

# Spherical dimension

**Bogdan Chornomaz**

*Department of Mathematics, Technion*

MARKYZ.KARABAS@GMAIL.COM

**Shay Moran**

*Departments of Mathematics, Computer Science, and Data and Decision Sciences, Technion*

SMORAN@TECHNION.AC.IL

*Google Research*

**Tom Waknine**

*Department of Mathematics, Technion*

TOM.WAKNINE@CAMPUS.TECHNION.AC.IL

## Abstract

We introduce and study the *spherical dimension*, a natural topological relaxation of the VC dimension that unifies several results in learning theory where topology plays a key role in the proofs. The spherical dimension is defined by extending the set of realizable datasets (used to define the VC dimension) to the continuous space of realizable distributions. In this space, a shattered set of size  $d$  (in the VC sense) is completed into a continuous object, specifically a  $d$ -dimensional sphere of realizable distributions. The spherical dimension is then defined as the dimension of the largest sphere in this space. Thus, the spherical dimension is at least the VC dimension.

The spherical dimension serves as a common foundation for leveraging the Borsuk-Ulam theorem and related topological tools. We demonstrate the utility of the spherical dimension in diverse applications, including disambiguations of partial concept classes, reductions from classification to stochastic convex optimization, stability and replicability, and sample compression schemes. Perhaps surprisingly, we show that the open question posed by Alon, Hanneke, Holzman, and Moran (FOCS 2021) of whether there exist non-trivial disambiguations for halfspaces with margin is equivalent to the basic open question of whether the VC and spherical dimensions are finite together.

**Keywords:** PAC Learning, VC Dimension, Borsuk-Ulam Theorem, Partial Concept Classes, Linear Classifiers with Margin, Extremal Classes, List Replicability, Stochastic Convex Optimization.

## 1. Introduction

The VC dimension is one of the most celebrated parameters in the theory of PAC learning, offering a simple yet powerful combinatorial framework for understanding the complexity of concept classes. Formally, the VC dimension of a binary-labeled concept class  $\mathcal{H}$  is the largest integer  $n$  such that there exist points  $x_1, \dots, x_n \in X$  where, for any labeling  $y_1, \dots, y_n \in \{\pm 1\}$ , the dataset  $(x_1, y_1), \dots, (x_n, y_n)$  is realizable by  $\mathcal{H}$ .

In this work, we study a continuous relaxation of the VC dimension. To define it, we first extend the discrete space of realizable datasets to the continuous space of realizable distributions. In this space, a shattered set corresponds to an embedding of an  $\ell_1$  sphere (see Figure 1 below). This observation motivates the definition of the spherical dimension as the maximum dimension of a sphere within the space of realizable distributions.

We will later demonstrate that this topological relaxation of the VC dimension has several disparate applications in learning theory, ranging from sample compression schemes to reductions from classification to optimization. But first, we present the formal definition. Let  $\mathcal{H}$  be a class

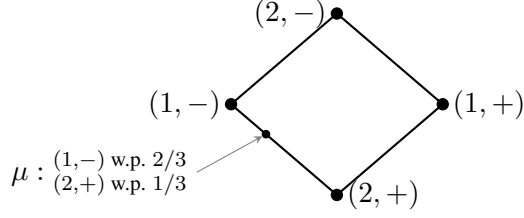


Figure 1: Space of realizable distributions of a four-concept class shattering two points.

of functions from domain  $X$  to label space  $Y = \{\pm 1\}$ , and let  $\Delta = \Delta(\mathcal{H})$  be the space of all possible distributions on  $X \times Y$ , realizable by  $\mathcal{H}$ . We equip  $\Delta$  with a *total variation distance*  $\text{TV}$  between the distributions, turning it into a metric space. We then define a natural “flipping the labels” map  $\rho \mapsto -\rho$  on the space of all distributions over  $X \times Y$ , that is, for any event  $E \subseteq X \times Y$ ,  $\mathbb{P}_{-\rho}(E) = \mathbb{P}_{\rho}(-E)$ , where  $-E = \{(x, -y) \mid (x, y) \in E\}$ . While  $\Delta$  is not necessarily closed under this map, we are concerned with a part of it that is. We thus define  $\Delta^{\text{ant}} = \Delta \cap -\Delta$ , on which this map becomes a continuous (with respect to  $\text{TV}$ ) fixed-point free involution<sup>1</sup>. Then, our main definition can be stated as:

**Definition 1 (Spherical dimension)** *We say that a finite class  $\mathcal{H}$  admits an  $n$ -sphere if there is a continuous map  $f: \mathbb{S}^n \rightarrow \Delta_{\mathcal{H}}^{\text{ant}}$  such that  $f$  respects the antipodalities on  $\mathbb{S}^n$  and  $\Delta_{\mathcal{H}}^{\text{ant}}$ , that is,  $f(-v) = -f(v)$  for all  $v \in \mathbb{S}^n$ . The spherical dimension  $\text{sd}(\mathcal{H})$  of  $\mathcal{H}$  is then defined as a maximal integer  $n$  such that  $\mathcal{H}$  admits an  $n$ -sphere.*

The restriction to finite classes here is for technical reasons. In Section 2 we give the full version, as well as certain gradations of this definition. All in all, due to space restrictions, it will be our usual policy to present the claims in the main part of the paper in a simplified, although formally correct, fashion, directing for the full statements to appendices.

The following lemma, extending the results in Chase et al. (2024b), enables us to illustrate the spherical dimension on some concrete easy classes<sup>2</sup>.

**Lemma 2** *For a class  $\mathcal{H}$ ,  $\text{sd}(\mathcal{H}) \geq \text{VC}(\mathcal{H}) - 1$  and  $\text{sd}(\mathcal{H}) \geq \text{VC}^{*\text{a}}(\mathcal{H}) - 2$ . Moreover, for the classes  $\mathcal{C}_n$  of binary hypercubes and universal classes  $\mathcal{U}_n$  (minimal classes witnessing  $\text{VC}$  and  $\text{VC}^*$ ),  $\text{sd}(\mathcal{C}_n) = n - 1$  and  $n - 2 \leq \text{sd}(\mathcal{U}_n) \leq n - 1$ .*

Recall that  $\mathcal{C}_n$  is a class of all  $2^n$  functions on  $[n]$ , and  $\mathcal{U}_n$  is a class of  $n$  functions on the set of  $2^n$  points, parametrized by subsets of  $[n]$ , witnessing all possible combinations of their values. Figure 2 below show the spheres witnessing  $\text{sd}(\mathcal{C}_n) \geq n - 1$  and  $\text{sd}(\mathcal{U}_n) \geq n - 2$ , for  $n = 3$ . As we will elaborate in Sections A.5 and A.6, they correspond to the boundary of an  $n$ -crosspolytope and the barycentric subdivision of the boundary of an  $(n - 1)$ -simplex.

We also know how to construct classes of spherical dimension asymptotically larger than both  $\text{VC}$  and  $\text{VC}^*$ , using *products of classes*:

1. Recall that an involution is a function  $f$  such that  $f(f(x)) = x$  for all  $x$ . A fixed-point free involution is an involution for which  $f(x) \neq x$  for all  $x$ .
2. In order to be technically correct, below and generally in the introduction we use the dual antipodal VC dimension,  $\text{VC}^{*\text{a}}$ , which we will only formally introduce in Section A.1. For the sake of understanding, it is safe to ignore this nuance and think of it as simply of  $\text{VC}^*$ .

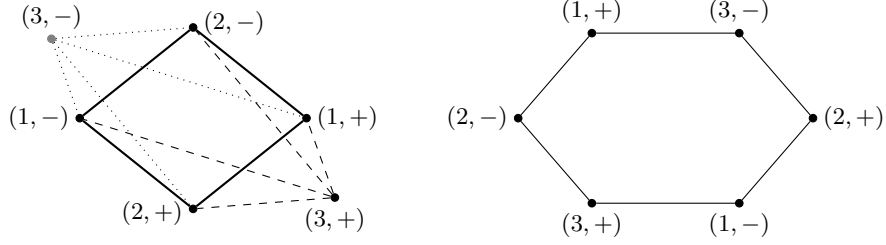


Figure 2: Spheres witnessing  $\text{sd}(\mathcal{C}_3) \geq 2$  (left) and  $\text{sd}(\mathcal{U}_3) \geq 1$  (right).

**Lemma 3** For classes  $\mathcal{H}_1$  and  $\mathcal{H}_2$ ,  $\text{VC}(\mathcal{H}_1 \times \mathcal{H}_2) = \text{VC}(\mathcal{H}_1) + \text{VC}(\mathcal{H}_2)$  and  $\text{sd}(\mathcal{H}_1 \times \mathcal{H}_2) \geq \text{sd}(\mathcal{H}_1) + \text{sd}(\mathcal{H}_2) + 1$ . In particular, for an  $m$ 'th direct power  $\mathcal{U}_n^m$  of the universal class  $\mathcal{U}_n$ , we have  $\text{VC}(\mathcal{U}_n^m) = m \lfloor \log n \rfloor$ ,  $n \leq \text{VC}^*(\mathcal{U}_n^m) \leq n + \lfloor \log m \rfloor$ , and  $m(n-1) - 1 \leq \text{sd}(\mathcal{U}_n^m) \leq mn$ .

In particular, for the class  $\mathcal{U}_n^n$ ,  $\text{VC}(\mathcal{U}_n^n) \sim \text{VC}^*(\mathcal{U}_n^n) \sim n$  (up to a log factor) and  $\text{sd}(\mathcal{U}_n^n) \gtrsim n^2$ .

In the proofs Lemmas 2 and 3 in Section B, we explicate some values for the classes  $\mathcal{C}_n$  and  $\mathcal{U}_n$ , and also give, in Lemma 31 in Section B.3, a fine-tuned classification of low-VC classes via spherical dimension.

Let us now discuss the utility of the spherical dimension.

### How we can use the spherical dimension

As of now, we know of four, rather diverse, applications of the spherical dimension in learning: (i) disambiguations of linear classifiers with margin, (ii) embedding into maximum/extremal classes, (iii) global stability and list replicability, and (iv) reductions to stochastic convex optimization (SCO). Here we outline these areas and applications of spherical dimension in them and push the technical discussion to Sections 3–6 (in the same order).

We note that, typically,  $\text{sd}$  is used to lower-bound the parameter of interest via an application of the Borsuk-Ulam (BU) theorem, or one of its relatives. The bounds that this application yields are usually linear, sometimes with a factor of  $1/2$ , and in many cases they are known to be sharp. We also note that while in the last two applications (list replicability and SCO) the connection to spherical dimension is rather straightforward (modulo published results and with some amount of technical work), the connection in the other two cases is, to our knowledge, either entirely new (disambiguations), or constitute substantial progress in a previously outlined direction (extremal classes).

