z}_2}(\Delta_{\mathcal{C}^{\pm}})$, where $\Delta_{\mathcal{C}^{\pm}}$ is equipped with the label-negation $\mathbb{Z}_2$-action described above.

Dual to the $\mathbb{Z}_2$-index is the $\mathbb{Z}_2$-*coindex* of a concept class $\mathcal{C}$, which is the dual invariant obtained by reversing the direction of the equivariant map. Namely, $\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C})$ is the largest $d$ for which there is a $\mathbb{Z}_2$-equivariant map $\mathbb{S}^d \to \Delta_{\mathcal{C}^\pm}$. The $\mathbb{Z}_2$-coindex was previously studied under the name *spherical dimension* in [CMW25]. Note that, by the Borsuk–Ulam theorem,

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \le \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}).$$

**Connection to sign-rank.** As observed in [FHV26], the sign-rank decomposition (1) naturally produces a $\mathbb{Z}_2$-equivariant map from $\Delta_{\mathcal{C}^\pm}$ to $\mathbb{S}^{d-1}$: Suppose $\mathrm{signrk}(\mathcal{C}) \le d$, with embeddings $\{u_c\}_{c\in\mathcal{C}}$ and $\{v_x\}_{x\in\mathcal{X}}$ as in (1). The antipodal completion does not increase sign-rank, since we may set $u_{-c} := -u_c$. The domain embedding $x \mapsto v_x$ extends by linearity to a map $\psi\colon \Delta_{\mathcal{C}^\pm} \to \mathbb{R}^d$, defined as $\psi(\mu) := \mathbb{E}_{(x,b)\sim\mu}[b\, v_x]$. If $\mu$ is realizable by $c \in \mathcal{C}^\pm$, then every $(x,b) \sim \mu$ satisfies $\langle u_c, bv_x\rangle = |\langle u_c, v_x\rangle| > 0$ with probability 1, and therefore, $\psi(\mu) \neq 0$. Hence, we can normalize $\Phi(\mu) := \psi(\mu)/\|\psi(\mu)\|$ to obtain a continuous map $\Phi\colon \Delta_{\mathcal{C}^\pm} \to \mathbb{S}^{d-1}$. Since $\psi$ is linear and label-negation reverses the signed weights, $\psi(-\mu) = -\psi(\mu)$, and hence $\Phi(-\mu) = -\Phi(\mu)$, i.e., $\Phi$ is $\mathbb{Z}_2$-equivariant. Therefore,

$$\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}) \;\le\; \mathrm{signrk}(\mathcal{C}) - 1. \tag{2}$$

In light of the construction above, the $\mathbb{Z}_2$-index can be viewed as a topological relaxation of sign-rank: while sign-rank requires the map $\psi$ from realizable distributions to $\mathbb{R}^d \setminus \{0\}$ to be *linear*, the $\mathbb{Z}_2$-index asks only for a *continuous* $\mathbb{Z}_2$-equivariant map with no linearity requirement.

**List replicability.** There are various formalizations of replicability in learning theory, most of which build off of the well-established notion of probably approximately correct (PAC) learning. An $(\epsilon,\delta)$-*PAC learning* algorithm for a concept class $\mathcal{C}$ receives $n := n_{\mathcal{C}}(\epsilon,\delta)$ i.i.d. examples from an unknown realizable distribution $\mu$ and, with probability at least $1-\delta$, produces a hypothesis $h\colon \mathcal{X} \to \{\pm 1\}$ with *population loss*

$$\mathrm{loss}_\mu(h) \;:=\; \mathbb{P}_{(x,b)\sim\mu}[h(x) \neq b] \;\le\; \epsilon.$$

Given $L \in \mathbb{N}$, such an algorithm is *$L$-list-replicable* if for every realizable distribution $\mu \in \Delta_{\mathcal{C}}$, the output hypothesis belongs to a small list $\mathcal{L}_\mu = \{h_1,\ldots,h_L\}$ with probability at least $1-\delta$. The list may depend on $\mu$, but its size must not. The *list replicability number* $\textsc{lr}(\mathcal{C})$, introduced by [CMY23, DPVWV23], is the smallest $L$ for which an $L$-list-replicable $(\epsilon,\delta)$-PAC learner exists for all $\epsilon,\delta > 0$.

## 1.1 Our results

### 1.1.1 List replicability bridges $\mathbb{Z}_2$-index and sign-rank.

Our main result connects the $\mathbb{Z}_2$-index to list replicability.

**Theorem 1.1** (Index is controlled by list replicability)**.** *For every concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ over a finite domain $\mathcal{X}$,*

$$\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}) \;\le\; 2\,\textsc{lr}(\mathcal{C}) - 1.$$

It is shown in [BHH$^+$26a] that, over finite domains, $\textsc{lr}(\mathcal{C}) \le \mathrm{signrk}(\mathcal{C})$. Combining this with Theorem 1.1 and the Borsuk–Ulam inequality $\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \le \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$ gives the following chain of inequalities:

$$\frac{1}{2}\,\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \;\le\; \frac{1}{2}\,\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}) \;<\; \textsc{lr}(\mathcal{C}) \;\le\; \mathrm{signrk}(\mathcal{C}). \tag{3}$$

Frick et al. [FHV26, Question 7] asked whether there exists a function $f$ such that

$$\mathrm{signrk}(\mathcal{C}) \le f(\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}))$$

for every finite concept class $\mathcal{C}$. We use Theorem 1.1 in conjunction with existing results about sign-rank to show that no such bound exists.

**Theorem 1.2** (Sign-rank is not bounded by index)**.** *There exists a family of $N \times N$ sign matrices whose $\mathbb{Z}_2$-index is at most 5, while their sign-rank grows polynomially in $N$.*

The matrices are incidence matrices of finite projective planes. Their sign-rank is known to be polynomially large by the lower bound of Alon, Moran, and Yehudayoff [AMY16]. We show in Theorem 3.2 that their list replicability number is at most 3. Theorem 1.1 then gives $\mathrm{Ind}_{\mathbb{Z}_2} \le 5$.

#### 1.1.2 Extremal classes.

Recall that a set $S \subseteq \mathcal{X}$ is *shattered* by $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ if $\mathcal{C}|_S = \{\pm 1\}^S$, and the *VC dimension* of $\mathcal{C}$ is the size of the largest set that it shatters. An immediate topological ramification of shattering a set $S$ of size $d$ is that the set of distributions in $\Delta_{\mathcal{C}^{\pm}}$ supported on $S$ is homeomorphic to $\mathbb{S}^{d-1}$ (via the identification of distributions with $\ell_1$-unit vectors in $\mathbb{R}^{\mathcal{X}}$), and therefore $\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \ge d-1$. Consequently, every concept class satisfies

$$\mathrm{VC}(\mathcal{C}) - 1 \le \mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \le \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}).$$

A set $S \subseteq \mathcal{X}$ is *strongly shattered* by $\mathcal{E} \subseteq \{\pm 1\}^{\mathcal{X}}$ if there is a fixed labelling $a \in \{\pm 1\}^{\mathcal{X} \setminus S}$ with

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

A class $\mathcal{E}$ is *extremal* if every shattered set is also strongly shattered. Many natural concept classes are extremal or admit natural extremal extensions; see [CCMW22, Section 3.2] for a list of examples.

Extremal classes are among the few cases where list replicability is well understood. Blondal et al. [BHH$^+$26b] showed that for an extremal class $\mathcal{E} \subseteq \{\pm 1\}^{\mathcal{X}}$ over a finite domain,

$$\mathrm{LR}(\mathcal{E}) = \begin{cases} \mathrm{VC}(\mathcal{E}) & \text{if } \mathcal{E} = \{\pm 1\}^{\mathcal{X}}, \\ \mathrm{VC}(\mathcal{E}) + 1 & \text{otherwise.} \end{cases} \tag{4}$$

Combined with Theorem 1.1 and the general bound $\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \ge \mathrm{VC}(\mathcal{C}) - 1$, this pins down all three parameters up to a factor of two: for every extremal class $\mathcal{E}$,

$$\mathrm{VC}(\mathcal{E}) - 1 \ \le\ \mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{E}) \ \le\ \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{E}) \ \le\ 2\,\mathrm{LR}(\mathcal{E}) - 1 \le 2\,\mathrm{VC}(\mathcal{E}) + 1. \tag{5}$$

#### 1.1.3 Height and eluder dimension.

Invariants which capture a notion of height for topological spaces (such at the Stiefel–Whitney height) are standard tools for bounding the $\mathbb{Z}_2$-index/coindex. In the same vein, Frick et al. [FHV26] introduced a combinatorial notion of height as a way to upper-bound the $\mathbb{Z}_2$-index of a sign matrix. Using this approach, they showed that random $N \times N$ sign matrices and the Hadamard matrix have $\mathbb{Z}_2$-index $O(\log N)$.

We instead use a version of their height definition adapted to partial concept classes $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ (both versions are equivalent up to a constant factor). For two partial concepts $h_1, h_2 \in \{\pm 1, \star\}^{\mathcal{X}}$, define their *junction* by

$$(h_1 \cap h_2)(x) = \begin{cases} h_1(x) & \text{if } h_1(x) = h_2(x), \\ \star & \text{otherwise.} \end{cases}$$

We write $g \preceq h$ if $g = g \cap h$. Write $g \prec h$ if $g \preceq h$ and $g \neq h$. Define the *junction closure* of a partial concept class $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ as,

$$\mathcal{J}(\mathcal{C}) = \left\{ \bigcap_{c \in S} c : \varnothing \neq S \subseteq \mathcal{C} \right\}$$

The *height* of $\mathcal{C}$, denoted $\mathrm{H}(\mathcal{C})$, is the length $m$ of the longest strict chain $h_1 \prec h_2 \prec \cdots \prec h_m$ in $\mathcal{J}(\mathcal{C})$.[1]

---

[1] Given $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$, let $\mathcal{S}_{\mathcal{C}} := \{c^+, c^- : c \in \mathcal{C}\}$, where $c^+ := \{x : c(x) = 1\}$ and $c^- := \{x : c(x) = -1\}$, and let $\mathcal{I}_{\mathcal{C}}$ be the set of all finite non-empty intersections of elements of $\mathcal{S}_{\mathcal{C}}$. Frick et al. [FHV26] define the height $h_{\mathrm{FHV}}(\mathcal{C})$ as the length of the longest strict chain in $\mathcal{I}_{\mathcal{C}}$. Up to the convention of whether the empty set is included in $\mathcal{I}_{\mathcal{C}}$,

$$h_{\mathrm{FHV}}(\mathcal{C}) \ \le\ \mathrm{H}(\mathcal{C}^{\pm}) \ \le\ 2\,h_{\mathrm{FHV}}(\mathcal{C}) - 1.$$

In view of the aforementioned bound of Frick et al. on the $\mathbb{Z}_2$-index by height, and our Theorem 1.1 bounding the $\mathbb{Z}_2$-index by LR, it is natural to ask how the two upper-bounding quantities, height and LR, compare. Our next result clarifies this by showing that height also upper-bounds LR.

**Theorem 1.3** (Height controls list replicability). *For every finite partial concept class $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$,*

$$\textsc{lr}(\mathcal{C}) \leq \mathrm{H}(\mathcal{C}).$$

The proof of Theorem 1.3, as found in Section 4, is algorithmic. We construct a learner whose output hypothesis is likely to lie on a single chain in the junction closure of the class. Since every such chain has length at most $\mathrm{H}(\mathcal{C})$, the learner is $\mathrm{H}(\mathcal{C})$-list-replicable.

The combinatorial notion of height given by $\mathrm{H}(\mathcal{C})$ has a useful equivalence to a learning-theoretic parameter known as eluder dimension. The eluder dimension was introduced by Russo and Van Roy [RVR13] as a sequential notion of independence for function classes. Informally, it is the maximum length of a sequence of points such that, at each step, the label at the next point is not determined by the labels on the preceding points.

Fix a finite partial concept class $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ and a base concept $c' \in \mathcal{C}$. Write $\mathrm{supp}(c') := \{x \in \mathcal{X} : c'(x) \neq \star\}$. The *eluder dimension* of $\mathcal{C}$ *relative* to $c'$, denoted $\mathrm{Edim}(\mathcal{C}, c')$, is the largest $m$ for which there exist points $x_1, \dots, x_m \in \mathrm{supp}(c')$ and concepts $c_1, \dots, c_m \in \mathcal{C}$ such that, for every $i \in [m]$,

$$c_i(x_i) \neq c'(x_i), \ (c_i(x_i) = \star \text{ allowed})$$

while at every earlier position $j < i$,

$$c_i(x_j) = c'(x_j).$$

The *eluder dimension* of $\mathcal{C}$ is $\mathrm{Edim}(C) \coloneqq \sup_{c' \in \mathcal{C}} \mathrm{Edim}(\mathcal{C}, c')$.

**Proposition 1.4** (Height and eluder dimension are equivalent). *For any finite partial concept class $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$,*

$$\mathrm{H}(\mathcal{C}) = \mathrm{Edim}(\mathcal{C}) + 1.$$

We prove a slightly stronger form of this in Proposition 5.3. The result allows us to translate the bound $\textsc{lr}(\mathcal{C}) \leq \mathrm{H}(\mathcal{C})$ into the existing literature surrounding eluder dimension, as will be seen in the following section.

### 1.1.4 The star number.

The *star number* is a classical combinatorial parameter from learning theory that measures the size of the largest local star around a target labelling [HY15].

**Definition 1.5** (Star number). *For a class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$ and a center $h \in \{\pm 1\}^{\mathcal{X}}$, let $s_h(\mathcal{C})$ denote the largest $m$ for which there exist distinct points $x_1, \dots, x_m \in \mathcal{X}$ and concepts $c_0, c_1, \dots, c_m \in \mathcal{C}$ such that*

$$c_i(x_j) = h(x_j) \qquad \Longleftrightarrow \qquad i \neq j$$

*for every $i \in \{0, \dots, m\}$ and $j \in [m]$.*

Thus $c_0$ agrees with $h$ on all of $x_1, \dots, x_m$, while each other $c_i$ disagrees with $h$ exactly at $x_i$. We consider both extremes over the choice of center:

$$s_{\min}(\mathcal{C}) \coloneqq \min_{h \in \{\pm 1\}^{\mathcal{X}}} s_h(\mathcal{C}) \qquad \text{and} \qquad s_{\max}(\mathcal{C}) \coloneqq \max_{h \in \{\pm 1\}^{\mathcal{X}}} s_h(\mathcal{C}).$$

Hanneke [Han24] showed that the star number characterizes intersection-closed structure: the minimum star number $s_{\min}(\mathcal{C})$ of a concept class $\mathcal{C}$ equals the least possible VC dimension of a *generalized intersection-closed* class $\mathcal{H}$ containing $\mathcal{C}$.

**Definition 1.6** (Generalized intersection-closed)**.** *We say that $\mathcal{H} \subseteq \{\pm 1\}^{\mathcal{X}}$ is* generalized intersection-closed *if there exists a center $h_\star \in \{\pm 1\}^{\mathcal{X}}$ such that, for every nonempty finite $\mathcal{A} \subseteq \mathcal{H}$, the concept*

$$\left(\bigwedge_{h_\star} \mathcal{A}\right)(x) := \begin{cases} h_\star(x), & \text{if } c(x) = h_\star(x) \text{ for every } c \in \mathcal{A}, \\ -h_\star(x), & \text{otherwise} \end{cases}$$

*also belongs to $\mathcal{H}$.*

This notion generalizes the classical definition of an intersection-closed class, which corresponds to the case where $h_\star$ is the all-1 hypothesis.

We show that all finite concept classes can be list-replicably learned with list size $O(s_{\min})$, by combining Hanneke's characterization with known embeddings into extremal concept classes.

**Theorem 1.7.** *For every finite binary concept class $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$,*

$$\textsc{lr}(\mathcal{C}) \leq 11 s_{\min}(\mathcal{C}) + 1.$$

*Proof.* By Hanneke's characterization of intersection-closed classes, $\mathcal{C}$ embeds into a generalized intersection-closed class $\mathcal{H}$ satisfying

$$\textsc{vc}(\mathcal{H}) = s_{\min}(\mathcal{C});$$

see [Han24, Theorem 19 and Remark 24].

Rubinstein and Rubinstein show that every generalized intersection-closed class of VC dimension $d$ embeds into an extremal class of VC dimension at most $11d$; see [RR22, Theorems 4.1 and 4.4]. Applying this result to $\mathcal{H}$, we obtain an extremal class $\mathcal{E} \subseteq \{\pm 1\}^{\mathcal{X}}$ such that $\mathcal{C} \subseteq \mathcal{H} \subseteq \mathcal{E}$ and

$$\textsc{vc}(\mathcal{E}) \leq 11\, \textsc{vc}(\mathcal{H}) = 11 s_{\min}(\mathcal{C}).$$

This combined embedding statement is also recorded explicitly in [Han24, Corollary 27].

Finally, by Equation (4), every finite extremal class satisfies

$$\textsc{lr}(\mathcal{E}) \leq \textsc{vc}(\mathcal{E}) + 1.$$

By monotonicity of the list replicability under taking subclasses, we conclude

$$\textsc{lr}(\mathcal{C}) \leq \textsc{lr}(\mathcal{E}) \leq \textsc{vc}(\mathcal{E}) + 1 \leq 11 s_{\min}(\mathcal{C}) + 1. \qquad \square$$

**Remark 1.8.** *It is easy to see that the maximum star number is upper bounded by eluder dimension and height*[2]*. In particular, for every concept class $\mathcal{C}$, we have*

$$s_{\min}(\mathcal{C}) \leq s_{\max}(\mathcal{C}) \leq \mathrm{Edim}(\mathcal{C}) = \mathrm{H}(\mathcal{C}) - 1. \tag{6}$$

*Thus, for total classes, Theorem 1.7 is a strengthening, ignoring constant factors, of the height/eluder bound $\textsc{lr}(\mathcal{C}) \leq \mathrm{H}(\mathcal{C})$ of Theorem 1.3. The height bound, however, applies in the more general setting of partial concept classes.*

**Star number and sign-rank** Li, Kamath, Foster, and Srebro [LKFS22] asked whether there exist concept classes with bounded $s_{\max}$ but arbitrarily large sign-rank. They also conjectured the stronger statement that there exist concept classes with constant Edim and arbitrarily large sign rank. We disprove both of these conjectures: sign-rank is bounded by a function of $s_{\max}$ alone.

---

[2] In fact, Li, Kamath, Foster, and Srebro [LKFS22] showed that the eluder dimension is characterized by the maximum of star number and *threshold dimension*: $\max\{s_{\max}(\mathcal{C}), \mathrm{Tdim}(\mathcal{C})\} \leq \mathrm{Edim}(\mathcal{C}) \leq 4^{\max\{s_{\max}(\mathcal{C}), \mathrm{Tdim}(\mathcal{C})\}}$.

**Theorem 1.9.** *For every sign matrix $A$ with $s_{\max}(A) \le d$,*

$$\mathrm{signrk}(A) \;\le\; 2\,f(d)^2 \;\le\; 2\,(2d^2)^{2(2d^2)^{4d}},$$

*where $f$ is the function from [Atm22, Theorem 2.1].*

*Proof.* Consider the bipartite graph $G_{\mathcal{C}}$ with parts $\mathcal{C}$ and $\mathcal{X}$, where $c \in \mathcal{C}$ is adjacent to $x \in \mathcal{X}$ if and only if $c(x) = -1$. Let $d := s_{\max}(\mathcal{C})$. Then since $s_{\mathbf{1}}(\mathcal{C}) \le d$ and $s_{-\mathbf{1}}(\mathcal{C}) \le d$, where $\mathbf{1}$ and $-\mathbf{1}$ denote the all-ones and all-minus-ones hypotheses[3], neither $G_{\mathcal{C}}$ nor its bipartite complement contains an induced matching of size $d$. Atminas [Atm22, Theorem 2.1] showed that there is a function $f(d) \le (2d^2)^{(2d^2)^{4d}}$ such that whenever a bipartite graph $G = (A \cup B, E)$ and its bipartite complement both omit induced matchings of size $d$, the parts admit partitions $A = A_1 \cup \cdots \cup A_u$ and $B = B_1 \cup \cdots \cup B_u$ with $u \le f(d)$ such that every induced subgraph $G[A_i, B_j]$ is $2K_2$-free (as induced subgraph).

We combine this with the following standard fact about $2K_2$-free bipartite graphs.

**Proposition 1.10** ([MP95]; see [HK07, Theorem 6.3])**.** *If a bipartite graph $G = (A \cup B, E)$ contains no induced matching with two edges, then there is an ordering $B = \{b_1, \ldots, b_m\}$ such that $ab_i \in E$ implies $ab_j \in E$ for all $j \ge i$.*

The adjacency pattern of Proposition 1.10 is realized by points and thresholds on a line, so each block $G[A_i, B_j]$ in Atminas's partition, viewed as a sign matrix, has sign-rank at most 2. The matrix $A$ is partitioned into a $u \times u$ grid of such blocks with $u \le f(d)$, and since sign-rank is subadditive under both horizontal and vertical concatenation, we conclude $\mathrm{signrk}(A) \le 2u^2 \le 2f(d)^2$. ☐

### 1.1.5 Separation of coindex and list replicability for partial classes.

All three parameters, $\mathrm{Ind}_{\mathbb{Z}_2}$, $\mathrm{coInd}_{\mathbb{Z}_2}$, and LR, extend naturally to partial concept classes $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$. A distribution $\mu$ over $\mathcal{X} \times \{\pm 1\}$ is realizable by such a $c$ if it is supported on pairs $(x, c(x))$ with $c(x) \ne \star$.

One advantage of allowing partial concept classes is that they are flexible enough to encode known classic topological examples. In particular, building on an observation of Frick et al. [FHV26], standard free $\mathbb{Z}_2$-spaces with bounded $\mathbb{Z}_2$-coindex but unbounded $\mathbb{Z}_2$-index can be realized by partial concept classes. Combining this with Theorem 1.1, we obtain a dimension-free separation between $\mathrm{coInd}_{\mathbb{Z}_2}$ and LR.

**Corollary 1.11.** *There exist partial concept classes $\mathcal{C} \subseteq \{\pm 1, \star\}^n$ with*

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) = O(1) \qquad \textit{and} \qquad \mathrm{LR}(\mathcal{C}) = \omega(1).$$

We outline the argument below, leaving the details for Section 6. The argument builds on an observation of Frick, Hosseini, and Vasileuski that every finite free $\mathbb{Z}_2$-simplicial complex arises as $\Delta_{\mathcal{C}^{\pm}}$ for some partial concept class $\mathcal{C}$ [FHV26]. Moreover, by Illman's equivariant triangulation theorem [Ill78], every compact smooth free $\mathbb{Z}_2$-manifold admits a finite $\mathbb{Z}_2$-equivariant triangulation. This gives a recipe for importing standard examples from equivariant topology to partial concept classes. Applying this reduction to the spaces $\mathbb{RP}^{2N-1}$, with the free $\mathbb{Z}_2$-action induced by multiplication by $i$ on $\mathbb{C}^N$, one obtains partial concept classes with bounded $\mathrm{coInd}_{\mathbb{Z}_2}$ and unbounded $\mathrm{Ind}_{\mathbb{Z}_2}$ [FHV26]. Combined with Theorem 1.1, this gives a dimension-free separation between $\mathrm{coInd}_{\mathbb{Z}_2}$ and LR.

### 1.1.6 List replicability under joins and concatenations.

We next record quantitative composition properties of list replicability.

Given $\mathcal{C}_1 \subseteq \{\pm 1\}^{X_1}$ and $\mathcal{C}_2 \subseteq \{\pm 1\}^{X_2}$ over disjoint finite domains $X_1$ and $X_2$, define their *join* $\mathcal{C}_1 * \mathcal{C}_2 \subseteq \{\pm 1\}^{X_1 \sqcup X_2}$ as

$$\mathcal{C}_1 * \mathcal{C}_2 \;:=\; \big\{c \in \{\pm 1\}^{X_1 \sqcup X_2} : c|_{X_1} \in \mathcal{C}_1 \text{ and } c|_{X_2} \in \mathcal{C}_2\big\}.$$

---

[3] Conversely, since every $h \in \{\pm 1\}^{\mathcal{X}}$ satisfies $s_h(\mathcal{C}) \le s_{\mathbf{1}}(\mathcal{C}) + s_{-\mathbf{1}}(\mathcal{C})$, we have $s_{\max}(\mathcal{C}) \le s_{\mathbf{1}}(\mathcal{C}) + s_{-\mathbf{1}}(\mathcal{C})$.

The terminology is motivated by the fact that $\Delta_{(\mathcal{C}_1 * \mathcal{C}_2)^\pm}$ is homeomorphic to the topological join $\Delta_{\mathcal{C}_1^\pm} * \Delta_{\mathcal{C}_2^\pm}$: since $X_1$ and $X_2$ are disjoint, every realizable distribution for $\mathcal{C}_1 * \mathcal{C}_2$ is a convex combination of a distribution supported on $X_1$ and one supported on $X_2$ [CMW25, Lemma 23]. Note that the $\mathbb{Z}_2$-index is subadditive, and the coindex is superadditive under joins (see, e.g., [Mat03]),

$$\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}_1 * \mathcal{C}_2) \;\le\; \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}_1) + \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}_2) + 1, \tag{7}$$

and

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}_1 * \mathcal{C}_2) \;\ge\; \mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}_1) + \mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}_2) + 1. \tag{8}$$

In light of Theorem 1.1, it is natural to ask how list replicability behaves under joins. It is not difficult to prove $\textsc{lr}(\mathcal{C}_1 * \mathcal{C}_2) \le (\textsc{lr}(\mathcal{C}_1) + 1) \cdot (\textsc{lr}(\mathcal{C}_2) + 1)$ by essentially running the list-replicable algorithms for $\mathcal{C}_1$ and $\mathcal{C}_2$ separately on the corresponding parts of the sample and combining their outputs into a single hypothesis on $X_1 \sqcup X_2$: the resulting list consists of all pairs of hypotheses from the two lists. We improve this multiplicative bound to an additive one, which is sharp.

**Theorem 1.12.** *Given two concept classes $\mathcal{C}_1$ and $\mathcal{C}_2$ over disjoint finite domains, we have*

$$\textsc{lr}(\mathcal{C}_1 * \mathcal{C}_2) \le \textsc{lr}(\mathcal{C}_1) + \textsc{lr}(\mathcal{C}_2).$$

The proof uses a quantile-style coupling to synchronize the choices of the two LR learners. Instead of taking all pairs of possible outputs, it aligns the two output distributions on a common interval, so only linearly many pairs can appear. The full proof can be found in Section 7.

For matrices, Theorem 1.12 gives useful decomposition rules. Given two concept classes $\mathcal{C}_1$ and $\mathcal{C}_2$ over the same domain $X$, it follows from the definition that $\textsc{lr}(\mathcal{C}_1 \cup \mathcal{C}_2) \le \textsc{lr}(\mathcal{C}_1) + \textsc{lr}(\mathcal{C}_2)$, since one can run list-replicable learners for both classes and take the union of the resulting lists (see also Lemma 7.1). This operation corresponds to the vertical concatenation of the associated sign matrices. On the other hand, horizontal concatenation is controlled by joins: if $A_1$ and $A_2$ have the same number of rows, then the concept class associated with $[A_1 \;\; A_2]$ is naturally a subclass of the join of the two corresponding concept classes. We thus obtain subadditivity of list replicability under both horizontal and vertical concatenation of matrices.

**Corollary 1.13.** *For sign matrices with matching dimensions,*

$$\textsc{lr}\big(\begin{bmatrix} A_1 & A_2 \end{bmatrix}\big) \;\le\; \textsc{lr}(A_1) + \textsc{lr}(A_2) \qquad \textit{and} \qquad \textsc{lr}\left(\begin{bmatrix} A_3 \\ A_4 \end{bmatrix}\right) \;\le\; \textsc{lr}(A_3) + \textsc{lr}(A_4).$$

## 1.2 Related work

**Sign-rank.** The sign-rank of a matrix was introduced by Paturi and Simon [PS86], who observed that $\textsc{vc}(A) \le \mathrm{signrk}(A)$ as a consequence of the VC dimension of half-spaces. Sign-rank has since become a fundamental quantity in theoretical computer science, with connections to learning theory, communication complexity, circuit complexity, combinatorics, discrete geometry, and Banach space theory; see [HHP$^+$22]. Shortly after its introduction, Alon, Frankl, and Rödl [AFR85] used bounds on the number of connected components of real algebraic varieties [Mil64, Tho65, War68] to prove linear lower bounds on the sign-rank of random matrices.

For explicit matrices, the VC dimension bound remained the state of the art for nearly two decades, until the breakthrough of Forster [For02], who proved the $n \times n$ Hadamard matrix $H_n$ satisfies $\mathrm{signrk}(H_n) \ge \sqrt{n}$, the first super-logarithmic lower bound on the sign-rank of an explicit matrix.