### DISAMBIGUATIONS OF LINEAR CLASSIFIERS WITH MARGIN

In Alon et al. (2022) the authors initiated a systematic study of PAC learnability for partial concept classes<sup>3</sup>. Their motivation for studying this setup was that it naturally captures some practical data-dependent assumptions, not covered by the standard PAC learning model. For example that the data is located on a low-dimensional subspace, or is separable with a margin (such as in the classical Perceptron algorithm (Rosenblatt, 1958)).

While the authors did establish that, just as for the PAC learnability, the learnability of partial concept classes is equivalent to the finite VC dimension, they also note a range of stark differences

3. Although the first, to our knowledge, paper where this learning setting was considered is Long (2001).

specific to this setup. For example, while in the classical setting the learning can always be done by an *empirical risk minimization* (ERM) algorithm, it is provably not so for partial concept classes. Moreover, there are easily learnable partial classes for which the class of hypotheses returned by the learner necessarily has an infinite VC dimension.

The central notion in Alon et al. (2022) is a *disambiguation* of a partial class  $H$ : this is a class  $H'$  of total functions such that each partial function in  $H$  is extended by some function in  $H'$ . A natural question to ask here is whether for a given partial class there is a disambiguation of comparable VC dimension. However, the authors proved that it is, in a strong sense, false: There is a partial class of  $VC = 1$  such that any of its disambiguation has an infinite VC dimension. Effectively, the existence of such a class is the reason for the failure of ERM, along with some other impossibility results, for partial classes. This construction, however, relies on a recent breakthrough in graph theory and communication complexity (Göös (2015); Ben-David et al. (2016); Balodis et al. (2023)) and provides little insight into the structure of such a class. They conjectured, however (Open Question 21), that the partial class of linear classifiers with margin, the very same class as used in connection to the Perceptron model, can give such a construction: for large margin, its VC dimension is trivially 1 (independently of the euclidean dimension!), and for all “natural” disambiguations the VC grows with the dimension of the underlying space.

In this paper, we prove that modulo some technical details, their OQ 21: whether the class of linear classifiers with margin, for some nontrivial margin, can be disambiguated with a class of bounded VC dimension, is equivalent to asking if  $sd$  is bounded by a function of VC dimension (that is, if  $sd$  and  $VC$  go to infinity together). Moreover, our connection is not just qualitative, but quantitative, and, in fact, very precise:

**Theorem 4 (Rough equivalence between disambiguations and spherical dimension)** *For integers  $a$ ,  $b$ , and  $n$ , the following are equivalent<sup>4</sup>:*

- *There is a class with  $VC \leq a$ ,  $VC^{*a} \leq b$ , and  $sd_{\text{simp}} \geq n$ ;*
- *There is a disambiguation of a class of linear classifiers of an  $n$ -dimensional unit sphere with some nontrivial margin, for which  $VC \leq a$  and  $VC^{*a} \leq b$ .*

#### EMBEDDINGS INTO EXTREMAL CLASSES

When one has to study a complicated object, a very natural and common sense idea is to transform it, without losing too much precision, into a nice structured object. In mathematics, it is manifested, for example, in approximating smooth functions with polynomials, considering a convex hull of a set instead of the set itself, you name it.

In learning theory, especially from the perspective of one of the most long-standing open problems, the *sample compression schemes conjecture* Floyd and Warmuth (1995); Warmuth (2003), good candidates for such “nice structured objects” are *extremal classes*<sup>5</sup>. An unavoidably incomplete list of works along the lines of resolving sample compression conjecture by embedding a low-VC class into an extremal class of comparable VC dimension includes Ben-David and Litman

4. Similarly to how we used  $VC^{*a}$  in Lemma 2 above, here we use simplicial spherical dimension  $sd_{\text{simp}}$ , without properly introducing it, for formal correctness, and suggest treating simply as  $sd$ .

5. Extremal, or shattering-extremal, classes are those that meet Pajor’s inequality with equality; they generalize a perhaps more known family of *maximum* classes that reach equality in Sauer-Shelah-Perles bound, see Section 4 for technical discussion.

(1998); Kuzmin and Warmuth (2007); Rubinstein et al. (2009); Rubinstein and Rubinstein (2012); Rubinstein et al. (2015); Moran and Warmuth (2016); Chepoi et al. (2020, 2021, 2022); Chalopin et al. (2022); Rubinstein and Rubinstein (2022); Chalopin et al. (2023). While the aimed at “comparable VC dimension” would usually be linearly bounded, anything subexponential would also be good, at least in the sense of beating the best known bound for the sample compression in Moran and Yehudayoff (2016).

While for certain natural families of classes such embedding can be done (for example Rubinstein and Rubinstein (2022) constructed such embeddings of linearly bounded VC for intersection-closed classes), in recent work (Chase et al., 2024a) the authors proved that certain classes cannot be embedded into extremal without at least an exponential blowup in the VC dimension. Their main structural result, leading to this conclusion, was that extremal classes satisfy the inequality  $VC^* \leq 2VC + 1$  (contrasting with the general exponential upper bound on  $VC^*$  in terms of VC, Assouad (1983)). The proof of this bound in Chase et al. (2024a) was topological, utilizing the properties of *cubical complexes* of extremal classes, as well as a generalization of the topological Radon theorem. Here we prove that this bound, again, for extremal classes, can be strengthened in terms of spherical dimension to  $sd \leq 2VC - 1$ . Furthermore, this bound is a consequence of a strong topological similarity between the space of realizable distributions and the cubical complex of an extremal class, which, structurally, is our main result in this direction. This is best illustrated by picture, such as Figure 14 in Section D; however, a fair amount of technical work needs to be done in order to give meaning to it.

Curiously enough, this result, together with a well-known fact that every  $VC = 1$  class can be extended to an extremal  $VC = 1$  class, when applied to the setup of disambiguations, proves that any class that disambiguates classifiers with a margin on a 2-dimensional sphere should have VC dimension at least 2.

We finally note that, apart from the above connection to the sample compression schemes conjecture, extremal classes found to be useful in other problems in learning theory, such as proper optimal learners (Bousquet et al., 2020), and also in other areas of mathematics such as extremal combinatorics (Bollobás et al., 1989; Bollobás and Radcliffe, 1995), discrete geometry (Lawrence, 1983) and functional analysis (Pajor, 1985) (on many occasions, these classes were independently rediscovered, see Moran (2012)).

## GLOBAL STABILITY AND LIST REPLICABILITY

The study of replicability in PAC learning was initiated in the work of Impagliazzo et al. (2022) to answer, on the theory side, the growing demand for reproducibility in science (Baker, 2016). Their definition of replicability, however, assumed shared internal randomness between the runs of the learner. As observed in Dixon et al. (2023), without this assumption replicability becomes essentially non-achievable and, to overcome this obstruction, they introduced *list replicability*, on which we concentrate in this paper. Concurrently, Chase et al. (2023) proved an equivalence (by reciprocity) of list replicability with *global stability*, introduced in Bun et al. (2020), where it was an essential ingredient in the proof of establishing the differentially-private (DP) learnability of online-learnable classes (together with the earlier results in Alon et al. (2019), this established an equivalence between DP and online learnability). We also note that, apart from an independent interest and the above connection to DP, the notions of stability and replicability also have deep ties to pseudo-determinism (Gat and Goldwasser, 2011). We refer to Kalavasis et al. (2024) for a further review of the field, with a focus on computational efficiency.

It was then observed in [Chase et al. \(2024b\)](#) (with a precursor in [Chase et al. \(2023\)](#)) that for finite classes the list replicability number  $\text{LR}$  has a characterization in terms of overlap of certain closed covers of the space of realizable distributions. Further down this road, they showed that  $\text{LR}$  can be linearly lower-bounded by both VC and dual VC ( $\text{VC}^*$ ) dimensions, even for weak learners. In this paper, we show that both these bounds are subsumed by the bound  $\text{sd} \leq 2\text{LR} - 3$ , which, similarly, holds even for weak learners. We note, however, that we consider [Chase et al. \(2024b\)](#) as a crucial stepping stone for our present setup.

We further note that the same topological approach was used in [Chase et al. \(2024b\)](#) to show that, surprisingly, *agnostic* replicable learning is impossible (unless in trivial cases). A recent work ([Blonda et al., 2025](#)) also elaborates on some relaxations of the notion of replicability in the agnostic setting.

## REDUCTIONS TO STOCHASTIC CONVEX OPTIMIZATION

While, from the statistical perspective, PAC learnability is fully characterized by the VC dimension, the things are getting much more complicated once we also consider a computational aspect of learning (see, for example, Chapter 8 in [Shalev-Shwartz and Ben-David \(2014\)](#)). One reasonable way of dealing with it is by *reducing* the learning problem of interest to the one that can be efficiently solved ([Pitt and Warmuth, 1990](#)). This venue was recently explored in [Chornomaz et al. \(2024\)](#), where the authors focused on reductions to learning with halfspaces ([Forster et al., 2003](#); [Ben-David et al., 2002](#); [Linial and Shraibman, 2009](#); [Kamath et al., 2020](#)) or stochastic convex optimization (SCO) problems ([Shalev-Shwartz et al., 2009](#)), with a special focus on the, broadly understood, dimensional complexity of such reductions.

In particular, in [Chornomaz et al. \(2024\)](#) the authors proved that the reduction of a class to a  $d$ -dimensional SCO problem requires  $d$  to be at least the VC dimension of this class and, somewhat counterintuitive to the natural interpretation of VC as the “number of parameters”, for certain classes requires  $d$  to be at least exponential in VC. Their approach combines the Borsuk-Ulam theorem with convexity considerations and thus fits naturally into our framework. In particular, here we prove that their result can be refined in terms of lower-bounding an SCO dimension of a class (defined as a minimal  $d$  for which the above reduction is possible) with  $\text{sd}$ . We further note that the SCO dimension can, in turn, be seen as a refinement of *dimension complexity* (see, e.g. [Kamath et al. \(2020\)](#)), also known, especially from the communication complexity side, as *sign-rank* ([Paturi and Simon, 1986](#); [Forster, 2002](#); [Alon et al., 2016](#); [Hatami et al., 2022](#)), resulting in the bound  $\text{sd} - 1 \leq \text{SCO} \leq \text{sign-rank}$ .

We finally note that the study of reductions in [Chornomaz et al. \(2024\)](#) is broader than the particular thread that we concentrate on here and also extends to probabilistic and approximate reductions which, to our knowledge, do not fall under our current perspective.