**Topological methods.** The use of topological obstructions in combinatorics has a rich history, with Lovász's proof of Kneser's conjecture [Lov78] as a landmark example; see [Mat03] for a comprehensive treatment. The common theme is to associate a $\mathbb{Z}_2$-space to a combinatorial object and extract consequences from equivariant invariants such as the $\mathbb{Z}_2$-index and coindex [MZ02, CLSW04, ST06, STV09]. The sign-rank lower bound of Frick, Hosseini, and Vasileuski [FHV26] via the $\mathbb{Z}_2$-index fits squarely within this framework.

**List replicability.** Replicability, the requirement that an algorithm produce consistent outcomes when repeated under similar conditions, has become a vibrant research area in learning theory, with various rigorous formulations introduced and studied [BLM20, MM22, CMY23, BGH$^+$23, KVYZ23, EKK$^+$23, EKM$^+$23, MSS23, EHKS23, KKL$^+$24, KKMV23].

A key notion in this area is *global stability*, which emerged from the study of differentially private and online learning [BLM20, ABL$^+$22]. Chase, Moran, and Yehudayoff [CMY23] later reformulated global stability in the equivalent language of *list replicability*. Subsequent work has revealed that this notion is intrinsically linked to the geometry and topology of the space of realizable distributions [CMY23, CCMY24, BGHH25, CMW25, BHH$^+$26a, BHH$^+$26b, BHH$^+$26c].

**Extremal and intersection-closed concept classes.** The study of extremal and intersection-closed classes is motivated by the fact that their combinatorial structure gives rise to natural learning algorithms, sharp PAC sample-complexity bounds, and simple sample-compression constructions [MW16, CCMW22, CCH$^+$24, BHH$^+$26b, HSW90, HLW94, FW95, BDE98, Kuh99, DJ03, AO07, Dar15, Han16, BHQS21, RR22, Han24].

**Relationships between parameters.** Frick, Hosseini, and Vasileuski [FHV26] introduced the $\mathbb{Z}_2$-index and the closely related parameter $\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C})$, and showed

$$\textsc{vc}(\mathcal{C}) - 1 \;\leq\; \mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) \;\leq\; \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}) \;\leq\; \mathrm{signrk}(\mathcal{C}) - 1.$$

Chase, Moran, and Yehudayoff [CMY23, Theorem 3] proved that $\textsc{vc}(\mathcal{C}) \leq \textsc{lr}(\mathcal{C})$ for every concept class $\mathcal{C}$. The inequality $\textsc{lr}(\mathcal{C}) \leq \mathrm{signrk}(\mathcal{C})$ was conjectured in [CMY23], verified there for $\mathrm{signrk}(\mathcal{C}) = 2$, and later resolved in full by Blondal et al. [BHH$^+$26a].

Finally, the *Littlestone dimension* is a refinement of the VC dimension that characterizes the optimal mistake bound in online learning; in particular, the VC dimension is a lower bound on the Littlestone dimension. A celebrated result of Bun, Livni, and Moran [BLM20, ABL$^+$22] shows that every class with Littlestone dimension $d$ satisfies

$$\textsc{lr}(\mathcal{C}) \;\leq\; 2^{2^{O(d)}}.$$

## 1.3 Concluding remarks and open problems

**How large can the $\mathbb{Z}_2$-index be?** The lower bound $\textsc{vc}(\mathcal{C}) - 1 \leq \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$ provides $N \times N$ sign matrices with $\mathrm{Ind}_{\mathbb{Z}_2}(A) \geq \log N$, namely those with $\textsc{vc}(A) = \log N$.

Since a typical $N \times N$ sign matrix has sign-rank $\Omega(N)$, (3) leaves a wide gap between this logarithmic lower bound and the polynomial behaviour of sign-rank. It is natural to ask whether the $\mathbb{Z}_2$-index or the list replicability number of an $N \times N$ sign matrix can be super-logarithmic, or even polynomial, in $N$. Evidence so far has pointed in the negative direction: Frick et al. [FHV26] showed that a random $N \times N$ sign matrix satisfies $\mathrm{Ind}_{\mathbb{Z}_2}(A) = O(\log N)$ with high probability, and that the $N \times N$ Hadamard matrix, the standard example of an explicit matrix with large sign-rank, also satisfies $\mathrm{Ind}_{\mathbb{Z}_2}(H_N) = O(\log N)$. Nevertheless, the logarithmic barrier can be surpassed.

**Proposition 1.14.** *There exist $N \times N$ sign matrices $A$ with*

$$2\,\textsc{lr}(A) - 1 \geq \mathrm{Ind}_{\mathbb{Z}_2}(A) \geq \mathrm{coInd}_{\mathbb{Z}_2}(A) \;\geq\; \Omega\!\left(\frac{\log^2 N}{\log\log N}\right).$$

*Proof of Proposition 1.14, originally from [CMW25].* Consider the $m \times 2^m$ matrix $\mathcal{U}_m$, containing a column for every sign pattern on the $m$ rows. Chornomaz, Moran, and Waknine show that it has $\mathbb{Z}_2$-coindex of at least $m - 2$. Since the coindex is superadditive with respect to joins (recall Section 1.1.6), we take the join of $k = m/\log_2 m$ copies of $\mathcal{U}_m$.

$$\mathcal{U}_m^k = \mathcal{U}_m * \cdots * \mathcal{U}_m,$$

and obtain $\operatorname{coInd}_{\mathbb{Z}_2}(\mathcal{U}_m^k) \geq (m-2)(m)/\log m$. Note $\mathcal{U}_m^k$ has $m^{m/\log m} = 2^m$ rows and $(m/\log m)2^m$ columns. Therefore, it is contained in an $N \times N$ sign matrix $A$, where $N = (m/\log m)2^m$, with

$$2\,\textsc{lr}(A) - 1 \geq \operatorname{Ind}_{\mathbb{Z}_2}(A) \geq \operatorname{coInd}_{\mathbb{Z}_2}(A) \geq \Omega\left(\frac{m^2}{\log m}\right) = \Omega\left(\frac{\log^2 N}{\log\log N}\right). \qquad \square$$

The lower bound provided in Proposition 1.14 is currently the strongest known lower bound on the list replicability and the $\mathbb{Z}_2$-index of any $N \times N$ matrix. We conjecture that this bound is essentially tight (see also [FHV26, Question 37]).

**Conjecture 1.15.** *Every $N \times M$ sign matrix $A$ satisfies*

$$\textsc{lr}(A) = O(\log N \log M).$$

**Upper bounds by a function of VC dimension?** Perhaps the most intriguing open question in the study of list replicability is whether $\textsc{lr}(\mathcal{C})$ can be upper bounded by a function of VC dimension. Originally posed in [CMY23], this question was resolved for extremal concept classes in [BHH$^+$26b], where it was shown that $\textsc{lr}(\mathcal{C}) = \Theta(\textsc{vc}(\mathcal{C}))$. However, the question remains open for arbitrary finite concept classes.

Moreover, our result that $\operatorname{Ind}_{\mathbb{Z}_2}(\mathcal{C}) = O(\textsc{lr}(\mathcal{C}))$ suggests a natural addition to the question:

**Problem 1.16.** *Can $\operatorname{Ind}_{\mathbb{Z}_2}(\mathcal{C})$ or $\textsc{lr}(\mathcal{C})$ be bounded by a function of $\textsc{vc}(\mathcal{C})$?*

Note that a positive answer to Problem 1.16 would also imply $\operatorname{coInd}_{\mathbb{Z}_2}(\mathcal{C})$ is bounded by a function of $\textsc{vc}(\mathcal{C})$. This question has already been asked [CMW25] as it has important implications for the learnability of *large-margin half-spaces*.

**Embeddings.** One potential approach to resolving Problem 1.16 is to embed the concept class $\mathcal{C}$ into an extremal or intersection-closed concept class while keeping the VC dimension small. Recall that, in both settings, $\operatorname{Ind}_{\mathbb{Z}_2}$ and $\textsc{lr}$ are bounded above linearly by $\textsc{vc}$. The problem of embedding into extremal concept classes has been widely studied and remains open [MW16, CCH$^+$24].

The analogous problem for intersection-closed classes was studied by Hanneke in [Han24] through the minimum star number $s_{\min}$. Hanneke showed that if $\textsc{vc}(\mathcal{C}) = 1$, then $s_{\min}(\mathcal{C}) = 1$; equivalently, every concept class of VC dimension 1 can be embedded into a generalized intersection-closed class of VC dimension 1. On the other hand, Hanneke also observed that there exists a family of finite concept classes with VC dimension 3 and arbitrarily large minimum star number $s_{\min}$. This implies that, in general, one cannot embed an arbitrary concept class into an intersection-closed class without incurring a blow-up in VC dimension.

Hanneke asked what happens in the case $\textsc{vc}(\mathcal{C}) = 2$. We show that $s_{\min}$ can be arbitrarily large in this case as well.

**Theorem 1.17.** *For every $k \geq 2$, there is a finite total concept class $C_k$ with*

$$\textsc{vc}(C_k) = 2 \qquad \textit{and} \qquad s_{\min}(C_k) \geq k.$$

*Proof.* Let $X_k = [k] \times [k]$, and write $R_i = \{i\} \times [k]$ for the $i$th row. For convenience, we will treat concepts as subsets of $\mathcal{X}$, that is $c(x) = 1$ if $x \in c$ and $c(x) = -1$ otherwise. Define

$$C_k = \{\varnothing\} \cup \{R_i : i \in [k]\} \cup \{R_i \setminus \{x\} : i \in [k],\ x \in R_i\}.$$

First, $\textsc{vc}(C_k) \geq 2$. Indeed, if $x, y \in R_i$ are distinct, then $\varnothing$, $R_i$, $R_i \backslash \{x\}$, $R_i \backslash \{y\}$ realize all four labelings on $\{x, y\}$. Also, since any concept can only have 1's on one row $R_i$, it is not hard to see $\textsc{vc}(C_k) \leq 2$. We conclude $\textsc{vc}(\mathcal{C}_k) = 2$.

Now fix any center function $h : X_k \to \{-1, 1\}$. We show that $s_h(C_k) \geq k$.

If every row $R_i$ contains some point $x_i$ with $h(x_i) = -1$, let

$$S = \{x_1, \ldots, x_k\}.$$

Then $\varnothing$ agrees with $h$ on $S$, and for each $i \in [k]$, the concept $R_i$ flips exactly the point $x_i$ on $S$. Hence $S$ is a star of size $k$ centered at $h$.

Otherwise, some row $R_i$ is entirely labeled 1 by $h$. Take $S = R_i$. Then the concept $R_i$ agrees with $h$ on $S$, and for each $x \in R_i$, the concept $R_i \setminus \{x\}$ flips exactly $x$ on $S$. Hence $S$ is again a star of size $k$ centered at $h$.

Thus $s_h(\mathcal{C}_k) \geq k$ for every $h$, so $s_{\min}(\mathcal{C}_k) \geq k$. □

## 1.4 Technical overview of Theorem 1.1 and Theorem 1.2

In this section, we briefly describe how we prove Theorem 1.1 and Theorem 1.2. Complete proofs are in Section 3.

**Bounding index by LR.** Theorem 1.1 states that for any concept class $\mathcal{C}$,

$$\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}) \leq 2\,\mathrm{LR}(\mathcal{C}) - 1.$$

To prove this, we first use the fact that a list-replicable algorithm with list size $L$ gives an antipodal-free open cover of the distribution space $\Delta_{\mathcal{C}^\pm}$, where each point is contained in at most $2L$ open sets (see Theorem 2.11). This bounded overlap allows us to use a partition of unity to construct a continuous map from $\Delta_{\mathcal{C}}$ into $\mathbb{R}^{2L} \setminus \{0\}$.

The antipodal symmetry of the cover ensures that this map is $\mathbb{Z}_2$-equivariant, while antipodal-freeness ensures that the image avoids the origin. After normalizing, we obtain a $\mathbb{Z}_2$-equivariant map

$$\Delta_{\mathcal{C}} \xrightarrow{\mathbb{Z}_2} \mathbb{S}^{2L-1}.$$

Therefore, by the definition of the $\mathbb{Z}_2$-index,

$$\mathrm{Ind}_{\mathbb{Z}_2}(\Delta_{\mathcal{C}}) \leq 2L - 1.$$

**Separating index from sign-rank.** The separation of sign-rank and index in Theorem 1.2 hinges on a particular family of matrices $B_q$ for which $\mathrm{signrk}(B_q)$ grows polynomially in $q$, while $\mathrm{Ind}_{\mathbb{Z}_2}(B_q)$ is bounded. This family was first used by Alon, Moran, and Yehudayoff to separate sign-rank and VC dimension [AMY16].

**Definition 1.18** (Finite Projective Plane)**.** *Let* $\mathrm{PG}(2, q)$ *be the finite projective plane of order* $q$*. It is an incidence geometry* $(\mathcal{P}, \mathcal{L})$ *with* $N = q^2 + q + 1$ *points and lines such that:*

1. *Any two distinct lines* $\ell_1, \ell_2 \in \mathcal{L}$ *intersect in exactly one point* $p \in \mathcal{P}$*.*
2. *Any two distinct points* $p_1, p_2 \in \mathcal{P}$ *are contained in exactly one line* $\ell \in \mathcal{L}$*.*

*Let* $B_q$ *be the* $N \times N$ *indicator matrix of this incidence geometry. That is,* $B_q$ *has rows indexed by lines and columns indexed by points, and*

$$(B_q)_{ij} = \begin{cases} 1 & \text{if } p_j \in \ell_i \\ -1 & \text{if } p_j \notin \ell_i. \end{cases}$$

Matrices of this form were useful for the separation of VC dimension and sign-rank because of the polynomial growth of $\mathrm{signrk}(B_q)$, a result which we will apply directly for our separation of index and sign-rank.

**Theorem 1.19** ([AMY16])**.** *The* $N \times N$ *indicator matrix* $B_q$ *of* $\mathrm{PG}(2, q)$ *satisfies the inequality*

$$\mathrm{signrk}(B_q) \geq \frac{q^2 - 1}{\sqrt{q}(q-1)} \geq N^{\frac{1}{4}}.$$

The remaining half of our separation requires bounding $\mathrm{Ind}_{\mathbb{Z}_2}(\Delta_{B_q})$ by a constant. This is done by combining Theorem 1.1 with a list-replicable algorithm for $B_q$, which we give in Theorem 3.2.

**The algorithm.** Here we treat the matrix $B_q$ as a concept class with concepts labeled by lines and domain points labeled by points. We design a simple list-replicable algorithm for $B_q$ that shows

$$\mathrm{LR}(B_q) \leq 3.$$

The algorithm proceeds as follows. If the sample contains two distinct points labeled 1, then there is a unique line passing through both of them, so the algorithm outputs that line. Otherwise, if the sample contains only a single point labeled 1 and that point has low sampling probability, the algorithm ignores it, since doing so affects the error only negligibly with high probability. Finally, if the sample contains exactly one point labeled 1 and that point has high sampling probability, the algorithm outputs the indicator function of that point. The complete proof of this step can be found in Theorem 3.2.

Combining the two steps, we have that the family $B_q$ has $\mathrm{Ind}_{\mathbb{Z}_2}(B_q) \leq 5$, but $\mathrm{signrk}(B_q) = \Omega(\sqrt{q}) = \Omega(N^{1/4})$.

# 2 Preliminaries

In this section, we collect all topological and learning-theoretic tools that we use. We first introduce basic notions from $\mathbb{Z}_2$-equivariant topology, which allows us to formally define the $\mathbb{Z}_2$-index of a sign matrix and its corresponding concept class. Afterwards, we go through some fundamental learning theoretical definitions and results, including list replicability and its topological and algorithmic interpretations.

## 2.1 The $\mathbb{Z}_2$-topological framework

**Definition 2.1** (Simplicial Complex)**.** *An* abstract simplicial complex $\mathcal{K}$ *on a vertex set* $V$ *is a collection of finite subsets of* $V$ *that are closed under taking subsets. Its elements are called* simplices *and its* dimension *is* $\max_{\sigma\in\mathcal{K}} |\sigma| - 1$. *The* geometric realization *of* $\mathcal{K}$ *is the topological space*

$$\|\mathcal{K}\| := \left\{ x \in \mathbb{R}^V : x = \sum_{v\in V} \lambda_v e_v,\ \lambda_v \geq 0,\ \sum_{v\in V} \lambda_v = 1,\ \mathrm{supp}(x) \in \mathcal{K} \right\},$$

*where* $e_v$ *denotes the standard basis vector indexed by* $v$ *and* $\mathrm{supp}(x) = \{v : \lambda_v > 0\}$.

**Definition 2.2** (Free $\mathbb{Z}_2$-Simplicial Complex)**.** *A* free $\mathbb{Z}_2$-simplicial complex *is a pair* $(\mathcal{K}, \nu)$ *where* $\mathcal{K}$ *is an abstract simplicial complex and* $\nu$ *is a* fixed-point-free simplicial involution, *i.e., a simplicial map* $\nu\colon \mathcal{K} \to \mathcal{K}$ *satisfying* $\nu \circ \nu = \mathrm{id}_{\mathcal{K}}$ *and* $\nu(\sigma) \neq \sigma$ *for every* $\sigma \in \mathcal{K}$.

**Definition 2.3** (Free $\mathbb{Z}_2$-Space)**.** *A* free $\mathbb{Z}_2$-space *is a pair* $(X, \nu)$ *where* $X$ *is a topological space and* $\nu$ *is a continuous* fixed-point-free involution, *i.e., a continuous map* $\nu\colon X \to X$ *satisfying* $\nu \circ \nu = \mathrm{id}_X$ *and* $\nu(x) \neq x$ *for every* $x \in X$.

The geometric realization of a free $\mathbb{Z}_2$-simplicial complex is naturally a free $\mathbb{Z}_2$-space. Indeed, if $\mathcal{K}$ is a free $\mathbb{Z}_2$-simplicial complex with involution $\nu$, the map

$$\sum_{v\in V} \lambda_v e_v \mapsto \sum_{v\in V} \lambda_v e_{\nu(v)}$$

is an involution on $\|\mathcal{K}\|$.

**Example 2.4.** *The* $n$*-dimensional sphere* $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$ *equipped with the canonical involution* $\nu(x) = -x$ *is a free* $\mathbb{Z}_2$*-space.*

**Definition 2.5** (Antipodal Map)**.** *Let* $(X, \nu)$ *and* $(Y, \omega)$ *be free* $\mathbb{Z}_2$*-spaces. We say a map* $f\colon X \to Y$ *is* antipodal *if it commutes with the involutions on each space, that is* $f \circ \nu = \omega \circ f$. *We write* $f\colon X \xrightarrow{\mathbb{Z}_2} Y$ *to denote that* $f$ *is an antipodal map between* $X$ *and* $Y$.

The antipodal terminology comes from the fact that $f$ cannot have $f(x) = f(\nu(x))$ for any $x \in X$ due to the involutions $\nu$ and $\omega$ being fixed-point-free.

**Definition 2.6** ($\mathbb{Z}_2$-index)**.** *The* $\mathbb{Z}_2$-index *of a free* $\mathbb{Z}_2$*-space* $(X, \nu)$ *is the minimum integer* $n$ *such that there is an antipodal map from* $X$ *into the* $n$*-dimensional sphere.*

$$\mathrm{Ind}_{\mathbb{Z}_2}(X) \coloneqq \min\{n \in \mathbb{N} : \exists f \colon X \xrightarrow{\mathbb{Z}_2} \mathbb{S}^n\}.$$

It is also natural to instead consider the smallest sphere that maps antipodally into $X$, giving us the notion of $\mathbb{Z}_2$-coindex.

**Definition 2.7** ($\mathbb{Z}_2$-coindex)**.** *The* $\mathbb{Z}_2$*-coindex of a* $\mathbb{Z}_2$ *space* $X$ *is the maximum* $n$ *such that there is an antipodal map from* $\mathbb{S}^n$ *into* $X$*.*

$$\mathrm{coInd}_{\mathbb{Z}_2}(X) \coloneqq \max\{n \in \mathbb{N} : \exists f \colon \mathbb{S}^n \xrightarrow{\mathbb{Z}_2} X\}.$$

The following is a generalization of the classical Borsuk–Ulam Theorem.

**Theorem 2.8** (Borsuk–Ulam)**.** *Let* $(X, \nu)$ *be a free* $\mathbb{Z}_2$*-space. Then* $\mathrm{coInd}_{\mathbb{Z}_2}(X) \leq \mathrm{Ind}_{\mathbb{Z}_2}(X)$*.*

We refer the reader to Jiří Matoušek's excellent textbook on topological methods in combinatorics for more details on $\mathbb{Z}_2$-spaces and their applications to combinatorics [Mat03, Section 5].

## 2.2 The $\mathbb{Z}_2$-index of sign matrices and concept classes

To prove our results in full generality, we will work in the broader framework of partial sign matrices and partial concept classes.

If $A \in \{1, -1, \star\}^{M \times N}$ is a partial sign matrix, the *sign-rank* of $A$, denoted $\mathrm{signrk}(A)$, is the minimum rank of a real matrix $B$ that captures the sign patterns of $A$. That is

$$\mathrm{sign}(B_{ij}) = A_{ij} \qquad \text{for every } i \in [M],\, j \in [N] \text{ with } A_{ij} \neq \star.$$

**Partial concept classes.** Partial sign matrices can also be interpreted as *partial concept classes.* Given a domain $\mathcal{X}$, a partial concept class $\mathcal{C} \subseteq \{+1, -1, \star\}^{\mathcal{X}}$ is a family of functions, called *concepts*, mapping $\mathcal{X} \to \{+1, -1, \star\}$. These are fundamental objects in learning theory because they encode bias, which is a prerequisite for any meaningful definition of learning.

The partial concept class of a partial sign matrix $A \in \{1, -1, \star\}^{M \times N}$ is a family $\mathcal{C}_A \subseteq \{+1, -1, \star\}^{[N]}$ with (partial) concepts given by rows of $A$. That is,

$$c_i \colon j \mapsto A_{ij}.$$

**Space of realizable distributions.** We say that a distribution $\mu \sim \mathcal{X} \times \{\pm 1\}$ is *realizable* by a partial concept class $\mathcal{C}$ if there exists a partial concept $c$ such that $c(x) = b$ for each $(x, b)$ in the support of $\mu$. We denote the set of all realizable distributions $\mu$ by

$$\Delta_{\mathcal{C}} \coloneqq \{\mu : \mu \text{ realizable by } \mathcal{C}\}.$$

When equipped with *the total variation* (TV) *distance*, $\Delta_{\mathcal{C}}$ forms a metric space. $\Delta_{\mathcal{C}}$ has a natural geometric realization as a subset of the $\ell_1$-sphere.

$$\Delta_{\mathcal{C}} \coloneqq \left\{\mu \in \mathbb{R}^{\mathcal{X}} :\ \|\mu\|_1 = 1 \text{ and } \exists c \in \mathcal{C} \text{ with } c(x) = b\mu(x)\ \forall (x, b) \in \mathrm{supp}(\mu)\right\} \subseteq \mathbb{R}^{\mathcal{X}}. \tag{9}$$

Note that each partial concept $c \in \mathcal{C}$ corresponds to a simplex $\sigma_c \in \Delta_{\mathcal{C}}$:

$$\sigma_c = \mathrm{conv}(\{b \times e_x :\ x \in \mathcal{X}, c(x) = b\})$$

where $\{e_x\}_{x\in\mathcal{X}}$ is the standard basis of $\mathbb{R}^{\mathcal{X}}$.

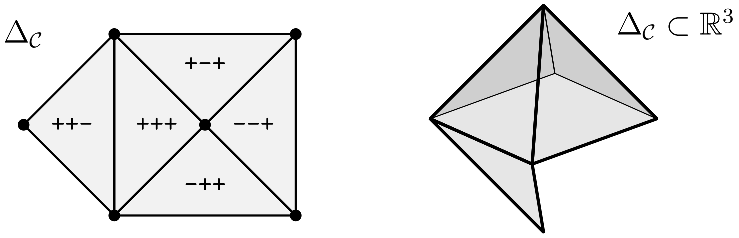

Figure 1: Two views of the simplicial complex $\Delta_{\mathcal{C}}$ for $\mathcal{C} = \{$++−, +++, +−+, −−+, −++$\}$.

The following properties of $\Delta_{\mathcal{C}}$ are easy to deduce and can also be found in [CMW25, BHH+26b, FHV26].

- $\Delta_{\mathcal{C}}$ is a finite compact simplicial complex.
- $\Delta_{\mathcal{C}}$ has vertex set contained in $\mathcal{X} \times \{\pm 1\}$ and is a subcomplex of the cross-polytope boundary $\Delta_{\{\pm 1\}^{\mathcal{X}}}$.
- Each maximal simplex of $\Delta_{\mathcal{C}}$ equals $\sigma_c$ for some $c \in \mathcal{C}$.
- If $\mathcal{C}^{\pm} = C \cup -C$, then $\Delta_{\mathcal{C}^{\pm}}$ is a free $\mathbb{Z}_2$-space with a natural $\mathbb{Z}_2$-action given by negating labels: $\mu \mapsto -\mu$, where $-\mu$ assigns mass $\mu(x, b)$ to $(x, -b)$.

**Definition 2.9** (Index/Coindex)**.** *Let $A$ be a partial sign matrix and let $\mathcal{C}_A$ be its associated partial concept class. Then we define*

$$\mathrm{Ind}_{\mathbb{Z}_2}(A) \coloneqq \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}_A) \coloneqq \mathrm{Ind}_{\mathbb{Z}_2}(\Delta_{\mathcal{C}_A^{\pm}})$$

$$\mathrm{coInd}_{\mathbb{Z}_2}(A) \coloneqq \mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}_A) \coloneqq \mathrm{coInd}_{\mathbb{Z}_2}(\Delta_{\mathcal{C}_A^{\pm}})$$

## 2.3 List replicability

The motivation for our discussion of the concept class of a sign matrix in the previous section is that this object allows us to leverage existing *list replicability* techniques from learning theory.

A *learning rule* is a (possibly randomized) function $\mathcal{A}$ that maps any *sample* $S \in \bigcup_{n=0}^{\infty} (\mathcal{X} \times \{\pm 1\})^n$ to a *hypothesis* $\mathcal{A}(S) \in \{\pm 1\}^{\mathcal{X}}$. The error or *population loss* of a hypothesis with respect to a distribution $\mu$ over $\mathcal{X} \times \{\pm 1\}$ is measured as

$$\mathrm{loss}_{\mu}(h) = \mathbb{P}_{(x,y)\sim\mu}[h(x) \neq y].$$

**Definition 2.10** (List Replicability, [CMY23, DPVWV23])**.** *A learning rule $\mathcal{A}$ is an $(\epsilon, L)$-*list-replicable learner *for a partial concept class $\mathcal{C}$ if for every $\delta > 0$ there exists a sample complexity $n \coloneqq n(\delta)$ such that the following holds. For every distribution $\mu$ realizable by $\mathcal{C}$, there exists a list of hypothesis $h_1, \ldots, h_L \in \{\pm 1\}^{\mathcal{X}}$ such that*

$$\mathrm{loss}_{\mu}(h_i) \leq \epsilon \;\forall i \quad \textit{and} \quad \mathbb{P}_{S\sim\mu^n}[\mathcal{A}(S) \in \{h_1, \ldots, h_L\}] \geq 1 - \delta.$$

*The $\epsilon$-*list replicability number *of $\mathcal{C}$ is*

$$\textsc{lr}(\mathcal{C}, \epsilon) \coloneqq \min\{L : \exists (\epsilon, L)\textit{-list-replicable learner for } \mathcal{C}\},$$

*with $\textsc{lr}(\mathcal{C}, \epsilon) = \infty$ if none exists. The* list replicability number *of $\mathcal{C}$ is*

$$\textsc{lr}(\mathcal{C}) \coloneqq \sup_{\epsilon > 0} \textsc{lr}(\mathcal{C}, \epsilon).$$

*We say $\mathcal{C}$ is* list-replicable *if $\textsc{lr}(\mathcal{C}) < \infty$.*

One of the advantages of list replicability is that this purely learning theoretic definition also admits a topological description via closed covers of $\Delta_{\mathcal{C}}$. Recall that for each $h \in \{\pm\}^{\mathcal{X}}$ we defined

$$\sigma_h = \mathrm{conv}(\{b \times e_x \, : \, x \in \mathcal{X}, h(x) = b\})$$

where $\{e_x\}_{x\in\mathcal{X}}$ is the standard basis of $\mathbb{R}^{\mathcal{X}}$.

Now for each hypothesis $h \in \{\pm 1\}^{\mathcal{X}}$ let $B_\epsilon(\sigma_h)$ denote set of all points in $\Delta_{\{\pm 1\}^{\mathcal{X}}}$ (equivalently the cross-polytope boundary) that are within total variation distance less than $\epsilon$ from the simplex $\sigma_h$. Denote by $\overline{B_\epsilon}(\sigma_h)$ the respective closure. We call the sets $B_\epsilon(\sigma_h)$ and $\overline{B_\epsilon}(\sigma_h)$ the open and closed *$\epsilon$-loss* sets of $h$ because they contain all distributions $\mu \in \Delta_{\mathcal{C}}$ for which $\mathrm{loss}_\mu(h) < \epsilon$ and $\mathrm{loss}_\mu(h) \le \epsilon$ respectively.