## 2. Spherical dimension

By an  $n$ -dimensional sphere  $\mathbb{S}^n$  we understand the standard  $n$ -dimensional unit sphere  $\mathbb{S}^n = \{v \in \mathbb{R}^{n+1} \mid \|v\| = 1\}$ . For a topological space  $X$ , a map  $f: X \rightarrow X$  is called an *antipodality map* if it is continuous; fixed-point free, that is,  $f(x) \neq x$  for all  $x \in X$ ; and involutive, that is,  $f(f(x)) = x$  for all  $x \in X$ . An *antipodal space* is a pair  $(X, \nu)$  where  $X$  is a topological space and  $\nu: X \rightarrow X$  is an antipodality map. In particular, we consider the sphere  $\mathbb{S}^n$  as an antipodal space with its standard



antipodality map  $v \mapsto -v$ . For antipodal spaces  $(X, \nu)$  and  $(Y, \omega)$ , a continuous map  $f: X \rightarrow Y$  is called *equivariant* if it commutes with antipodalities, that is, if  $f \circ \nu(x) = \omega \circ f(x)$  for all  $x \in X$ .

For a simplicial complex<sup>6</sup>  $\mathbb{Q}$ , we say that a vertex map  $\nu: V(\mathbb{Q}) \rightarrow V(\mathbb{Q})$  is an antipodality map if it is i) simplicial, ii) an involution, and iii) there is no simplex containing a vertex and its  $\nu$ -image. Similarly to antipodal space, we define an *antipodal simplicial complex* as a pair  $(\mathbb{Q}, \nu)$ . Finally, we say that an antipodal simplicial complex  $(\mathbb{Q}^n, \nu)$  is a *simplicial  $n$ -sphere* if the antipodal space  $(\|\mathbb{Q}^n\|, \nu)$ , where  $\|\mathbb{Q}^n\|$  is the *geometric realization* of  $\mathbb{Q}^n$  and  $\nu$  is extended to it by linearity, is homeomorphic, as an antipodal space, to  $(\mathbb{S}^n, v \mapsto -v)$ .

As outlined in Section 1, for a binary class  $\mathcal{H}$ , we turn the *set of realizable distributions* of  $\mathcal{H}$  into a topological space  $\Delta_{\mathcal{H}}$  by equipping it with the *total variation distance*  $\text{TV}(\mu, \nu)$ . For a, not necessarily realizable, distribution  $\mu$ , its *antipodal* distribution  $-\mu$  is defined as  $\mathbb{P}_{-\mu}(E) = \mathbb{P}_{\mu}(-E)$  for all measurable  $E \subseteq X \times Y$ ; Informally, the antipodality map  $\mu \mapsto -\mu$  flips the labels of the samples. We then define  $\Delta^{\text{ant}}(\mathcal{H}) = \Delta(\mathcal{H}) \cap -\Delta(\mathcal{H})$ . Trivially,  $\Delta^{\text{ant}}$  is a closed subset of  $\Delta$ , and  $\mu \mapsto -\mu$  is the continuous *fixed-point free* involution on it. We refer the reader to Section A for a thorough review of the background of learning theory, relevant topological notions, and the details of the above construction.

We are now ready to define all the versions of the spherical dimensions, extending Definition 1:

**Definition 5 (Three spherical dimensions)** *We say that a class  $\mathcal{H}$  admits an infinitely supported  $n$ -sphere if there is an equivariant map from  $\mathbb{S}^n$  to  $\Delta_{\mathcal{H}}^{\text{ant}}$  with their standard antipodalities.*

*Furthermore,  $\mathcal{H}$  admits a (finitely supported)  $n$ -sphere if there is such map to  $\Delta_{\mathcal{H}'}^{\text{ant}}$ , where  $\mathcal{H}' = \mathcal{H}|_{X'}$  for some finite  $X' \subseteq X$ . Finally, it admits a simplicial  $n$ -sphere if this equivariant map is simplicial from a simplicial  $n$ -sphere to  $\Delta_{\mathcal{H}'}^{\text{ant}}$ , treated as a simplicial complex.*

*The spherical dimension  $\text{sd}(\mathcal{H})$  of  $\mathcal{H}$  is defined as a maximal integer  $n$  such that  $\mathcal{H}$  admits an  $n$ -sphere. Additionally, we say that  $\text{sd}(\mathcal{H}) = \infty$  if  $\mathcal{H}$  admits  $n$ -spheres for arbitrary large  $n$  and  $\text{sd}(\mathcal{H}) = -1$  if  $\Delta_{\mathcal{H}}^{\text{ant}} = \emptyset$ , that is, if  $\mathcal{H}$  does not even admit a 0-sphere. The infinitely supported and simplicial spherical dimensions  $\text{sd}_{\infty}(\mathcal{H})$  and  $\text{sd}_{\text{simp}}(\mathcal{H})$  are defined similarly using respective  $n$ -spheres.*

The fact that  $\Delta_{\mathcal{H}}^{\text{ant}}$  for finitely supported<sup>7</sup> classes is a simplicial complex, implicitly used in the definition of  $\text{sd}_{\text{simp}}$ , will be substantiated in Section A.3. Finally, we note that in the above  $\text{sd}$  is defined via *finitely supported* classes, while in the introduction we defined it for finite classes (which might not be finitely supported). The fact that they are equivalent needs proof, which we give in Corollary 28 in Section B.1. We finally note that, trivially,  $\text{sd}_{\text{simp}} \leq \text{sd} \leq \text{sd}_{\infty}$ , and that our central definition is that of  $\text{sd}$ . In particular, we say that a class admits an  $n$ -sphere if it admits a finitely supported  $n$ -sphere.

While the particularities of using the spherical dimension differ across the four outlined applications, it typically involves using a variant of the Borsuk-Ulam theorem or one of its relatives, out of which the most illustrative, to our view, is the following:

**Theorem 23 (Local BU, main theorem in Jaworowski (2000))** *Let  $\mathbb{P}^d$  be a finite simplicial complex of dimension at most  $d$ . If, for  $n \geq 1$ ,  $f: \mathbb{S}^n \rightarrow \mathbb{P}^d$  is continuous and  $2d \leq n$  then  $f$  collapses a pair of antipodal points, that is, there is  $x \in \mathbb{S}^n$  so that  $f(x) = f(-x)$ .*

6. We refer to Section A.2 for a review of standard topological notions, such as simplicial complex.

7. We say that  $\mathcal{H}$  is finitely supported if its domain is finite.

For example, its application to the problem of embedding into extremal classes involves constructing a map from  $\Delta$  into a cube complex  $\mathbb{C}$  of an extremal class  $\mathcal{E}$ , arguing that it does not collapse antipodal points, and inferring, by Theorem 23, that  $2 \dim(\mathbb{C}) = 2\text{VC}(\mathcal{E}) \geq \text{sd}(\mathcal{E}) + 1$ . We postpone a further discussion till Section A.4.

### 3. Disambiguations

The main question and the background follow Alon et al. (2022). A *partial class*  $\mathcal{P}$  on  $X$  is a subset of  $\{-, +, *\}^X$ , where  $*$  is treated as “undefined”. The standard involution  $y \mapsto -y$  on  $\{-, +, *\}$  is defined as  $-- = +$ ,  $-+ = -$ , and  $-* = *$ . In this section, we usually refer to regular classes, that is, to subsets of  $\{-, +\}^X$ , as to *total classes*. The definitions of partial concepts and hypotheses, and distributions realizable by a partial class, are formulated similarly as for total classes. For a partial hypothesis  $h \in \{-, +, *\}^X$ , the *support*  $\text{supp}(h)$  is  $\text{supp}(h) = \{x \in X \mid h(x) \neq *\}$ . Partial hypotheses are naturally partially ordered by *extension*, where  $h_2$  extends  $h_1$ , denoted  $h_1 \leq h_2$ , if  $\text{supp}(h_1) \subseteq \text{supp}(h_2)$  and for all  $x \in \text{supp}(h_1)$ ,  $h_1(x) = h_2(x)$ . A set  $S \subseteq X$  is shattered by  $\mathcal{P}$  if every binary pattern on  $S$  is realizable by  $\mathcal{P}$ , that is, if for every  $s \in \{-, +\}^S$  there is  $h \in \mathcal{P}$  such that  $s = h|_S$ ; as usual, VC dimension of  $\mathcal{P}$  is the maximal size of a set shattered by  $\mathcal{P}$ .

A total class  $\mathcal{H}$  *disambiguates*<sup>8</sup>  $\mathcal{P}$  if any partial concept in  $\mathcal{P}$  can be extended to a total concept in  $\mathcal{H}$ , that is:

$$\text{for any } h \in \mathcal{P} \text{ there is } \bar{h} \in \mathcal{H} \text{ s.t. } h(x) = \bar{h}(x) \text{ for all } x \in \text{supp}(h).$$

The one of the main results in Alon et al. (2022) is then as follows:

**Theorem 6 (Theorem 1 in Alon et al. (2022))** *There is a partial class  $\mathcal{P}$  on a countable domain such that  $\text{VC}(\mathcal{P}) = 1$  and for any disambiguation  $\mathcal{H}$  of  $\mathcal{P}$ ,  $\text{VC}(\mathcal{H}) = \infty$ .*

While the above theorem establishes the principal impossibility of disambiguations with finite VC, it is reasonable to ask how big a VC of disambiguation should be for some natural partial classes, such as linear classifiers with margin. For this end, let us define the family of partial classes  $\mathcal{L}_{n,\varepsilon}$ , where  $\mathcal{L}$  stands for “linear classifier”, for integer  $n$  and  $0 < \varepsilon < \pi/2$  as:

- The domain of  $\mathcal{L}_{n,\varepsilon}$  is the  $n$ -dimensional unit sphere  $\mathbb{S}^n$ ;
- The partial functions of  $\mathcal{L}_{n,\varepsilon}$  are parametrized by  $u \in \mathbb{S}^n$ , and we call the respective partial functions  $u_\varepsilon$ ;
- For  $u, v \in \mathbb{S}^n$ ,  $u_\varepsilon(v) = +$  if  $d(u, v) \leq \varepsilon$ ,  $-$  if  $d(-u, v) \leq \varepsilon$ , and  $*$  otherwise.