A recent line of work in [CMY23, CCMY24, BHH$^+$26b] has shown a correspondence between list-replicable learners for $\mathcal{C}$ and closed/open covers of $\Delta_{\mathcal{C}}$. In both cases, list size corresponds to the *overlap degree* of the cover, which we define as the maximum number of sets in the cover with a common non-empty intersection.

**Theorem 2.11** ([CCMY24, Corollary 23], [BHH$^+$26b, Theorem A])**.** *Let $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ be a finite partial concept class and let $\epsilon > 0$.*

- (Closed cover characterization) *The $\epsilon$-list replicability number $\textsc{lr}(\mathcal{C}, \epsilon)$ equals the minimum integer $L$ for which there exists a closed cover $\mathcal{F} = \{F_h : h \in \{\pm 1\}^{\mathcal{X}}\}$ of $\Delta_{\mathcal{C}}$ with overlap degree at most $L$ and $F_h \subseteq \overline{B}_\varepsilon(\sigma_h)$ for every $h \in \{\pm 1\}^{\mathcal{X}}$.*
- (Open cover characterization) *If $L := \textsc{lr}(\mathcal{C}, \epsilon)$, then for every $\eta > \epsilon$ there exists an open cover $\mathcal{U} = \{U_h : h \in \{\pm 1\}^{\mathcal{X}}\}$ of $\Delta_{\mathcal{C}}$ with overlap degree at most $L$ such that $U_h \subseteq B_\eta(\sigma_h)$ for every $h \in \{\pm 1\}^{\mathcal{X}}$. Conversely, if for some $\eta > 0$ and $L \in \mathbb{N}$ there exists an open cover $\mathcal{U} = \{U_h : h \in \{\pm 1\}^{\mathcal{X}}\}$ of $\Delta_{\mathcal{C}}$ with overlap degree at most $L$ such that $U_h \subseteq B_\eta(\sigma_h)$ for every $h \in \{\pm 1\}^{\mathcal{X}}$, then there exists some $\epsilon < \eta$ such that $\textsc{lr}(\mathcal{C}, \epsilon) \le L$.*

The closed cover characterization of list replicability was shown in [CMY23] for total concept classes, but the argument immediately generalizes to partial concept classes. The open cover characterization is closely related, and was studied in a stronger form as *simplicial covering dimension* in [BHH$^+$26b].

# 3 Separating sign-rank and index

Recall from Section 1.4 that our separation of sign-rank and index in Theorem 1.2 leverages the family of incidence matrices $B_q$ for the finite projective planes $\mathrm{PG}(2, q)$. We are guaranteed the polynomial growth of $\mathrm{signrk}(B_q)$ from Theorem 1.19, so this section will collect the proofs of Theorem 1.1 and Theorem 1.2 to bound $\mathrm{Ind}_{\mathbb{Z}_2}(B_q)$ by a constant.

## 3.1 List replicability number upper bounds index

In this section, we prove our main technical result Theorem 1.1, stated here in a slightly stronger form.

**Theorem 3.1.** *For every (partial) concept class $\mathcal{C}$ and error parameter $\varepsilon \in (0, \frac{1}{2})$,*

$$\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}) \ \le\ 2\,\textsc{lr}(\mathcal{C}, \varepsilon) - 1.$$

The argument begins by producing a closed cover of $\Delta_{\mathcal{C}}$ with small overlap degree, as guaranteed by Theorem 2.11. We use basic topology to extend that cover to an open cover of $\Delta_{\mathcal{C}^\pm}$ which respects the $\mathbb{Z}_2$ structure of that space. This step is necessary because $\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$ is by definition a property of $\Delta_{\mathcal{C}^\pm}$ rather than $\Delta_{\mathcal{C}}$.

From here, we map $\Delta_{\mathcal{C}^\pm}$ into low-dimensional Euclidean space, which we in turn project into a sphere.

$$\Delta_{\mathcal{C}^\pm} \xrightarrow{\mathbb{Z}_2} \mathbb{R}^{2L} \setminus 0 \xrightarrow{\mathbb{Z}_2} \mathbb{S}^{2L-1}.$$

By definition, the composition of these maps witnesses an upper bound on $\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$. The construction of the map $\Delta_{\mathcal{C}^\pm} \xrightarrow{\mathbb{Z}_2} \mathbb{R}^{2L}$ was discovered by considering the nerve complex associated with the open cover of $\Delta_{\mathcal{C}^\pm}$. This is a simplicial complex that encodes intersection data, making it a natural candidate to analyze overlap degree.

*Proof of Theorem 3.1.* For convenience, set $L \coloneqq \textsc{lr}(\mathcal{C}, \varepsilon)$. Let $\mathcal{A}$ be an $(\varepsilon, L)$-list-replicable learner for $\mathcal{C}$. By the closed cover characterization of list replicability in *Theorem* 2.11, $\mathcal{A}$ induces a closed cover of $\Delta_{\mathcal{C}}$ given by $\mathcal{F} = \{F_h : h \in \{\pm 1\}^{\mathcal{X}}\}$, where

$$\text{the overlap degree of } \mathcal{F} \text{ is } L \qquad \text{and} \qquad F_h \subseteq \overline{B_\varepsilon}(\sigma_h) \text{ for all } h \in \{\pm 1\}^{\mathcal{X}}.$$

Recall that $\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$ is defined to be the $\mathbb{Z}_2$-index of $\Delta_{\mathcal{C}^\pm}$. Since $\Delta_{\mathcal{C}}$ is closed subspace of $\Delta_{\mathcal{C}^\pm}$, each set $F_h$ remains closed under the natural inclusion $\Delta_{\mathcal{C}} \hookrightarrow \Delta_{\mathcal{C}^\pm}$. Hence, the family

$$\mathcal{F}' \coloneqq \{F'_h : \; F'_h = F_h \cup (-F_{-h}) \text{ for some } h \in \{\pm 1\}^{\mathcal{X}}\}.$$

is a closed cover of $\Delta_{\mathcal{C}^\pm}$ with overlap degree at most $2L$ satisfying $F'_h \subseteq \overline{B_\epsilon}(\sigma_h)$ and $F'_h = -F'_{-h}$ for all $h$.

Now pick any $\epsilon_0 \in (\epsilon, 1/2)$. Because $\Delta_{\mathcal{C}^\pm}$ is compact, we can extend each closed set $F'_h$ to an open set $U_h \subseteq B_{\epsilon_0}(\sigma_h)$ without increasing the overlap degree of the family and while preserving the antipodal symmetry $U_h = -U_{-h}$[4]. Then $\mathcal{U} \coloneqq \{U_h : h \in \{\pm 1\}^{\mathcal{X}}\}$ is a finite open cover of $\Delta_{\mathcal{C}^\pm}$ with overlap degree at most $2L$ such that $U_h \subseteq \overline{B_{\epsilon_0}}(\sigma_h)$ and $U_h = U_{-h}$ for all $h$.

Since $\Delta_{\mathcal{C}^\pm}$ is a compact metric space, there exists a partition of unity $\{\phi_h\}_h$ subordinate to the open cover $\mathcal{U}$ [Rud87, Theorem 2.13]. That is, there exists a collection of maps $\phi_h \colon \Delta_{\mathcal{C}^\pm} \to [0,1]$ such that

$$\operatorname{supp} \phi_h \subseteq U_h \qquad \text{and} \qquad \sum_h \phi_h \equiv 1.$$

We can make an antipodally symmetric version of $\{\phi_h\}_h$ by setting

$$\phi^*_h(x) \coloneqq \frac{1}{2}(\phi_h(x) + \phi_{-h}(-x)).$$

Note that the antipodal symmetry $U_h = -U_{-h}$ guarantees that $\operatorname{supp} \phi^*(x) \subseteq U_h$.

Next, let $\{p_h\}_h \subseteq \mathbb{R}^{2L}$ be a set of points in general position such that $p_{-h} = -p_h$. That is, a subset $P \subset \{p_h\}$ of size $2L$ contains zero in its convex span if and only if $P$ contains a pair $\{p_h, p_{-h}\}$ for some $h$.

Using the antipodally symmetric partition of unity $\phi^*$ and the points $\{p_h\}_h$, we define an antipodal map $f : \Delta_{\mathcal{C}^\pm} \xrightarrow{\mathbb{Z}_2} \mathbb{R}^{2L}$ by

$$f(x) = \sum_h \phi^*_h(x) p_h.$$

We claim that $0 \notin \mathrm{im}(f)$. Indeed, for every $x \in \Delta_{\mathcal{C}^\pm}$, the overlap degree of $\mathcal{U}$ ensures $\phi^*_h(x)$ is nonzero for at most $2L$ hypotheses $h$. Additionally, $\varepsilon_0 < 1/2$ implies that no $U_h$ contains both $x$ and $-x$, so $\phi^*$ must be zero for at least one of the two points. Thus, the coefficients of $p_h$ and $p_{-h}$ cannot both be positive. It follows that $f(x)$ is a convex combination of at most $2L$ points of $P$, which does not contain both $p_h$ and $p_{-h}$. Since $P$ is in general position, $0 \notin \mathrm{im}(f)$.

We may therefore project the image of $f$ into the sphere $\mathbb{S}^{2L-1} \subset \mathbb{R}^{2L}$ to get a map

$$\frac{f}{\|f\|_2} : \Delta_{\mathcal{C}^\pm} \xrightarrow{\mathbb{Z}_2} \mathbb{S}^{2L-1}.$$

Exhibiting such an antipodal map into the sphere shows that $\mathrm{Ind}_{\mathbb{Z}_2}(\Delta_{\mathcal{C}^\pm}) \leq 2L - 1$ by definition of index. □

---

[4] Let $d \colon \Delta_{\mathcal{C}^\pm} \to [0,1]$ be a function measuring the total variation distance $d(x)$ from a point $x \in \Delta_{\mathcal{C}}$ to the $L+1$'st closest set $F'_h$. By compactness, this function achieves a strictly positive minimum $\beta$. Take $U_h \coloneqq \cup_{x \in F'_h} B_{\beta/2}(x)$.

## 3.2 A list-replicable algorithm for $\mathrm{PG}(2, q)$

As defined in Definition 1.18, recall the incidence geometry $\mathrm{PG}(2, q) = (\mathcal{P}, \mathcal{L})$. Let $\mathcal{C}_q$ denote the corresponding concept class

$$\begin{aligned} \mathcal{C}_q &:= \{c_\ell : \ell \in \mathcal{L}\} \\ c_\ell : \mathcal{P} &\to \{\pm 1\} \\ p &\mapsto \begin{cases} 1 & \text{if } p \in \ell \\ -1 & \text{otherwise.} \end{cases} \end{aligned}$$

We exhibit a list-replicable algorithm for $\mathcal{C}_q$. By sampling random points $x \in \mathcal{P}$ and whether they lie on a line $\ell$, the learner's goal is to output a predictor of whether future points lie on the same line.

**Theorem 3.2.** *The list replicability number of $\mathcal{C}_q$ satisfies $\mathrm{LR}(\mathcal{C}_q) \leq 3$.*

*Proof.* We extend the concept class $\mathcal{C}_q$ to a larger hypothesis class $\mathcal{H}_q$

$$\mathcal{H}_q = \{c_\ell : \ell \in \mathcal{L}\} \cup \{c_p : p \in \mathcal{P}\} \cup \{c_{-1}\},$$

where $c_{-1}$ is the all-minus hypothesis, and $c_p$ is the indicator function for a single point $p$, evaluating to 1 on $p$ and $-1$ everywhere else.

Since $\mathcal{H}_q$ is finite, we have by the union bound and Hoeffding's inequality[5] that there exists an $n(\epsilon, \delta)$ such that for any $\mathcal{D} \in \Delta_{\mathcal{C}_q}$, we can estimate the population loss of all hypotheses in $\mathcal{H}_q$ simultaneously:

$$\mathbb{P}_{S \sim \mathcal{D}^n}\left[\sup_{h \in \mathcal{H}_q} |\mathrm{loss}_S(h) - \mathrm{loss}_{\mathcal{D}}(h)| \leq \frac{\epsilon}{8}\right] \geq 1 - \delta. \tag{10}$$

Next, we define the learning rule $\mathcal{A}$. For any $\epsilon, \delta$, let its sample size be $n(\epsilon, \delta)$ so as to satisfy (10).

**Algorithm 1** The learning rule $\mathcal{A}$

```
Sample $S = (S_1, \ldots, S_n) \sim \mathcal{D}^n$.
for $p \in \mathcal{P}$ do
    Let $\hat{\mathcal{D}}(p) = \frac{1}{n}|\{i \in [n] : S_i = (p, 1)\}|$.
end for
Let $P_0 \leftarrow \{p \in \mathcal{P} : \hat{\mathcal{D}}(p) > 0\}$
Let $P_1 \leftarrow \{p \in \mathcal{P} : \hat{\mathcal{D}}(p) > 7\epsilon/8\}$
if $|P_0| \geq 2$ then
    Output $c_\ell$, where $\ell$ is the unique line containing all points in $P_0$.
else if $|P_1| = 1$ then
    Output $c_p$, where $P_1 = \{p\}$.
else
    Output $c_{-1}$.
end if
```

We start by confirming that this algorithm is a PAC learner.

By (10), with probability $\geq 1 - \delta$,

$$\sup_{h \in \mathcal{H}_q} |\mathrm{loss}_S(h) - \mathrm{loss}_{\mathcal{D}}(h)| \leq \frac{\epsilon}{8}.$$

---

[5] Let $c \in \mathbb{R}$ and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be independent random variables with $\boldsymbol{x}_i \in [-c, c]$ and $\mathbb{E}[\boldsymbol{x}_i] = 0$. For any $t > 0$,

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} \boldsymbol{x}_i\right| \geq t\right] \leq 2e^{-\frac{t^2}{2c}}.$$

Denote this event by $E$.

**Case 1.** Whenever $|P_0| \geq 2$, the hypothesis output is correct and has no loss.

**Case 2.** If $P_0 = P_1 = \{p\}$, then $\mathcal{A}(S) = c_p$ and $\mathrm{loss}_S(c_p) = 0$. Hence, given $E$, we have

$$\mathrm{loss}_{\mathcal{D}}(c_p) \leq \frac{\epsilon}{8}.$$

**Case 3.** If $|P_0| = 0$, then once again, the hypothesis output has no empirical loss, so given $E$,

$$\mathrm{loss}_{\mathcal{D}}(c_{-1}) \leq \frac{\epsilon}{8}.$$

**Case 4.** Finally, if $P_0 = \{p\}$ but $P_1 = \varnothing$, then given $E$, we have

$$\mathrm{loss}_{\mathcal{D}}(c_{-1}) \leq \mathrm{loss}_S(c_{-1}) + \frac{\epsilon}{8} \leq \frac{7\epsilon}{8} + \frac{\epsilon}{8} = \epsilon.$$

Next, we show the list replicability of $\mathcal{A}$.

**Claim 3.3.** *Given $E$, the algorithm $\mathcal{A}$ cannot output $c_p$ and $c_{p'}$ for two distinct points $p \neq p'$.*

*Proof.* Suppose that $\mathcal{A}$ outputs $c_p$ on some sample satisfying $E$. Then $P_0 = P_1 = \{p\}$, so $\mathrm{loss}_S(c_{-1}) = \hat{\mathcal{D}}(p) > 7\epsilon/8$ and $\mathrm{loss}_S(c_p) = 0$. By $E$,

$$\mathrm{loss}_{\mathcal{D}}(c_{-1}) > \frac{7\epsilon}{8} - \frac{\epsilon}{8} = \frac{3\epsilon}{4}, \qquad \mathrm{loss}_{\mathcal{D}}(c_p) \leq \frac{\epsilon}{8}.$$

Moreover, since $c_p$ is output, the point $p$ appears with label 1, so $p$ lies on any line realizing $\mathcal{D}$. Hence

$$\mathcal{D}(p) = \mathbb{P}_{(x,y)\sim\mathcal{D}}[x = p] = |\,\mathrm{loss}_{\mathcal{D}}(c_{-1}) - \mathrm{loss}_{\mathcal{D}}(c_p)| > \frac{5\epsilon}{8}.$$

Now let $S'$ be any sample satisfying $E$. Again using $E$ for the two hypotheses $c_{-1}$ and $c_p$,

$$\hat{\mathcal{D}}_{S'}(p) = |\,\mathrm{loss}_{S'}(c_{-1}) - \mathrm{loss}_{S'}(c_p)| \geq |\,\mathrm{loss}_{\mathcal{D}}(c_{-1}) - \mathrm{loss}_{\mathcal{D}}(c_p)| - \frac{2\epsilon}{8} > \frac{3\epsilon}{8}.$$

Thus, every sample satisfying $E$ contains at least one copy of $(p, 1)$.

Therefore, if both $c_p$ and $c_{p'}$ could be output on samples satisfying $E$, then every sample satisfying $E$ would contain both $(p, 1)$ and $(p', 1)$. But then $|P_0| \geq 2$, and the algorithm would output the unique line through $p$ and $p'$, not a point hypothesis. This contradiction proves the claim. □

By Claim 3.3, we see that when $E$ holds, $\mathcal{A}$ has population loss less than $\epsilon$, and outputs one of at most 3 different hypotheses. Therefore,

$$\textsc{lr}(\mathcal{H}_q) \leq 3.$$

□

## 4 A height-based list replicability algorithm

In [FHV26], Frick et al. used a combinatorial notion of the height of a simplicial complex to upper bound the index of a concept class. We show that this height is an upper bound for list replicability as well. Recall from Section 1.1.3 the definition of the junction closure $\mathcal{J}(\mathcal{C})$ of a partial concept class $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$:

$$\mathcal{J}(\mathcal{C}) = \left\{ \bigcap_{c \in S} c : \varnothing \neq S \subseteq \mathcal{C} \right\}$$

Then, the height $\mathrm{H}(\mathcal{C})$ measures the length of the longest inclusion chain in $\mathcal{J}(\mathcal{C})$. We restate Theorem 1.3 for completeness.

**Theorem 1.3** (Height controls list replicability)**.** *For every finite partial concept class* $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$*,*

$$\textsc{lr}(\mathcal{C}) \leq \mathrm{H}(\mathcal{C}).$$

Frick et al. applied their result to show that with high probability, random sign matrices have index $O(\log n)$.

**Theorem 4.1** (Random sign matrices have $O(\log n)$ height and index, [FHV26])**.** *Let* $A \in \{\pm 1\}^{N \times N}$ *be a sign matrix with entries sampled independently and uniformly at random.*

*Then, there exists some constant* $C > 0$ *such that with probability* $1 - o(1)$*,*

$$\mathrm{H}(A) \leq C \cdot \log N,$$

*and therefore*

$$\mathrm{Ind}_{\mathbb{Z}_2} \leq 2C \cdot \log N.$$

**Corollary 4.2** (Random sign matrices have $O(\log n)$ list replicability)**.** *For* $A \in \{\pm 1\}^{N \times N}$ *with entries sampled independently and uniformly at random, with probability* $1 - o(1)$*,*

$$\textsc{lr}(A) \leq C \cdot \log N$$

*as well.*

*Proof.* The corollary follows directly from Theorems 1.3 and 4.1. ☐

To prove Theorem 1.3, we give an algorithm that, with high probability, will only ever output $c$ contained within a single chain of $\mathcal{J}(\mathcal{C})$. Therefore, the list replicability of this algorithm will be naturally bounded by the height of $\mathcal{C}$.

The algorithm is fairly similar to that described in Theorem 3.2. We estimate the error of every partial hypothesis in $\mathcal{J}(\mathcal{C})$, and pick the smallest one that has "low error". In particular, our threshold for "low error" diminishes as the size of the hypothesis grows, so that we can guarantee that if any two partial hypotheses have low enough error, so does their junction. This way, we prevent any anti-chains in our output set.

*Proof of Theorem 1.3.* Let the domain $\mathcal{X}$ of $\mathcal{C}$ have size $|\mathcal{X}| = N$.

Since $\mathcal{J}(\mathcal{C})$ is finite, by Hoeffding's inequality, for every $\epsilon, \delta > 0$, there exists an $n(\epsilon, \delta)$ such that for any $\mathcal{D} \in \Delta_{\mathcal{C}}$, we can estimate the population loss of all hypotheses in $\mathcal{J}(\mathcal{C})$ simultaneously:

$$\mathbb{P}_{S \sim \mathcal{D}^n}\left[\sup_{c \in \mathcal{J}(\mathcal{C})} |\mathrm{loss}_S(c) - \mathrm{loss}_{\mathcal{D}}(c)| \leq \frac{\epsilon}{4^N}\right] \geq 1 - \delta. \tag{11}$$

We define the learning rule $\mathcal{A}$. For any $\epsilon, \delta$, let its sample size be $n(\epsilon, \delta)$ so as to satisfy (11). Denote by E the event that all losses are estimated within $\epsilon 4^{-N}$. Denote by $|c|$ the number of non-$\star$ points in the concept.

**Algorithm 2** The learning rule $\mathcal{A}$

```
Sample $S = (S_1, \dots, S_n) \sim \mathcal{D}^n$.
for $c \in \mathcal{J}(\mathcal{C})$, ordered from smallest to largest by size $|c|$ do
    if $\mathrm{loss}_S(c) \leq \epsilon \cdot 4^{-|c|}$ then
        Output $c$, or any completion of $c$.
    end if
end for
Output ERROR (the algorithm never reaches this state)
```

First of all, it is clear that $\mathcal{A}$ is a PAC learner. If E holds, then

$$\mathrm{loss}_{\mathcal{D}}(\mathcal{A}(S)) \leq \frac{\epsilon}{4^N} + \mathrm{loss}_S(\mathcal{A}(S)) \leq \frac{\epsilon}{4^N} + \frac{\epsilon}{4^{|\mathcal{A}(S)|}} \leq \epsilon.$$

Since $\mathcal{A}(S)$ can be a partial hypothesis, outputting any completion of it will never make the error grow.

**Claim 4.3.** *Let event E hold. Then, if $c_1$ and $c_2$ both satisfy the output requirement in line 3 of Algorithm 2, so does $c_1 \cap c_2$.*

*Proof.* Without loss of generality, if $c_1 \subseteq c_2$, then this holds trivially. Otherwise, $|c_1 \cap c_2| < \min\{|c_1|, |c_2|\}$.

Note that $c_1 \cap c_2$ is only correct on a sample if both $c_1$ and $c_2$ are. Thus,

$$\begin{aligned}
\mathrm{loss}_{\mathcal{D}}(c_1 \cap c_2) &= \mathbb{P}_{(x,y)\sim\mathcal{D}}[c_1(x) \neq y \text{ or } c_2(x) \neq y] \\
&\leq \mathbb{P}_{(x,y)\sim\mathcal{D}}[c_1(x) \neq y] + \mathbb{P}_{(x,y)\sim\mathcal{D}}[c_2(x) \neq y] \\
&\leq \mathrm{loss}_{\mathcal{D}}(c_1) + \mathrm{loss}_{\mathcal{D}}(c_2) \\
&\leq \mathrm{loss}_S(c_1) + \mathrm{loss}_S(c_2) + \frac{2\epsilon}{4^N} \\
&\leq \frac{\epsilon}{4^{|c_1|}} + \frac{\epsilon}{4^{|c_2|}} + \frac{2\epsilon}{4^N} \\
&\leq \frac{4\epsilon}{4^{\min\{|c_1|,|c_2|\}}} \\
&\leq \frac{\epsilon}{4^{|c_1 \cap c_2|}}
\end{aligned}$$

It is clear that $\mathcal{J}(\mathcal{C})$ is closed under taking junctions, so $c_1 \cap c_2 \in \mathcal{J}(\mathcal{C})$, and would be output first if both $c_1$ and $c_2$ fit the conditions. ☐

As a result of this claim, so long as E holds, no two concepts in an antichain will be output. Thus, so long as $\mathcal{A}$ doesn't output ERROR, when E holds, it will only output hypotheses from a single chain, so at most $\mathrm{H}(\mathcal{C})$ different hypotheses.

Finally, $\mathcal{A}$ will never output ERROR, for the distribution $\mathcal{D}$ comes from $\Delta_{\mathcal{C}}$, and thus some $c \in \mathcal{C} \subseteq \mathcal{J}(\mathcal{C})$ has 0 error on it. So if all else fails, that $c$ always fits the conditions to be output by $\mathcal{A}$. ☐

Notice that this algorithm can, instead of outputting partial hypotheses, output any completion desired. In particular, many of the partial hypotheses in $\mathcal{J}(\mathcal{C})$ may have common completions. Therefore, by using specific completions, better list replicability bounds may be obtained.

# 5 Height and eluder dimension

We now prove a more general form of Proposition 1.4, stated in the introduction, that the height parameter coincides with the eluder dimension up to an additive constant. We keep the notation from Section 1.1.3: $\mathcal{J}(\mathcal{C})$ is the junction closure of $\mathcal{C}$, and $g \preceq h$ means that $g$ is a restriction of $h$. For a partial concept $h$, write $\mathrm{supp}(h) := \{x \in \mathcal{X} : h(x) \neq \star\}$.

**Definition 5.1** (Height relative to a base concept)**.** *Fix a finite partial concept class $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ and a base concept $c' \in \mathcal{C}$. Define the height of $\mathcal{C}$ relative to $c'$ by*

$$\mathrm{H}(\mathcal{C}; c') := \sup\{m : \exists h_1, \ldots, h_m \in \mathcal{J}(\mathcal{C}) \text{ such that } h_1 \prec h_2 \prec \cdots \prec h_m = c'\}.$$

This rooted version recovers the height from the introduction:

$$\mathrm{H}(\mathcal{C}) = \sup_{c' \in \mathcal{C}} \mathrm{H}(\mathcal{C}; c').$$

Also, recall the definition of the eluder dimension. We extend the definition by [LKFS22] to partial concept classes.

**Definition 5.2** ([LKFS22])**.** *Fix a finite partial concept class* $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ *and a base concept* $c' \in \mathcal{C}$*. The eluder dimension of* $\mathcal{C}$ *relative to* $c'$*, denoted* $\mathrm{Edim}(\mathcal{C}, c')$*, is the largest* $m$ *such that there exist points* $x_1, \ldots, x_m \in \mathrm{supp}(c')$ *and concepts* $c_1, \ldots, c_m \in \mathcal{C}$ *such that, for every* $i \in [m]$*,*

$$c_i(x_i) \neq c'(x_i), \ (c_i(x_i) = \star \text{ allowed})$$

*while for every earlier point* $x_j$*,* $j < i$*,*

$$c_i(x_j) = c'(x_j).$$

*Define the eluder dimension of* $\mathcal{C}$ *by* $\mathrm{Edim}(C) \coloneqq \sup_{c' \in \mathcal{C}} \mathrm{Edim}(\mathcal{C}, c')$*.*

**Proposition 5.3** (Height equals eluder dimension)**.** *For every finite partial concept class* $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ *and every base concept* $c' \in \mathcal{C}$*,*

$$\mathrm{H}(\mathcal{C}, c') = \mathrm{Edim}(\mathcal{C}, c') + 1.$$

*In particular,*

$$\mathrm{H}(\mathcal{C}) = \mathrm{Edim}(\mathcal{C}) + 1.$$

*Proof.* We prove both inequalities.

First suppose $\mathrm{Edim}(\mathcal{C}, c') = m$. For $i = 0, 1, \ldots, m$, define

$$h_i = \bigcap \left(\{c'\} \cup \{c_{i+1}, c_{i+2}, \ldots, c_m\}\right),$$

and note that

$$h_0 \prec h_1 \prec \cdots \prec h_m = c'.$$

Therefore there is a strict chain of $m + 1$ partial concepts, so

$$\mathrm{H}(\mathcal{C}, c') \geq m + 1.$$

Conversely, suppose $\mathrm{H}(\mathcal{C}, c') = m$. Then there exist $h_1 \prec \cdots \prec h_m = c'$ in $\mathcal{J}(\mathcal{C})$. For each $i = 1, \ldots, m-1$, since $h_i \prec h_{i+1}$, choose a point $x_i \in X$ such that

$$h_i(x_i) = \star \qquad \text{and} \qquad h_{i+1}(x_i) \in \{\pm 1\}.$$

Because $h_{i+1} \preceq c'$, we have $h_{i+1}(x_i) = c'(x_i)$. Suppose $h_i = \bigcap_{c \in S_i} c$ for some $S_i \subseteq C$. There exists some $c_i \in S_i$ such that

$$c_i(x_i) \neq c'(x_i) \quad \text{and} \quad c_i(x_j) = c'(x_j) \quad \text{for all } j < i.$$

Thus the domain points $x_1, \ldots, x_{m-1}$ together with the concepts $c_1, \ldots, c_{m-1}$ form an eluder sequence with base concept $c'$. Therefore

$$\mathrm{Edim}(\mathcal{C}, c') \geq m - 1.$$

Combining the two inequalities,

$$\mathrm{H}(\mathcal{C}, c') = \mathrm{Edim}(\mathcal{C}, c') + 1.$$

□

## 6 Separation of coindex and list replicability

In this section, we discuss Corollary 1.11, which we restate below.