Although the particularities of the distance  $d$  on  $\mathbb{S}^n$  are not that important, we consider  $d$  to be the geodesic metric on  $\mathbb{S}^n$ . Alternatively,  $d(u, v)$  is the angle, in radians, between  $u$  and  $v$ . Thus,  $\varepsilon < \pi/2$  ensures that these classes are well-defined, and it is not hard to verify that for  $\varepsilon < \pi/4$ ,  $\text{VC}(\mathcal{L}_{n,\varepsilon}) = 1$ . We additionally note that  $\mathcal{L}_{n,\varepsilon_1}$  disambiguates  $\mathcal{L}_{n,\varepsilon_2}$  whenever  $\varepsilon_1 \geq \varepsilon_2$ ; here disambiguation of a partial class by partial is defined analogously, however normally by disambiguation we mean disambiguation by a total class. Finally, for all  $n \leq m$ ,  $\mathcal{L}_{n,\varepsilon}$  can be obtained from  $\mathcal{L}_{m,\varepsilon}$  by restricting the domain of the latter.

8. In Alon et al. (2022) this was called *strongly disambiguates*.



The family  $\mathcal{L}_{n,\varepsilon}$  is closely related to the family considered in Open Question 4 in Alon et al. (2022), although with some technical differences. In particular, the classes in OQ 21 are parametrized with *margin*  $\gamma$ , which relates to our  $\varepsilon$  as  $\gamma = \cos(\varepsilon)$ . Thus, while OQ 21 is concerned with  $\gamma \rightarrow 1$ , we are interested in  $\varepsilon \rightarrow 0$ . To better capture the fact that we are interested in the behavior of disambiguations for arbitrary small  $\varepsilon$ , we say that  $\mathcal{H}^n$  disambiguates  $\mathbb{S}^n$  if it disambiguates  $\mathcal{L}_{n,\varepsilon}$  for some  $\varepsilon > 0$ ; we note that  $\mathcal{L}_{n,0}$  can be trivially disambiguated by two functions, and thus the case  $\varepsilon = 0$  is excluded from consideration.

Our main result establishes the rough equivalence between disambiguations of  $\mathbb{S}^n$  and spherical dimension; recall that rough equivalence means that the respective parameters are the same from the perspective of Conjectures 14, 15, and 16.

**Theorem 4 (Rough equivalence between disambiguations and spherical dimension)** *For integers  $a, b$ , and  $n$ , the following are equivalent:*

- *There is a class with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ , and  $\text{sd}_{\text{simp}} \geq n$ ;*
- *There is a disambiguation of  $\mathbb{S}^n$  with  $\text{VC} \leq a$  and  $\text{VC}^{*a} \leq b$ .*

#### 4. Extremal classes

In this section, we are going to built upon Chase et al. (2024a). In particular, we refer to Sections 1.1, 1.2, and 2 there for an in-depth discussion of maximum and extremal classes, their relation to sample compression schemes, and the relevant topological notions. Here we repeat, in a minimalistic way, certain definitions and basic properties.

The key property, enabling the definition of an extremal class, is the Pajor inequality, perhaps better known in the form of its direct corollary, the Sauer-Shelah-Perles inequality, which establishes a polynomial upper bound on the size of classes of bounded VC dimension.

**Theorem 7 (Pajor (1985))** *A finite class  $\mathcal{H}$  shatters at least as many sets as it has.*

Extremal classes are then defined as follows:

**Definition 8 (Extremal Classes)** *A finite class  $\mathcal{E}$  is called extremal if it meets Pajor's inequality with equality.*

As long as we are concerned with embedding arbitrary classes into extremal, we define the following natural parameter: for a class  $\mathcal{H}$ , let  $\text{VC}^{\text{extr}}(\mathcal{H})$  be the minimal VC dimension of an extremal class, containing  $\mathcal{H}$ .

**Theorem 9** *For any finite extremal concept class  $\mathcal{E}$ , the space  $\Delta_{\mathcal{E}}$  can be mapped into the cubical complex  $\mathbb{C}_{\mathcal{E}}$ , of dimension  $\text{VC}(\mathcal{E})$ , without collapsing any antipodal points in  $\Delta_{\mathcal{E}}^{\text{ant}}$ .*

*In particular, every extremal concept class  $\mathcal{E}$  satisfies  $\text{sd}(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) - 1$  and  $\text{VC}^*(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) + 1$ . Consequently, for any class  $\mathcal{H}$ ,  $\text{sd}(\mathcal{H}) \leq 2\text{VC}^{\text{extr}}(\mathcal{H}) - 1$ , or, equivalently,  $\text{VC}^{\text{extr}}(\mathcal{H}) \geq \lceil (\text{sd}(\mathcal{H}) + 1)/2 \rceil$ .*

We refer to Section D.1 for the definition of the cubical complex of the class. The  $\text{VC}^*(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) + 1$  was the main result in Chase et al. (2024a) (Theorem B). It follows from  $\text{sd}(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) - 1$  via  $\text{sd} \geq \text{VC}^* - 2$  bound in Lemma 2. However, the bound in terms of  $\text{sd}$  has a wider range of applicability because, as noted in Lemma 3, there are classes for which  $\text{sd} \gg \text{VC}, \text{VC}^*$ . The

proof of this will be given in Section D. We also note that, structurally, we will prove a much stronger result that the corresponding map is the final map of a *deformation retraction* of  $\Delta_{\mathcal{E}}$  onto  $\mathbb{C}_{\mathcal{E}}$ , naturally embedded into it.

## 5. Stability and list replicability

In this section, we follow the setup from Chase et al. (2024b).

A learning rule  $\mathcal{A}$  for a class  $\mathcal{H}$  is called  $(\varepsilon, L)$ -*list replicable learner* for  $\mathcal{H}$  if for every  $\delta > 0$  there exists  $n = n(\varepsilon, L, \delta)$  such that for every distribution  $\mu$  that is realizable by  $\mathcal{H}$ , there exist hypotheses  $h_1, h_2, \dots, h_L$  such that  $\mathbb{P}_{S \sim \mu^n}[\mathcal{A}(S) \in \{h_1, \dots, h_L\}] \geq 1 - \delta$  and for all  $\ell \in [L]$ ,  $L_{\mu}(h_{\ell}) < \varepsilon$ .

We then define the  $\varepsilon$ -*list replicability number*  $\text{LR}(\mathcal{H}, \varepsilon)$  as minimal  $L$  for which  $\mathcal{H}$  has a  $(\varepsilon, L)$ -learner, and the *list replicability number*  $\text{LR}(\mathcal{H})$  as  $\text{LR}(\mathcal{H}) = \lim_{\varepsilon \rightarrow 0} \text{LR}(\mathcal{H}, \varepsilon)$ ; note that, trivially,  $\text{LR}(\mathcal{H}, \varepsilon)$  is non-increasing in  $\varepsilon$ . The following theorem refines one of the main results in Chase et al. (2024b), namely their Theorem D:

**Theorem 10** *Let  $\mathcal{H}$  be a concept class such that  $\text{sd}(\mathcal{H}) \geq 1$ . Then, for any  $0 < \varepsilon < 1/2$ , it holds*

$$\text{LR}(\mathcal{H}, \varepsilon) \geq \lceil (\text{sd}(\mathcal{H}) + 3)/2 \rceil.$$

*In particular,  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1 + \lceil \text{VC}(\mathcal{H})/2 \rceil$  whenever  $\text{VC}(\mathcal{H}) \geq 2$ , and  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1 + \lfloor \text{VC}^{*a}(\mathcal{H})/2 \rfloor$  (with no additional condition).*

*The only exception for the bounds in terms of  $\text{VC}$  and  $\text{sd}$  are the classes of the form  $\mathcal{H} = \{h_-, h_+\}$  such that  $h_-(x) = -h_+(x)$  for all  $x \in X$ , for example,  $\mathcal{C}_1$ , for which  $\text{LR}(\mathcal{H}, \varepsilon) = 1$ ,  $1 + \lceil \text{VC}(\mathcal{H})/2 \rceil = 1 + \lceil 1/2 \rceil = 2$ , and  $\lceil (\text{sd}(\mathcal{H}) + 3)/2 \rceil = \lceil (0 + 3)/2 \rceil = 2$ .*

We finally note that the condition  $\text{VC}(\mathcal{H}) \geq 2$  was overlooked in Chase et al. (2024b).

Curiously enough, Theorem E in Chase et al. (2024b) proves the upper bound  $\text{LR}(\mathcal{U}_n) \leq \lceil (n + 1)/2 \rceil$  by constructing an explicit list-replicable learner for this class. Together with  $\text{LR} \geq 1 + \lfloor \text{VC}^*/2 \rfloor$  this implies that  $\text{LR}(\mathcal{U}_n) = \lceil (n + 1)/2 \rceil$  and also gives near-tight bound on  $\text{sd}(\mathcal{U}_n)$  from Lemma 2.

## 6. Reductions to stochastic convex optimization

In this section, we follow the setup from Chornomaz et al. (2024).

A  $d$ -dimensional SCO problem can be defined by an abstract sample space  $Z$ , a convex set  $W \subseteq \mathbb{R}^d$ , and a collection of convex loss functions  $\{\ell_z\}_{z \in Z}$  from  $W$  to  $\mathbb{R}^+$ . The learner receives a sample of  $n$  examples from a distribution  $\zeta$  over  $Z$  and its goal is to output  $w \in W$  that minimizes the average loss  $L_{\zeta}(w) = \mathbb{E}_{z \sim \zeta} \ell_z(w)$ .

**Definition 11 (Reduction to SCO)** *Let  $\mathcal{H}$  be a concept class over domain  $X$ , and let  $\alpha > 0$  and  $\beta \geq 0$ . An  $(\alpha, \beta)$ -reduction from  $\mathcal{H}$  to a  $d$ -dimensional SCO problem  $(Z, W)$  consists of two maps  $r_{\text{in}} : X \times \{-, +\} \rightarrow Z$  and  $r_{\text{out}} : W \rightarrow \{-, +\}^X$  such that for every realizable distribution  $\mu \in \Delta(\mathcal{H})$  we have*

$$L_{r_{\text{in}}(\mu)}(w) \leq \inf_{w' \in W} L_{r_{\text{in}}(\mu)}(w') + \alpha \implies L_{\mu}(r_{\text{out}}(w)) \leq \beta,$$

where  $r_{\text{in}}(\mu)$  is the push-forward measure induced by sampling  $z \sim \mu$  and mapping it to  $r_{\text{in}}(z)$ .

We define the SCO dimension of  $\mathcal{H}$ , denoted  $\text{SCO}(\mathcal{H})$  to be the smallest number  $d$  such that there is an  $(\alpha, \beta)$ -reduction of  $\mathcal{H}$  to a  $d$ -dimensional SCO problem, for some  $\alpha > 0$  and  $0 < \beta < 1/2$ .