**Corollary 1.11.** *There exist partial concept classes* $\mathcal{C} \subseteq \{\pm 1, \star\}^n$ *with*

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}) = O(1) \qquad \text{and} \qquad \textsc{lr}(\mathcal{C}) = \omega(1).$$

This result follows from combining our Theorem 1.1 with a strategy of [FHV26]. We have shown that $\textsc{lr}(\mathcal{C})$ is bounded from below by $\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$, and it is know that the projective planes $\mathbb{RP}^{2N-1}$ are free $\mathbb{Z}_2$-spaces separating $\mathrm{Ind}_{\mathbb{Z}_2}$ and $\mathrm{coInd}_{\mathbb{Z}_2}$. To obtain that separation for $\mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C})$ and $\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C})$, we realize each $\mathbb{RP}^{2N-1}$ as the space of realizable distributions $\Delta_{\mathcal{C}^\pm}$ for some partial concept class $\mathcal{C}$.

The last step arises from an argument of Frick, Hosseini, and Vasileuski, which demonstrated that every finite free $\mathbb{Z}_2$-simplicial complex arises as the *sign complex* for some partial sign matrix [FHV26]. The sign complex is a simplicial complex associated with a partial sign matrix, and it parallels the role of the space of realizable distributions in our discussion of partial concept classes. The same proof in [FHV26] can be adapted to our setting:

**Lemma 6.1.** *Every finite free $\mathbb{Z}_2$-simplicial complex $\mathcal{K}$ is isomorphic to $\Delta_{\mathcal{C}^\pm}$ for a partial concept class $\mathcal{C}$.*

*Proof.* A fixed-point-free involution $\nu$ on a finite simplicial complex $\mathcal{K}$ pairs vertices of $\mathcal{K}$ as $(v, \nu(v))$. Since $\mathcal{K}$ is finite, we may index these pairs by $j \in [N]$. Likewise, the facets of $\mathcal{K}$ also come in pairs $(F, \nu(F))$, which we can index by $i \in [M]$.

Now define a partial concept class $\mathcal{C}_\mathcal{K} \subseteq \{+1, -1, \star\}^N$ of $M$ concepts given by

$$c_i(j) = \begin{cases} +1 & \text{if } v_j \in F_i \\ -1 & \text{if } \nu(v_j) \in F_i \\ \star & \text{otherwise.} \end{cases}$$

□

Moreover, nice enough $\mathbb{Z}_2$-free spaces can be given a $\mathbb{Z}_2$-free simplicial structure through *triangulation*. This is a classical result of Sören Illman.

**Theorem 6.2** (Illman's Theorem [Ill78])**.** *Every compact smooth free $\mathbb{Z}_2$-manifold admits a finite $\mathbb{Z}_2$-equivariant triangulation.*

Combining Lemma 6.1 and Theorem 6.2 allows us to import the crucial separating example $\mathbb{RP}^{2N-1}$.

*Proof of Corollary 1.11.* As stated in [Mat03, page 101] (also see [FHV26]), the projective space $\mathbb{RP}^{2N-1}$ can be equipped with the fixed-point-free involution induced by multiplication by $i$ on $\mathbb{C}^N$. Moreover, this gives the separation

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathbb{RP}^{2N-1}) = O(1) \qquad \text{and} \qquad \mathrm{Ind}_{\mathbb{Z}_2}(\mathbb{RP}^{2N-1}) = \omega(1).$$

By Theorem 6.2, there is an antipodal homeomorphism between the free $\mathbb{Z}_2$-space $\mathbb{RP}^{2N-1}$ and a simplicial complex $\mathcal{K}_N$. Since $\mathrm{Ind}_{\mathbb{Z}_2}$ and $\mathrm{coInd}_{\mathbb{Z}_2}$ are defined using topological properties of antipodal maps, it follows that $\mathcal{K}_N$ is also a separating example:

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{K}_N) = O(1) \qquad \text{and} \qquad \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{K}_N) = \omega(1).$$

Applying Lemma 6.1 yields a family of partial concept classes $\mathcal{C}_N$ with the property that $\Delta_{\mathcal{C}^\pm}$ is isomorphic to $\mathcal{K}_N$. Using once more that $\mathrm{Ind}_{\mathbb{Z}_2}$ and $\mathrm{coInd}_{\mathbb{Z}_2}$ are invariant under isomorphism completes the proof.

$$\mathrm{coInd}_{\mathbb{Z}_2}(\mathcal{C}_N) = O(1) \qquad \text{and} \qquad \mathrm{Ind}_{\mathbb{Z}_2}(\mathcal{C}_N) = \omega(1).$$

□

## 7 Composition properties for list replicability

In this section, we prove some properties of list replicability under joins and concatenations.

**Lemma 7.1** (Concatenation)**.** *Let* $\mathcal{C}_1, \mathcal{C}_2 \subseteq \{\pm 1\}^{\mathcal{X}}$ *be concept classes over the same finite domain* $\mathcal{X}$*. Then, for every* $\epsilon > 0$*,*

$$\textsc{lr}(\mathcal{C}_1 \cup \mathcal{C}_2, \epsilon) \leq \textsc{lr}(\mathcal{C}_1, \epsilon) + \textsc{lr}(\mathcal{C}_2, \epsilon).$$

*Consequently,*

$$\textsc{lr}(\mathcal{C}_1 \cup \mathcal{C}_2) \leq \textsc{lr}(\mathcal{C}_1) + \textsc{lr}(\mathcal{C}_2).$$

*Proof.* Fix $\varepsilon > 0$ and let $L_i := \textsc{lr}(\mathcal{C}_i, \varepsilon)$ for $i = 1, 2$.

By Theorem 2.11, for each $i \in \{1, 2\}$ there is a closed cover $\mathcal{F}_i = \{F^i_h : h \in \{\pm 1\}^{\mathcal{X}}\}$ of $\Delta_{\mathcal{C}_i}$ such that the overlap degree of $\mathcal{F}_i$ is at most $L_i$, and $F^i_h \subseteq B_\varepsilon(\sigma_h)$ for every $h \in \{\pm 1\}^{\mathcal{X}}$. Since $\Delta_{\mathcal{C}_1 \cup \mathcal{C}_2} = \Delta_{\mathcal{C}_1} \cup \Delta_{\mathcal{C}_2}$, define, for every hypothesis $h \in \{\pm 1\}^{\mathcal{X}}$, the closed set

$$F_h := F^1_h \cup F^2_h \subseteq \Delta_{\mathcal{C}_1 \cup \mathcal{C}_2}.$$

Then $\mathcal{F} := \{F_h : h \in \{\pm 1\}^{\mathcal{X}}\}$ is a closed cover of $\Delta_{\mathcal{C}_1 \cup \mathcal{C}_2}$ with overlap degree at most $L_1 + L_2$ satisfying $F_h \subseteq B_\varepsilon(\sigma_h)$,.

Again applying Theorem 2.11 gives

$$\textsc{lr}(\mathcal{C}_1 \cup \mathcal{C}_2, \epsilon) \leq L_1 + L_2.$$

Taking the supremum over $\epsilon > 0$ yields

$$\textsc{lr}(\mathcal{C}_1 \cup \mathcal{C}_2) \leq \textsc{lr}(\mathcal{C}_1) + \textsc{lr}(\mathcal{C}_2).$$

□

**Theorem 7.2** (Join of classes)**.** *Let* $\mathcal{C}_1 \subseteq \{\pm 1\}^{\mathcal{X}_1}$ *and* $\mathcal{C}_2 \subseteq \{\pm 1\}^{\mathcal{X}_2}$ *be concept classes over finite disjoint domains. Define*

$$\mathcal{C}_1 * \mathcal{C}_2 := \{c_1 \sqcup c_2 : c_1 \in \mathcal{C}_1,\ c_2 \in \mathcal{C}_2\} \subseteq \{\pm 1\}^{\mathcal{X}_1 \sqcup \mathcal{X}_2}.$$

*Then for every* $\varepsilon > 0$*,*

$$\textsc{lr}(\mathcal{C}_1 * \mathcal{C}_2, \varepsilon) \leq \textsc{lr}(\mathcal{C}_1, \varepsilon/4) + \textsc{lr}(\mathcal{C}_2, \varepsilon/4).$$

*Consequently,*

$$\textsc{lr}(\mathcal{C}_1 * \mathcal{C}_2) \leq \textsc{lr}(\mathcal{C}_1) + \textsc{lr}(\mathcal{C}_2).$$

*Proof.* Fix $\epsilon > 0$ and let $L_i := \textsc{lr}(\mathcal{C}_i, \varepsilon/4)$ for $i = 1, 2$. By the open-cover characterization of list replicability in Theorem 2.11, there are open covers

$$\mathcal{U} = \left\{U_h : h \in \{\pm 1\}^{\mathcal{X}_1}\right\} \qquad \text{and} \qquad \mathcal{V} = \left\{V_g : g \in \{\pm 1\}^{\mathcal{X}_2}\right\}$$

of $\Delta_{\mathcal{C}_1}$ and $\Delta_{\mathcal{C}_2}$, respectively, such that the overlap degree of $\mathcal{U}$ is at most $L_1$, the overlap degree of $\mathcal{V}$ is at most $L_2$, and $U_h \subseteq B_{\varepsilon/3}(\sigma_h)$, $V_g \subseteq B_{\varepsilon/3}(\sigma_g)$. In particular,

$$\mu_1 \in U_h \Rightarrow \mathrm{loss}_{\mu_1}(h) < \varepsilon/3, \qquad \mu_2 \in V_g \Rightarrow \mathrm{loss}_{\mu_2}(g) < \varepsilon/3. \tag{12}$$

Choose partitions of unity subordinate to these covers, $\{f^{(1)}_h\}_{h \in \{\pm 1\}^{\mathcal{X}_1}}$, $\{f^{(2)}_g\}_{g \in \{\pm 1\}^{\mathcal{X}_2}}$, so that $\operatorname{supp} f^{(1)}_h \subseteq U_h$, $\operatorname{supp} f^{(2)}_g \subseteq V_g$, and

$$\sum_{h \in \{\pm 1\}^{\mathcal{X}_1}} f^{(1)}_h \equiv 1, \qquad \sum_{g \in \{\pm 1\}^{\mathcal{X}_2}} f^{(2)}_g \equiv 1.$$

We construct an $(\varepsilon, L_1 + L_2)$-list-replicable learner for $\mathcal{C}_1 * \mathcal{C}_2$.

Choose $0 < \tau < \varepsilon/3$ and a continuous cutoff function $r : [0, 1] \to [0, 1]$ such that

$$r(s) = 0 \text{ for } s \leq \tau/2, \qquad r(s) = 1 \text{ for } s \geq \tau.$$

Fix arbitrary default hypotheses $h_0 \in \{\pm 1\}^{\mathcal{X}_1}$ and $g_0 \in \{\pm 1\}^{\mathcal{X}_2}$. Now choose arbitrary orderings $\{\pm 1\}^{\mathcal{X}_1} = \{h_1, \dots, h_M\}$, $\{\pm 1\}^{\mathcal{X}_2} = \{g_1, \dots, g_N\}$, where $M = 2^{|\mathcal{X}_1|}$ and $N = 2^{|\mathcal{X}_2|}$. The hypotheses $h_0$ and $g_0$ may repeat some $h_i$ or $g_j$; this causes no difficulty, since lists are sets of hypotheses and repetitions can only decrease their size.

Let $\mu \in \Delta_{\mathcal{C}_1 * \mathcal{C}_2}$. Write $t \coloneqq \mu(\mathcal{X}_1)$ and $1 - t = \mu(\mathcal{X}_2)$, and decompose

$$\mu = t\mu_1 + (1-t)\mu_2,$$

where $\mu_i \in \Delta_{\mathcal{C}_i}$ is the conditional distribution on $\mathcal{X}_i$ whenever the corresponding mass is nonzero.

Define probability vectors $p(\mu)$ on $\{0, 1, \dots, M\}$ and $q(\mu)$ on $\{0, 1, \dots, N\}$ by

$$p_0(\mu) \coloneqq 1 - r(t), \qquad p_i(\mu) \coloneqq r(t) f^{(1)}_{h_i}(\mu_1) \quad (1 \le i \le M),$$

and

$$q_0(\mu) \coloneqq 1 - r(1-t), \qquad q_j(\mu) \coloneqq r(1-t) f^{(2)}_{g_j}(\mu_2) \quad (1 \le j \le N).$$

Let

$$P_a(\mu) \coloneqq \sum_{i=0}^{a} p_i(\mu), \qquad Q_b(\mu) \coloneqq \sum_{j=0}^{b} q_j(\mu),$$

with $P_{-1}(\mu) = Q_{-1}(\mu) = 0$. Define intervals

$$I_a(\mu) \coloneqq [P_{a-1}(\mu), P_a(\mu)] \qquad (0 \le a \le M),$$

and

$$J_b(\mu) \coloneqq [Q_{b-1}(\mu), Q_b(\mu)] \qquad (0 \le b \le N).$$

Finally define

$$w_{a,b}(\mu) \coloneqq |I_a(\mu) \cap J_b(\mu)|.$$

In particular, every $w_{a,b}$ is continuous on $\Delta_{C_1 * C_2}$ and

$$\sum_{a=0}^{M} \sum_{b=0}^{N} w_{a,b}(\mu) = 1.$$

Define the good list for $\mu$ by

$$\operatorname{List}(\mu) \coloneqq \{h_a \sqcup g_b : w_{a,b}(\mu) > 0\},$$

where $h_a$ means the default $h_0$ when $a = 0$, and similarly $g_b$ means $g_0$ when $b = 0$.

**Claim 7.3.** *For every $\mu \in \Delta_{\mathcal{C}_1 * \mathcal{C}_2}$,*

$$|\operatorname{List}(\mu)| \le L_1 + L_2,$$

*and every hypothesis in* $\operatorname{List}(\mu)$ *has $\mu$-loss less than $\varepsilon$.*

*Proof.* Let

$$m(\mu) \coloneqq |\{a : p_a(\mu) > 0\}|, \qquad n(\mu) \coloneqq |\{b : q_b(\mu) > 0\}|.$$

For two interval partitions of $[0, 1]$ with $m(\mu)$ and $n(\mu)$ positive-length intervals, the number of positive-length intersections is at most

$$m(\mu) + n(\mu) - 1.$$

Indeed, sweeping from left to right, the active pair changes only when one crosses an endpoint of one of the two partitions.