The above definition is a compilation of Definition 1 and Example 4 (SCO learning task), Definition 6 (reductions), and Definition 12 (SCO-dimension) in Chornomaz et al. (2024). Informally, it aims at capturing the situation when we can design a weak learner for  $\mathcal{H}$  by transforming it to an SCO problem, solving the latter (with some accuracy), and then reinterpreting the result of the SCO as a learner's hypothesis. Then their Theorem 1 can be refined as

**Theorem 12** *For any class  $\mathcal{H}$  we have*

$$\text{sd}(\mathcal{H}) + 1 \leq \text{SCO}(\mathcal{H}) \leq \text{sign-rank}(\mathcal{H}).$$

Here the *sign-rank* is a one of the most popular parameters in the context of studying the expressivity of kernel methods and in communication complexity, dated back to Paturi and Simon (1986). Due to space restrictions, we move its discussion to Section B.5, where we prove Theorem 12. There we will also prove the next theorem, which turns out to be instrumental in proving  $\text{sd}(\mathcal{U}_n^m) \leq mn$  bound in Lemma 3.

**Theorem 13** *For any two concept classes  $\mathcal{H}_1, \mathcal{H}_2$  we have*

$$\text{sign-rank}(\mathcal{H}_1 \times \mathcal{H}_2) \leq \text{sign-rank}(\mathcal{H}_1) + \text{sign-rank}(\mathcal{H}_2).$$

## 7. Conclusion and future work

Most of our work can be summarized by the diagram in Figure 3 below<sup>9</sup>. As one can see,  $\text{sd}$  lower-bounds several parameters of interest, while in turn being lower-bounded by both VC and VC\* (up to an additive constant), but at times being asymptotically larger than the two (Lemma 3).

At the same time, we have very little understanding of how the spherical dimension can grow with VC and VC\*. The following three conjectures (from the easiest to the hardest) summarize what we expect to see and are based mostly on the examples in Lemmas 2 and 3. However, we have no means in sight of proving even the first most basic one (or disproving the hardest).

**Conjecture 14** *VC and spherical dimension go to infinity together. That is, there is increasing  $f: \mathbb{N} \rightarrow \mathbb{N}$  such that for any class  $\mathcal{H}$ ,  $\text{sd}(\mathcal{H}) \leq f(\text{VC}(\mathcal{H}))$ .*

**Conjecture 15** *Spherical dimension is at most exponential in both VC and VC\*.*

**Conjecture 16** *Up to log factors, the spherical dimension is upper-bounded by  $\text{VC} \cdot \text{VC}^*$ .*

The best we were able to achieve in the direction of proving Conjecture 14 is the following:

**Theorem 17** *If  $\text{VC}(\mathcal{H}) \leq 1$  then  $\text{sd}(\mathcal{H}) \leq 1$ . Alternatively, if  $\text{sd}(\mathcal{H}) \geq 2$  then  $\text{VC}(\mathcal{H}) \geq 2$ .*

---

9. Out of the things not covered in Figure 3, let us separately note the quantitative behaviour of the parameters of classes under products (Section B.6).

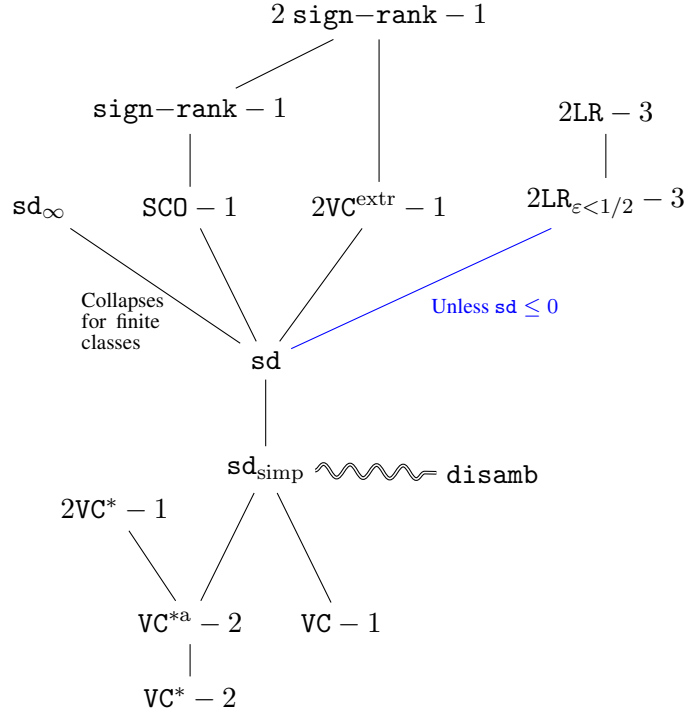


Figure 3: Hasse diagram for spherical dimensions and related values. For any class  $\mathcal{H}$ , the respective values are non-decreasing with respect to this partial order. The wavy line indicates rough equivalence. The parameter  $\text{disamb}$  is defined in the proof of Theorem 4; roughly, how big of a sphere can a class disambiguate.

Despite the modest statement, the proof relies on a highly nontrivial relation of spherical dimension and extremal classes, see Section 4, specifically Theorem 9.

We find it intriguing that the only known to us upper-bound of the form  $\text{VC} \cdot \text{VC}^*$  is from the famous construction of sample compression schemes (Moran and Yehudayoff, 2016); just as in Conjecture 16, this bound is up to log-factors. This parallel is further reinforced by the results in Section 4, where we use the spherical dimension to improve  $\text{VC}^* \leq 2\text{VC} + 1$  bound for extremal classes from Chase et al. (2024a), which, in turn, was used to prove that a certain attack on the sample compression schemes conjecture cannot succeed.

As it becomes apparent from Figure 3, one of the obvious venues towards proving Theorem 14 is to upper bound one of  $\text{SC0}$ , or  $\text{VC}^{\text{extr}}$ , or  $\text{LR}$  by a function of  $\text{VC}$ ; note that the sign-rank cannot be bounded in such way, as was shown in Alon et al. (2016). While there are no clear ways of doing this, it can be a viable prospect of attack on these conjectures, and might be of independent interest (for example, bounding  $\text{VC}^{\text{extr}}$  by a function of  $\text{VC}$  is Open Problem 1 in Chase et al. (2024a)). We also would like to point at a recent manuscript of Blondal, Hatami, Hatami, Lalov, and Tretiak (2025), which seems relevant from the perspective of upper-bounding  $\text{LR}$ .

Let us finish with two questions which we consider less important, but still relevant in shaping the theory of spherical dimension.

**Question 18** *Is there a separation between  $\text{sd}_{\text{simp}}$ ,  $\text{sd}$ , and  $\text{sd}_{\infty}$ ? That is, is there a class  $\mathcal{H}$  such that  $\text{sd}_{\text{simp}}(\mathcal{H}) < \text{sd}(\mathcal{H})$  or  $\text{sd}(\mathcal{H}) < \text{sd}_{\infty}(\mathcal{H})$ ? We expect to have the separation in the first case, but not in the second one.*

**Question 19** *Is  $\text{sd}$ , or one of its other two versions, computable? The question, to which we expect the negative answer, is motivated by the well-known undecidability result of [Markov \(1958\)](#).*

## Acknowledgments

Bogdan Chornomaz, Shay Moran, and Tom Waknine are supported by the European Union (ERC, GENERALIZATION, 101039692). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Shay Moran is also supported by Robert J. Shillman Fellowship, by ISF grant 1225/20, by BSF grant 2018385, by an Azrieli Faculty Fellowship, by Israel PBC-VATAT, and by the Technion Center for Machine Learning and Intelligent Systems (MLIS).

The authors also wish to express their gratitude to Ron Holzman, Noga Alon, and Roy Meshulam for fruitful discussions.

## References

- Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus VC dimension. In *Conference on Learning Theory*, pages 47–80. PMLR, 2016.
- Noga Alon, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private PAC learning implies finite Littlestone dimension. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 852–860, 2019.
- Noga Alon, Steve Hanneke, Ron Holzman, and Shay Moran. A theory of PAC learnability of partial concept classes. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 658–671. IEEE, 2022.
- Patrick Assouad. Densité et dimension. *Annales de l’institut Fourier*, 33(3):233–282, 1983.
- Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, 2016.
- Kaspars Balodis, Shalev Ben-David, Mika Göös, Siddhartha Jain, and Robin Kothari. Unambiguous DNFs and Alon–Saks–Seymour. *SIAM Journal on Computing*, pages FOCS21–157, 2023.
- Shai Ben-David. 2 notes on classes with vapnik-chervonenkis dimension 1. *arXiv preprint arXiv:1507.05307*, 2015.
- Shai Ben-David and Ami Litman. Combinatorial Variability of Vapnik-Chervonenkis Classes with Applications to Sample Compression Schemes. *Discrete Applied Mathematics*, 86(1):3–25, 1998. doi: 10.1016/S0166-218X(98)00000-6. URL [http://dx.doi.org/10.1016/S0166-218X\(98\)00000-6](http://dx.doi.org/10.1016/S0166-218X(98)00000-6).

- Shai Ben-David, Nadav Eiron, and Hans-Ulrich Simon. Limitations of learning via embeddings in Euclidean half spaces. *Journal of Machine Learning Research*, 3:441–461, 2002. URL <http://www.jmlr.org/papers/v3/bendavid02a.html>.
- Shalev Ben-David, Pooya Hatami, and Avishay Tal. Low-sensitivity functions from unambiguous certificates. *arXiv preprint arXiv:1605.07084*, 2016.
- Ari Blonda, Shan Gao, Hamed Hatami, and Pooya Hatami. Stability and list-replicability for agnostic learners. *arXiv preprint arXiv:2501.05333*, 2025.
- Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Borsuk-ulam and replicable learning of large-margin halfspaces. *arXiv preprint arXiv:2503.15294*, 2025.
- Béla Bollobás and Andrew J. Radcliffe. Defect Sauer results. *Journal of Combinatorial Theory, Series A*, 72(2):189–208, 1995.
- B. Bollobás, I. Leader, and A. J. Radcliffe. Reverse Kleitman Inequalities. *Proceedings of the London Mathematical Society*, s3-58(1):153–168, 01 1989. ISSN 0024-6115. doi: 10.1112/plms/s3-58.1.153. URL <https://doi.org/10.1112/plms/s3-58.1.153>.
- Karol Borsuk. Drei sätze über die  $n$ -dimensionale Euklidische sphäre. *Fundamenta Mathematicae*, 20(1):177–190, 1933.
- Olivier Bousquet, Steve Hanneke, Shay Moran, and Nikita Zhivotovskiy. Proper learning, Helly number, and an optimal SVM bound. In *Proceedings of the 33<sup>rd</sup> Conference on Learning Theory*, 2020.
- Martin R Bridson and André Haeffliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2013.
- John L Bryant. Piecewise linear topology. *Handbook of geometric topology*, pages 219–259, 2002.
- Mark Bun, Roi Livni, and Shay Moran. An equivalence between private classification and online prediction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402. IEEE, 2020.
- Jérémie Chalopin, Victor Chepoi, Shay Moran, and Manfred K. Warmuth. Unlabeled sample compression schemes and corner peelings for ample and maximum classes. *J. Comput. Syst. Sci.*, 127:1–28, 2022. doi: 10.1016/J.JCSS.2022.01.003. URL <https://doi.org/10.1016/j.jcss.2022.01.003>.
- Jérémie Chalopin, Victor Chepoi, Fionn Mc Inerney, Sébastien Ratel, and Yann Vaxès. Sample compression schemes for balls in graphs. *SIAM J. Discret. Math.*, 37(4):2585–2616, 2023. doi: 10.1137/22M1527817. URL <https://doi.org/10.1137/22m1527817>.
- Zachary Chase, Shay Moran, and Amir Yehudayoff. Replicability and stability in learning. *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 2023.
- Zachary Chase, Bogdan Chornomaz, Steve Hanneke, Shay Moran, and Amir Yehudayoff. Dual VC dimension obstructs sample compression by embeddings. In *COLT 2024 - The 37th Annual Conference on Learning Theory, Edmonton, Canada, June 30-July 3, 2024*, 2024a.



- 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*, pages 1769–1780, 2024b.
- Victor Chepoi, Kolja Knauer, and Manon Philibert. Two-dimensional partial cubes. *The Electronic Journal of Combinatorics*, pages 3–29, 2020.
- Victor Chepoi, Kolja Knauer, and Manon Philibert. Labeled sample compression schemes for complexes of oriented matroids. *CoRR*, abs/2110.15168, 2021. URL <https://arxiv.org/abs/2110.15168>.
- Victor Chepoi, Kolja Knauer, and Manon Philibert. Ample completions of oriented matroids and complexes of uniform oriented matroids. *SIAM Journal of Discrete Mathematics*, 36(1):509–535, 2022.
- Bogdan Chornomaz, Shay Moran, and Tom Waknine. On reductions and representations of learning problems in euclidean spaces. *arXiv preprint arXiv:2411.10784*, 2024.
- Peter Dixon, A. Pavan, Jason Vander Woude, and N. V. Vinodchandran. List and certificate complexities in replicable learning, 2023.
- Sally Floyd and Manfred Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(3):269–304, 1995.
- Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002. doi: 10.1016/S0022-0000(02)00019-3. URL [http://dx.doi.org/10.1016/S0022-0000\(02\)00019-3](http://dx.doi.org/10.1016/S0022-0000(02)00019-3).
- Jürgen Forster, Niels Schmitt, Hans-Ulrich Simon, and Thorsten Suttrop. Estimating the optimal margins of embeddings in Euclidean half spaces. *Machine Learning*, 51(3):263–281, 2003. doi: 10.1023/A:1022905618164. URL <http://dx.doi.org/10.1023/A:1022905618164>.
- Eran Gat and Shafi Goldwasser. Probabilistic search algorithms with unique answers and their cryptographic applications. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 18, pages 1–3, 2011.
- Mika Göös. Lower bounds for clique vs. independent set. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1066–1076. IEEE, 2015.
- Hamed Hatami, Kaave Hosseini, and Xiang Meng. A Borsuk-Ulam lower bound for sign-rank and its application. *Electron. Colloquium Comput. Complex.*, TR22-130, 2022. URL <https://eccc.weizmann.ac.il/report/2022/130>.
- Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
- Morris W Hirsch and John Milnor. Some curious involutions of spheres. *Bulletin of the American Mathematical Society*, 70(3):372–377, 1964.
- Russell Impagliazzo, Rex Lei, Toniann Pitassi, and Jessica Sorrell. Reproducibility in learning. In *STOC*, pages 818–831, 2022.

- Jan Jaworowski. Periodic coincidence for maps of spheres. *Kobe journal of mathematics*, 17(1): 21–26, 2000.
- Alkis Kalavasis, Amin Karbasi, Grigoris Velegkas, and Felix Zhou. On the computational landscape of replicable learning. *arXiv preprint arXiv:2405.15599*, 2024.
- Pritish Kamath, Omar Montasser, and Nathan Srebro. Approximate is good enough: Probabilistic variants of dimensional and margin complexity. In Jacob D. Abernethy and Shivani Agarwal, editors, *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125 of *Proceedings of Machine Learning Research*, pages 2236–2262. PMLR, 2020. URL <http://proceedings.mlr.press/v125/kamath20b.html>.
- Dima Kuzmin and Manfred K. Warmuth. Unlabeled compression schemes for maximum classes. *Journal of Machine Learning Research*, 8:2047–2081, 2007. URL <http://dl.acm.org/citation.cfm?id=1314566>.
- J. Lawrence. Lopsided sets and orthant-intersection of convex sets. *Pacific J. Math.*, 104:155–173, 1983.
- Nathan Linial and Adi Shraibman. Learning complexity vs communication complexity. *Combinatorics, Probability & Computing*, 18(1-2):227–245, 2009. doi: 10.1017/S0963548308009656. URL <http://dx.doi.org/10.1017/S0963548308009656>.
- Philip M Long. On agnostic learning with  $\{0, *, 1\}$ -valued and real-valued hypotheses. In *Computational Learning Theory: 14th Annual Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001 Amsterdam, The Netherlands, July 16–19, 2001 Proceedings 14*, pages 289–302. Springer, 2001.
- Lazar Lyusternik and Lev Shnirel'man. Topological methods in variational problems. *Issledovaniya Matematiki I Mekhaniki Pri O. M. G. U. Moscow*, 1930.
- Andrei Andreevich Markov. The insolubility of the problem of homeomorphy. In *Doklady Akademii Nauk*, volume 121, pages 218–220. Russian Academy of Sciences, 1958.
- Jiří Matoušek, Anders Björner, and Günter M Ziegler. *Using the Borsuk-Ulam theorem: lectures on topological methods in combinatorics and geometry*, volume 2003. Springer, 2003.
- S. Moran and M. Warmuth. Labeled compression schemes for extremal classes. In *Proceedings of the 27<sup>th</sup> International Conference on Algorithmic Learning Theory*, pages 34–49. Springer, 2016.
- Shay Moran. Shattering-extremal systems. *arXiv preprint*, 1211.2980, 2012.
- Shay Moran and Amir Yehudayoff. Sample compression schemes for VC classes. *Journal of the ACM (JACM)*, 63(3):1–10, 2016.
- Alain Pajor. *Sous-espaces  $\ell_1^n$  des Espaces de Banach*. Travaux en Cours. Hermann, Paris, 1985.
- Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *Journal of Computer and System Sciences*, 33(1):106–123, 1986.

- Leonard Pitt and Manfred K. Warmuth. Prediction-preserving reducibility. *J. Comput. Syst. Sci.*, 41(3):430–467, 1990. doi: 10.1016/0022-0000(90)90028-J. URL [https://doi.org/10.1016/0022-0000\(90\)90028-J](https://doi.org/10.1016/0022-0000(90)90028-J).
- Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- Benjamin I. P. Rubinstein and J. Hyam Rubinstein. A geometric approach to sample compression. *Journal of Machine Learning Research*, 13:1221–1261, 2012. URL <http://dl.acm.org/citation.cfm?id=2343686>.
- Benjamin I. P. Rubinstein, Peter L. Bartlett, and J. Hyam Rubinstein. Shifting: One-inclusion mistake bounds and sample compression. *J. Comput. Syst. Sci.*, 75(1):37–59, 2009. doi: 10.1016/J.JCSS.2008.07.005. URL <https://doi.org/10.1016/j.jcss.2008.07.005>.
- J Hyam Rubinstein, Benjamin IP Rubinstein, and Peter L Bartlett. Bounding embeddings of VC classes into maximum classes. *Measures of Complexity: Festschrift for Alexey Chervonenkis*, pages 303–325, 2015.
- Joachim Hyam Rubinstein and Benjamin I. P. Rubinstein. Unlabelled sample compression schemes for intersection-closed classes and extremal classes. In *NeurIPS*, 2022. URL [http://papers.nips.cc/paper\\_files/paper/2022/hash/54d6a55225cebbdc16fbb0e45c5bdf2b-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/54d6a55225cebbdc16fbb0e45c5bdf2b-Abstract-Conference.html).
- Michah Sageev. CAT(0) cube complexes and groups. *Geometric Group Theory, IAS/Park City Math Series, 21, AMS and IAS/PCMI*, pages 7–53, 2014.
- Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- Shai Shalev-Shwartz, Ohad Shamir, Nathan Srebro, and Karthik Sridharan. Stochastic convex optimization. In *COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009*, 2009. URL <http://www.cs.mcgill.ca/%7Ecolt2009/papers/018.pdf#page=1>.
- Manfred K. Warmuth. Compressing to VC dimension many points. In *COLT/Kernel*, pages 743–744, 2003. doi: 10.1007/978-3-540-45167-9\_60. URL [http://dx.doi.org/10.1007/978-3-540-45167-9\\_60](http://dx.doi.org/10.1007/978-3-540-45167-9_60).

## Appendix A. Background

### A.1. Learning theory toolbox

Here we pay the technical debt from the main part of the paper and formally define all the notions outlined in it. Notationwise, for an integer  $n$ , we write  $[n]$  to denote the set  $\{1, \dots, n\}$ . We also write  $A \sqcup B$  to denote a disjoint union of the sets  $A$  and  $B$ .

We usually consider a class  $\mathcal{H} \subseteq Y^X$ , where  $X$  is called the *domain* and  $Y = \{-, +\}$  the *label space* of  $\mathcal{H}$ . A function  $h \in Y^X$  is called a *hypothesis*, and  $h \in \mathcal{H}$  is called a *concept* of  $\mathcal{H}$ . Unless stated otherwise, we always consider the class  $\mathcal{H}$  to be nonempty, that is, containing at least one hypothesis. For a hypothesis  $h$  and a distribution  $\mu$  over  $X \times Y$ , we define the 0/1 *population loss*  $L_\mu(h)$  of  $h$  with respect to  $\mu$  as

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

The distribution  $\mu$  is called *realizable* by  $\mathcal{H}$  if

$$\inf_{h \in \mathcal{H}} L_\mu(h) = 0.$$

In case  $X$  is finite, the definition of realizability can be rectified as:  $\mu$  is realizable by  $\mathcal{H}$  if there is  $h \in \mathcal{H}$  such that  $h(x) = y$  for all  $(x, y) \in \text{supp}(\mu)$ . Here  $\text{supp}(\mu)$  is the *support* of  $\mu$ ,  $\text{supp}(\mu) = \{(x, y) \mid \mathbb{P}_\mu(x, y) > 0\}$ .