If $t \ge \tau$, then $p_0(\mu) = 0$ and at most $L_1$ of the non-default $p_i(\mu)$ are positive. If $t \le \tau/2$, then $p_0(\mu) = 1$. If $\tau/2 < t < \tau$, then at most $L_1 + 1$ of the $p_i(\mu)$'s are positive. The same statements hold for $q(\mu)$ with

$1-t$ in place of $t$. Since $\tau < 1/3$, the two transition regimes $\tau/2 < t < \tau$ and $\tau/2 < 1-t < \tau$ cannot occur simultaneously. Hence

$$|\operatorname{List}(\mu)| \le m(\mu) + n(\mu) - 1 \le L_1 + L_2.$$

Now take any glued hypothesis $h_a \sqcup g_b$ in $\operatorname{List}(\mu)$. Then $p_a(\mu) > 0$ and $q_b(\mu) > 0$. If $a \neq 0$, then $f^{(1)}_{h_a}(\mu_1) > 0$, and therefore $\mu_1 \in U_{h_a}$; by (12), $\operatorname{loss}_{\mu_1}(h_a) < \varepsilon/3$. If $a = 0$, then $p_0(\mu) > 0$, so $r(t) < 1$, and hence $t < \tau < \varepsilon/3$.

Similarly, if $b \neq 0$, then $\operatorname{loss}_{\mu_2}(g_b) < \varepsilon/3$, while if $b = 0$, then $1 - t < \tau$. The case $a = b = 0$ is impossible, since it would imply $t < \tau$ and $1 - t < \tau < \varepsilon/3$. Therefore

$$\operatorname{loss}_{\mu}(h_a \sqcup g_b) = t \operatorname{loss}_{\mu_1}(h_a) + (1-t) \operatorname{loss}_{\mu_2}(g_b) < \varepsilon/3 + \varepsilon/3 < \varepsilon.$$

This proves the claim. ☐

We now define the learning rule.

**Algorithm 3** The join learning rule $\mathcal{A}$

```
Sample $S = (S_1, \dots, S_n) \sim \mu^n$.
Construct the empirical distribution $\widehat{\mu}$ on $(\mathcal{X}_1 \sqcup \mathcal{X}_2) \times \{\pm 1\}$.
Compute the weights $w_{a,b}(\widehat{\mu})$ for all $0 \le a \le M$ and $0 \le b \le N$.
Output $h_a \sqcup g_b$ with probability $w_{a,b}(\widehat{\mu})$.
```

It remains to show that for sufficiently large $n$, the output belongs to $\operatorname{List}(\mu)$ with probability at least $1 - \delta$, uniformly over $\mu$.

For a fixed $\mu \in \Delta_{\mathcal{C}_1 * \mathcal{C}_2}$ define

$$F_\mu(\nu) \coloneqq \sum_{\substack{0 \le a \le M,\ 0 \le b \le N: \\ h_a \sqcup g_b \notin \operatorname{List}(\mu)}} w_{a,b}(\nu).$$

This is the probability that the algorithm, when run with empirical distribution $\nu$, outputs a hypothesis outside $\operatorname{List}(\mu)$. Since the sum is finite and the functions $w_{a,b}$ are continuous, $F_\mu$ is continuous. Moreover,

$$F_\mu(\mu) = 0.$$

The compactness of $\Delta_{\mathcal{C}_1 * \mathcal{C}_2}$ implies the finite family of functions $w_{a,b}$ is uniformly equicontinuous. More explicitly, choose $\rho > 0$ so that each $w_{a,b}$ changes by at most $\delta/(2R)$ on $\rho$-balls, where $R = (M+1)(N+1)$. Thus this $\rho$ is independent of $\mu$, and whenever $d_{\mathtt{TV}}(\nu, \mu) < \rho$, we have

$$F_\mu(\nu) < \delta/2.$$

Since $\mathcal{X}_1 \sqcup \mathcal{X}_2$ is finite, standard uniform convergence of empirical distributions gives an $n_0 = n_0(\rho, \delta, \mathcal{X}_1, \mathcal{X}_2)$ such that for every realizable $\mu$ and every $n \ge n_0$,

$$\mathbb{P}_{S \sim \mu^n}\left[d_{\mathtt{TV}}(\widehat{\mu}, \mu) < \rho\right] \ge 1 - \delta/2.$$

Consequently, for every $\mu \in \Delta_{\mathcal{C}_1 * \mathcal{C}_2}$ and $n \ge n_0$,

$$\begin{aligned}\mathbb{P}_{S \sim \mu^n}\left[\mathcal{A}(S) \notin \operatorname{List}(\mu)\right] &\le \mathbb{P}\left[d_{\mathtt{TV}}(\widehat{\mu}, \mu) \ge \rho\right] + \mathbb{E}\left[F_\mu(\widehat{\mu}) \mathbf{1}_{\{d_{\mathtt{TV}}(\widehat{\mu}, \mu) < \rho\}}\right] \\ &\le \delta/2 + \delta/2 = \delta.\end{aligned}$$

Together with Claim 7.3, this shows that $\mathcal{A}$ is an $(\varepsilon, L_1 + L_2)$-list-replicable learner for $\mathcal{C}_1 * \mathcal{C}_2$. Therefore

$$\operatorname{LR}(\mathcal{C}_1 * \mathcal{C}_2, \varepsilon) \le L_1 + L_2 = \operatorname{LR}(\mathcal{C}_1, \varepsilon/3) + \operatorname{LR}(\mathcal{C}_2, \varepsilon/3).$$

Taking the supremum over $\varepsilon > 0$ proves the final assertion. ☐

# References

[ABL+22] 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.

[AFR85] Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science, FOCS 1985*, pages 277–280. IEEE Computer Society, 1985.

[AMY16] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus VC dimension. In *Conference on Learning Theory*, pages 47–80. PMLR, 2016.

[AO07] Peter Auer and Ronald Ortner. A new PAC bound for intersection-closed concept classes. *Machine Learning*, 66(2):151–163, 2007.

[Atm22] Aistis Atminas. Classes of graphs without star forests and related graphs. *Discrete Mathematics*, 345(12):113089, 2022.

[BDE98] Shai Ben-David and Nadav Eiron. Self-directed learning and its relation to the VC-dimension and to teacher-directed learning. *Machine Learning*, 33(1):87–104, 1998.

[BGH+23] 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 *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 520–527, 2023.

[BGHH25] Ari Blondal, Shan Gao, Hamed Hatami, and Pooya Hatami. Stability and list-replicability for agnostic learners. In *Conference on Learning Theory*, pages 380—400. PMLR, 2025.

[BHH+26a] Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Borsuk-Ulam and replicable learning of large-margin halfspaces. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pages 529–540, 2026.

[BHH+26b] Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Simplicial Covering Dimension of Extremal Concept Classes. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:24, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[BHH+26c] Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Tight list replicability bounds via a novel sphere covering theorem, 2026. `arXiv:2606.06148`.

[BHQS21] Avrim Blum, Steve Hanneke, Jian Qian, and Han Shao. Robust learning under clean-label attack. In *Conference on Learning Theory*, pages 591–634. PMLR, 2021.

[BLM20] Mark Bun, Roi Livni, and Shay Moran. An equivalence between private classification and online prediction. In *IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402, 2020.

[CCH+24] 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, 2024.

[CCMW22] 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.

[CCMY24] Zachary Chase, Bogdan Chornomaz, Shay Moran, and Amir Yehudayoff. Local Borsuk-Ulam, stability, and replicability. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1769–1780, New York, NY, USA, 2024. Association for Computing Machinery.

[CLSW04] Péter Csorba, Carsten Lange, Ingo Schurr, and Arnold Wassmer. Box complexes, neighborhood complexes, and the chromatic number. *Journal of Combinatorial Theory, Series A*, 108(1):159–168, 2004.

[CMW25] Bogdan Chornomaz, Shay Moran, and Tom Waknine. Spherical dimension. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 1259–1313. PMLR, 30 Jun–04 Jul 2025.

[CMY23] Zachary Chase, Shay Moran, and Amir Yehudayoff. Stability and Replicability in Learning. In *IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2430–2439. IEEE Computer Society, 2023.

[Dar15] Malte Darnstädt. The optimal pac bound for intersection-closed concept classes. *Information Processing Letters*, 115(4):458–461, 2015.

[DJ03] Victor Dalmau and Peter Jeavons. Learnability of quantified formulas. *Theoretical Computer Science*, 306(1-3):485–511, 2003.

[DPVWV23] Peter Dixon, A. Pavan, Jason Vander Woude, and N. V. Vinodchandran. List and certificate complexities in replicable learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NeurIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

[EHKS23] Eric Eaton, Marcel Hussing, Michael Kearns, and Jessica Sorrell. Replicable reinforcement learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NeurIPS '23. Curran Associates Inc., 2023.

[EKK$^+$23] Hossein Esfandiari, Alkis Kalavasis, Amin Karbasi, Andreas Krause, Vahab Mirrokni, and Grigoris Velegkas. Replicable bandits. In *The Eleventh International Conference on Learning Representations*, 2023.

[EKM$^+$23] 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.

[FHV26] Florian Frick, Kaave Hosseini, and Aliaksei Vasileuski. A $\mathbb{Z}_2$-topological framework for sign-rank lower bounds, 2026. `arXiv:2604.01510`.

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. System Sci.*, 65(4):612–625, 2002. Special issue on complexity, 2001 (Chicago, IL).

[FW95] Sally Floyd and Manfred Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine learning*, 21(3):269–304, 1995.

[Han16] Steve Hanneke. Refined error bounds for several learning algorithms. *Journal of Machine Learning Research*, 17(135):1–55, 2016.

[Han24] Steve Hanneke. The star number and eluder dimension: Elementary observations about the dimensions of disagreement. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 2308–2359. PMLR, 2024.

[HHP$^+$22] Hamed Hatami, Pooya Hatami, William Pires, Ran Tao, and Rosie Zhao. Lower bound methods for sign-rank and their limitations. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, pages 22–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2022.

[HK07] Pinar Heggernes and Dieter Kratsch. Linear-time certifying recognition algorithms and forbidden induced subgraphs. *Nordic J. Comput.*, 14(1-2):87–108, 2007.

[HLW94] David Haussler, Nick Littlestone, and Manfred K Warmuth. Predicting {0, 1}-functions on randomly drawn points. *Information and Computation*, 115(2):248–292, 1994.

[HSW90] David Helmbold, Robert Sloan, and Manfred K Warmuth. Learning nested differences of intersection-closed concept classes. *Machine Learning*, 5(2):165–196, 1990.

[HY15] Steve Hanneke and Liu Yang. Minimax analysis of active learning. *The Journal of Machine Learning Research*, 16(1):3487–3602, 2015.

[Ill78] Sören Illman. Smooth equivariant triangulations of G-manifolds for G a finite group. *Mathematische Annalen*, 233(3):199–220, October 1978.

[KKL$^+$24] Alkis Kalavasis, Amin Karbasi, Kasper Green Larsen, Grigoris Velegkas, and Felix Zhou. Replicable learning of large-margin halfspaces. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

[KKMV23] Alkis Kalavasis, Amin Karbasi, Shay Moran, and Grigoris Velegkas. Statistical indistinguishability of learning algorithms. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

[Kuh99] Christian Kuhlmann. On teaching and learning intersection-closed concept classes. In *European Conference on Computational Learning Theory*, pages 168–182. Springer, 1999.

[KVYZ23] Amin Karbasi, Grigoris Velegkas, Lin Yang, and Felix Zhou. Replicability in reinforcement learning. *Advances in Neural Information Processing Systems*, 36:74702–74735, 2023.

[LKFS22] Gene Li, Pritish Kamath, Dylan J Foster, and Nati Srebro. Understanding the eluder dimension. *Advances in Neural Information Processing Systems*, 35:23737–23750, 2022.

[Lov78] László Lovász. Kneser's conjecture, chromatic number, and homotopy. *Journal of Combinatorial Theory, Series A*, 25(3):319–324, 1978.

[Mat03] Jiří Matoušek. *Using the Borsuk-Ulam Theorem: Lectures on Topological Methods in Combinatorics and Geometry.* Universitext. Springer, Berlin/Heidelberg, 2003.

[Mil64] J. Milnor. On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[MM22] Maryanthe Malliaris and Shay Moran. The unstable formula theorem revisited via algorithms. *arXiv preprint arXiv:2212.05050*, 2022.

[MP95] N. V. R. Mahadev and U. N. Peled. *Threshold graphs and related topics*, volume 56 of *Annals of Discrete Mathematics*. North-Holland Publishing Co., Amsterdam, 1995.

[MSS23] Shay Moran, Hilla Schefler, and Jonathan Shafer. The bayesian stability zoo. *Advances in Neural Information Processing Systems*, 36:61725–61746, 2023.

[MW16] Shay Moran and Manfred K Warmuth. Labeled compression schemes for extremal classes. In *International Conference on Algorithmic Learning Theory*, pages 34–49. Springer, 2016.

[MZ02] Jiri Matousek and Günter M Ziegler. Topological lower bounds for the chromatic number: A hierarchy. *arXiv preprint math/0208072*, 2002.

[PS86] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33(1):106–123, 1986.

[RR22] Joachim Rubinstein and Benjamin Rubinstein. Unlabelled sample compression schemes for intersection-closed classes and extremal classes. *Advances in Neural Information Processing Systems*, 35:13078–13090, 2022.

[Rud87] Walter Rudin. *Real and complex analysis.* McGraw-Hill, Inc., 1987.

[RVR13] Daniel Russo and Benjamin Van Roy. Eluder dimension and the sample complexity of optimistic exploration. *Advances in Neural Information Processing Systems*, 26, 2013.

[ST06] Gábor Simonyi and Gábor Tardos. Local chromatic number, Ky Fan's theorem, and circular colorings. *Combinatorica*, 26(5):587–626, 2006.

[STV09] Gábor Simonyi, Gábor Tardos, and Siniša Vrećica. Local chromatic number and distinguishing the strength of topological obstructions. *Transactions of the American Mathematical Society*, 361(2):889–908, 2009.

[Tho65] René Thom. Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 255–265. Princeton Univ. Press, Princeton, N.J., 1965.

[War68] Hugh E. Warren. Lower bounds for approximation by nonlinear manifolds. *Trans. Amer. Math. Soc.*, 133:167–178, 1968.