For a subset  $E \subseteq X \times Y$ , we define  $-E = \{(x, -y) \mid (x, y) \in E\}$ . For a distribution  $\mu$  over  $X \times Y$ , we define  $-\mu$  as a distribution defined by  $P_{-\mu}(E) = P_\mu(-E)$  for all measurable  $E$ . In particular, if  $X$  is finite, this is equivalent to saying that  $P_{-\mu}(x, y) = P_\mu(x, -y)$ , for all  $x, y \in X \times Y$ . We call  $\mu \mapsto -\mu$  an *antipodality map* on the set of distributions over  $X \times Y$ .

Let  $\Delta(\mathcal{H})$  be the space of all distributions realizable by  $\mathcal{H}$ , equipped with the *total variation distance*

$$\text{TV}(\mu, \nu) = \sup_E |\mu(E) - \nu(E)|,$$

where the supremum is over all measurable events  $E$ . It is easy to see that  $\Delta(\mathcal{H})$  is closed and the antipodality map  $\mu \mapsto -\mu$  is continuous with respect to TV. Let  $\Delta^{\text{ant}}(\mathcal{H}) = \Delta(\mathcal{H}) \cap -\Delta(\mathcal{H})$ . Trivially,  $\Delta^{\text{ant}}$  is a closed subset of  $\Delta$ , which is closed under the antipodality map, that is, such that  $-\mu \in \Delta^{\text{ant}}$  whenever  $\mu \in \Delta^{\text{ant}}$ . The following easy property turns out to be quite handy in this setup: For any distribution  $\mu$  and hypothesis  $h$ , it holds

$$L_\mu(h) + L_{-\mu}(h) = 1. \tag{1}$$

Informally, no hypothesis can be any good for both the distribution and its antipodal. In particular, it easily implies that the antipodality is *fixed-point free* on  $\Delta$ , that is, for no  $\mu \in \Delta$ ,  $\mu = -\mu$ .

For a hypothesis  $h$  over  $X$  and  $S \subseteq X$ , a *restriction*  $h|_S$  of  $h$  to  $S$  is a hypothesis  $h'$  over  $S$  defined as  $h'(s) = h(s)$  for all  $x \in S$ . Similarly, a restriction of  $\mathcal{H}$  to  $S$  is defined as  $\mathcal{H}|_S = \{h|_S \mid h \in \mathcal{H}\}$ . A class  $\mathcal{H}$  over  $X$  *shatters* a set  $S \subseteq X$  if for every *pattern*  $s$  on  $S$ , that is, for every  $s: S \rightarrow \{-, +\}$ , there is  $h \in \mathcal{H}$  such that  $h|_S = s$  (by convention, the empty set is always shattered). The VC dimension of  $\mathcal{H}$ , denoted by  $\text{VC}(\mathcal{H})$ , is the largest size of a set shattered by  $\mathcal{H}$ .

For a class  $\mathcal{H}$  over  $X$ , the *dual class*  $\mathcal{H}^*$  of  $\mathcal{H}$  is a class on a domain  $\mathcal{H}$  whose functions are parametrized by  $X$ , and, for  $x \in X \sim x \in \mathcal{H}^*$  and  $h \in \mathcal{H}$ , it holds  $x(h) = h(x)$ . The dual VC

dimension of  $\mathcal{H}$ ,  $\text{VC}^*(\mathcal{H})$ , is defined as the VC dimension of  $\mathcal{H}^*$ , that is,  $\text{VC}^*(\mathcal{H}) = \text{VC}(\mathcal{H}^*)$ . It also can be educational to define  $\text{VC}^*$  explicitly. A subset  $H \subseteq \mathcal{H}$  is dually-shattered if for every *dual pattern*  $r: H \rightarrow \{-, +\}$  there is  $x \in X$  such that  $r(h) = h(x)$  for all  $h \in H$ . Just as before,  $\text{VC}^*(\mathcal{H})$  is the maximal size of the set, dually shattered by  $\mathcal{H}$ .

Let us now formally define a relaxation of the VC dimension that works better with respect to our topological approach. We say that a class  $\mathcal{H}$  on a domain  $X$  *antipodally shatters*<sup>10</sup>  $S \subseteq X$  if for every binary pattern  $s$  on  $S$  there is  $h \in \mathcal{H}$  such that either  $h|_S = s$  or  $h|_S = -s$ . We define *antipodal VC dimension*  $\text{VC}^a(\mathcal{H})$  as the maximal size of a set, antipodally shattered by  $\mathcal{H}$ . The *dual antipodal VC dimension*  $\text{VC}^{*a}(\mathcal{H})$  is defined dually. We note that, as a matter of fact, the VC dimension is good for us as it is, but dual antipodal VC behaves better than dual VC, so we will mostly use only the latter relaxation.

While  $\text{VC}^a$  and  $\text{VC}^{*a}$  clearly are relaxations of VC and  $\text{VC}^*$ , in particular,  $\text{VC} \leq \text{VC}^a$  and  $\text{VC}^* \leq \text{VC}^{*a}$ , it is noteworthy that, as we will prove in Section B.2, Assouad’s bounds  $\lfloor \log \text{VC} \rfloor \leq \text{VC}^* \leq 2^{\text{VC}+1} - 1$  (see Assouad (1983)) can be extended to the antipodal VC dimensions:

**Lemma 20 (Assouad’s bounds for antipodal shattering)** *For a (partial) class  $\mathcal{H}$ , it holds*

$$\lfloor \log \text{VC}(\mathcal{H}) \rfloor \leq \lfloor \log \text{VC}^a(\mathcal{H}) \rfloor \leq \text{VC}^*(\mathcal{H}) \leq \text{VC}^{*a}(\mathcal{H}) \leq 2^{\text{VC}(\mathcal{H})+1} - 1,$$

and

$$\text{VC}^*(\mathcal{H}) \leq \text{VC}^{*a}(\mathcal{H}) \leq 2\text{VC}^*(\mathcal{H}) + 1.$$

Moreover, all inequalities between the neighboring values above are sharp (i.e., for each of these inequalities there exist classes for which they hold with equality).

## A.2. Topological toolbox

As it is customary, we consider all maps between topological spaces to be continuous without saying it explicitly. Topological spaces  $X$  and  $Y$  are called *homeomorphic*, denoted  $X \cong Y$ , if there is an isomorphism  $f: X \rightarrow Y$  such that both  $f$  and  $f^{-1}$  are continuous. We usually think of homeomorphic spaces as of the same space. We will now briefly review some basic definitions from simplicial geometry, following Chapter 1 in Matoušek et al. (2003). All sets in the definitions of simplicial complexes below are assumed to be finite.

An *abstract simplicial complex*  $K$  is a pair  $(V, S)$ , where  $V = V(K)$  is a set of *vertices* of  $K$ , and  $S = S(K) \subseteq 2^V$  is a *downward closed* set of simplices of  $K$ . Here downward closed means that for any  $s \in S(K)$  and any  $t \subseteq s$ , it holds  $t \in S(K)$ . An  $s \in S(K)$  is called an *abstract simplex*, and its *dimension* is  $\dim(s) = |s| - 1$ ; that is, for example, a 2-dimensional simplex contains 3 vertices. A simplex  $s \in S(K)$  is called *maximal* if it is not a proper subset of any other simplex in  $K$ . We also call a  $k$ -dimensional simplex a *k-simplex*. The dimension of  $K$  is the maximal dimension of its simplices.

A *geometric simplex*  $\sigma$  in the *ambient space*  $\mathbb{R}^d$  is a convex hull of an affinely independent set  $A \subseteq \mathbb{R}^d$ ;  $A$  is called the set of *vertices* of  $\sigma$ ,  $A = V(\sigma)$ . In this case, the *dimension* of  $\sigma$  is  $k = \dim(\sigma) = |A| - 1$  and  $\sigma$  is called a *k-simplex*. A convex hull of an arbitrary subset of  $V(\sigma)$  is called a *face* of  $\sigma$ , and is itself a simplex. A geometric simplicial complex  $\Omega$  is a pair  $(V, S)$ , where

10. Antipodal shattering, under the same name, and antipodal VC, under the name *dual sign-rank*, was introduced in Alon et al. (2016). Their reasons for it, however, seem different from ours: As one can guess from the name, they defined a parameter dual to the sign-rank. However, “dual” here has a meaning different from ours.

$V \subseteq \mathbb{R}^d$  is a subset of  $\mathbb{R}^d$ , called the set of *vertices* of  $\Omega$ , and  $S$  is a set of simplices in  $\mathbb{R}^d$ , called the set of simplices of  $\Omega$ , such that:

- For every simplex  $\sigma \in S(\Omega)$ ,  $V(\sigma) \subseteq V(\Omega)$ ;
- Each face of every simplex  $\sigma \in S(\Omega)$  is a simplex of  $\Omega$ ;
- The intersection  $\sigma_1 \cap \sigma_2$  of any two simplices in  $S(\Omega)$  is a face of both  $\sigma_1$  and  $\sigma_2$ .

The *dimension* of  $\Omega$  is the maximal dimension of its simplices. The union of all simplices of  $\Omega$ , denoted  $\|\Omega\|$ , is called the *polyhedron* of  $\Omega$ .

A geometric simplicial complex  $\Omega$  is called a *geometric realization* of an abstract simplicial complex  $K$  if there is an isomorphism  $f: V(K) \rightarrow V(\Omega)$  such that  $\{f(s) \mid s \in S(K)\} = \{V(s) \mid s \in S(\Omega)\}$ . With some ambiguity, we then call the topological space  $\|\Omega\|$  a *geometric realization* of  $K$  and, for  $s \in K$ , define a *geometric realization* of  $s$ , denoted  $\|s\|$ , as a convex hull of the set  $\{f(v) \mid v \in s\}$ ; note that  $\|s\| \in S(\Omega)$ . It is well-known that for any two geometric realizations  $\Omega_1$  and  $\Omega_2$  of  $K$ ,  $\|\Omega_1\|$  is homeomorphic to  $\|\Omega_2\|$ , which enables us to talk about the geometric realization  $\|K\|$  of  $K$ .

A notion of the geometric simplicial complex is somewhat transitory, enabling us to unify the combinatorial information captured by the abstract simplicial complex  $K$  with the geometric information, captured by the space  $\|K\|$ . Thus, we will usually say simply *simplex* and *simplicial complex* to mean *abstract simplex* and *abstract simplicial complex*. For simplicial complexes  $K$  and  $L$ , a *simplicial map* is a map  $f: V(K) \rightarrow V(L)$  that maps simplices into simplices. Then  $f$  is canonically extended “by linearity” to a continuous map  $f: \|K\| \rightarrow \|L\|$ , where, as it is common, we denote the extended map by the same name  $f$ . It is also well-known and easy to check that the extension  $f: \|K\| \rightarrow \|L\|$  is an embedding (that is, one-to-one) if and only if  $f: V(K) \rightarrow V(L)$  is.

For a simplicial complex  $K$  and  $x \in \|K\|$ , let  $s \in S(K)$  be a simplex such that  $x \in \|s\|$ ; trivially, for every  $x$  there is always at least one such  $s$ . Then the *barycentric coordinates* of  $x$  in  $s$  is a unique function  $\alpha: V(s) \rightarrow [0, 1]$  such that  $\sum_{v \in V(s)} \alpha(v) = 1$  and  $\sum_{v \in V(s)} \alpha(v) \cdot v = x$ , where in the second sum  $v$ ’s are assumed to lie in the ambient space  $\mathbb{R}^d$ . It is easy to see that for  $x \in \|K\|$  there is a unique simplex  $s \in S(K)$  such that  $x \in \|s\|$  and  $x \notin \|t\|$  for any proper subsimplex  $t$  of  $s$ . This  $s$  is called the *support* of  $x$ , denoted  $s = \text{supp}(x)$ . It is also easy to see that  $s$  is the unique simplex containing  $x$ , such that all barycentric coordinates of  $x$  in  $s$  are nonzero.

Let us now have another round of discussion on antipodalities, with a partial overlap with Section 2. Recall that we define an *antipodal space* as a pair  $(X, \nu)$  where  $X$  is a topological space and  $\nu: X \rightarrow X$  is a continuous fixed-point free involution, that is,  $\nu \circ \nu(x) = x$  and  $\nu(x) \neq x$  for all  $x \in X$ . For antipodal spaces  $(X, \nu)$  and  $(Y, \omega)$ , a continuous map  $f: X \rightarrow Y$  is called *equivariant* if it commutes with antipodalities, that is, if  $f \circ \nu(x) = \omega \circ f(x)$  for all  $x \in X$ .

We say that  $(X, \nu)$  and  $(Y, \omega)$  are homeomorphic *as antipodal spaces*, denoted  $(X, \nu) \cong (Y, \omega)$ , if there is an equivariant homeomorphism between them, that is, if there is an isomorphism  $f: X \rightarrow Y$  such that both  $f$  and  $f^{-1}$  are continuous, and  $f \circ \nu(x) = \omega \circ f(x)$  for all  $x \in X$ . We note that this definition is a verbal repetition of Definition 5.2.1 in Matoušek et al. (2003), where antipodal space is called free  $\mathbb{Z}_2$ -space. We note that in the definition of a simplicial  $n$ -sphere  $\mathbb{Q}^n$  in Section 2, we required that  $(\|\mathbb{Q}^n\|, \nu)$  is homeomorphic to  $(\mathbb{S}^n, v \mapsto -v)$  as antipodal spaces. This detail is not negligible, as spheres, in general, admit non-standard antipodalities, see Hirsch and Milnor (1964).



That is,  $\mathbb{S}^n$  can be equipped with an antipodality  $\omega$  such that  $(\mathbb{S}^n, \omega) \not\cong (\mathbb{S}^n, v \mapsto -v)$ . In particular, it might be the case that  $\|\mathbb{Q}^n\| \cong \mathbb{S}^n$ , but  $(\|\mathbb{Q}^n\|, \nu) \not\cong (\mathbb{S}^n, v \mapsto -v)$ .

### A.3. Simplicial structure of $\Delta_{\mathcal{H}}$

Whenever the domain  $X$  of  $\mathcal{H}$  is finite,  $\Delta_{\mathcal{H}}$  is naturally a geometric simplicial complex with vertices  $X \times Y$  and maximal simplices  $\mathcal{H}$ , where  $h \in \mathcal{H}$  is treated as a subset of  $X \times Y$ : a distribution  $\mu \in \Delta_{\mathcal{H}}$  corresponds to a point in the geometric realization  $\|\text{supp}(\mu)\|$  of the simplex  $\text{supp}(\mu)$  with barycentric coordinates  $\{\mu(x, y) \mid (x, y) \in \text{supp}(\mu)\}$ <sup>11</sup>. In particular, the dimension of all maximal simplices of  $\Delta_{\mathcal{H}}$ , and, consequently, the dimension of  $\Delta_{\mathcal{H}}$  itself, is  $|X| - 1$ . Moreover,  $\Delta_{\mathcal{H}}$  consists of a finite number of simplices, and hence compact. Instead of doing a somewhat tedious formal proof of this, rather obvious, fact, we take a “proof by picture” approach and illustrate the simplicial structure of  $\Delta_{\mathcal{H}}$  in Figure 4 below.

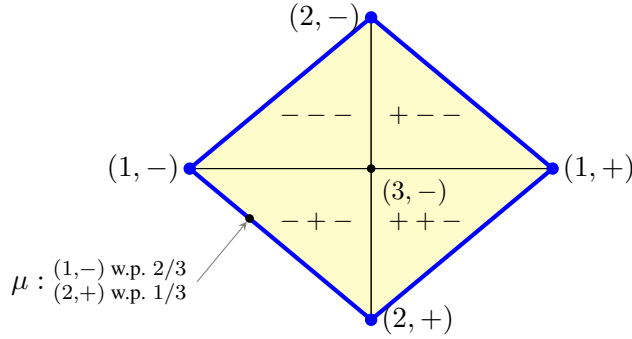


Figure 4: Simplicial complex  $\Delta_{\mathcal{H}}$  for a class  $\mathcal{H}$  on the domain  $\{1, 2, 3\}$  with 4 concepts:  $---$ ,  $-+-$ ,  $+--$ , and  $++-$ . The subcomplex  $\Delta_{\mathcal{H}}^{\text{ant}}$  is the boundary of  $\Delta(\mathcal{H})$ , drawn in thick blue lines.

It is also easy to see that  $\Delta_{\mathcal{H}}^{\text{ant}}$  is a (potentially empty) subcomplex of  $\Delta_{\mathcal{H}}$ . However, maximal simplices of  $\Delta_{\mathcal{H}}^{\text{ant}}$  no longer, in general, correspond to concepts of  $\mathcal{H}$ , in particular, the dimension of  $\Delta_{\mathcal{H}}^{\text{ant}}$  can be much smaller than  $|X| - 1$  (as in Figure 4 above). We also note that while typically we denote an abstract simplicial complex by  $K$  and its geometric realization by  $\|K\|$ , for historical consistency we do not use these “geometrization brackets” for  $\Delta_{\mathcal{H}}$ , and, with some ambiguity, use the same notation for it as for an abstract and geometrical simplicial complex.

### A.4. Discussion of spherical dimensions

Recall that for a class  $\mathcal{H}$ , we defined the spherical dimension  $\text{sd}(\mathcal{H})$  as a maximal  $n$  for which an  $n$ -sphere  $\mathbb{S}^n$  can be equivariantly mapped into  $\Delta^{\text{ant}}(\mathcal{H})$  that respects antipodalities. The notion of the spherical dimension is rather natural; in particular, in terms of Matoušek et al. (2003), Section 5,  $\text{sd}(\mathcal{H})$  is a  $\mathbb{Z}_2$ -coindex of  $\Delta^{\text{ant}}(\mathcal{H})$ . A typical application of  $\text{sd}$  then uses the BU theorem, or one of its relatives. Below we list our typical toolset for this approach.

11. Let us note the potential ambiguity: here  $\text{supp}(\mu)$  as defined in Section A.1, for  $\mu$  treated as a realizable distribution, coincides with  $\text{supp}(\mu)$  as defined in Section A.2, for  $\mu$  treated as a point in the geometric simplicial complex  $\Delta_{\mathcal{H}}$ .

**Theorem 21 (Borsuk-Ulam (BU) theorem, Borsuk (1933))** For  $0 \leq m < n$ , any map  $f: \mathbb{S}^n \rightarrow \mathbb{S}^m$  or  $f: \mathbb{S}^n \rightarrow \mathbb{R}^n$  collapses a pair of antipodal points.

**Theorem 22 (Lyusternik-Shnirel'man (LS) theorem, Lyusternik and Shnirel'man (1930))**

Let  $\mathcal{F}$  be an antipodal-free<sup>12</sup> cover of the  $n$ -dimensional sphere  $\mathbb{S}^n$ , for  $n \geq 0$ , such that every set  $A \in \mathcal{F}$  is either open or closed. Then,  $|\mathcal{F}| \geq n + 2$ . This is sharp, there is such a cover of size  $n + 2$ .

The above two classical theorems are equivalent to each other, and the statements of BU in terms of  $f: \mathbb{S}^n \rightarrow \mathbb{S}^m$  and  $f: \mathbb{S}^n \rightarrow \mathbb{R}^n$  are equivalent as well. Also, the following “local” versions are quite useful.

**Theorem 23 (Local BU, main theorem in Jaworowski (2000))** Let  $\mathbb{P}^d$  be a finite simplicial complex of dimension at most  $d$ . If, for  $n \geq 1$ ,  $f: \mathbb{S}^n \rightarrow \mathbb{P}^d$  is continuous and  $2d \leq n$  then there is  $x \in \mathbb{S}^n$  so that  $f(x) = f(-x)$ .

**Theorem 24 (Local LS, Theorem A in Chase et al. (2024b))** Let, for  $n \geq 1$ <sup>13</sup>,  $\mathcal{F}$  be a finite antipodal-free cover of the  $n$ -dimensional sphere  $\mathbb{S}^n$  such that either all sets in  $\mathcal{F}$  are open, or all are closed. Then there are at least  $\lceil (n + 3)/2 \rceil$  distinct sets in  $\mathcal{F}$  with nonempty intersection, and this bound is sharp.

An alternative way of stating a conclusion of Theorem 24 is to say that the *overlap-degree* of  $\mathcal{F}$  is at least  $\lceil (n + 3)/2 \rceil$ . Also note that, unlike in the original LS, the statement is no longer true if some sets in  $\mathcal{F}$  are open and some are closed. In fact, in Chase et al. (2024b) it is shown that for open/closed case the overlap is at least  $\lceil t/2 \rceil$  for  $t = \lceil (n + 3)/2 \rceil$ , and that this bound is sharp. We also note that Theorem 24 (Local LS), except for the bounds in open/closed case, follows, in a straightforward way, from Theorem 23 (Local BU). We finally note that the list of the four theorems above is not exhaustive, for example, in Chornomaz et al. (2024) the authors used a version of BU for closed convex relations (their Theorem 7) or, in Chase et al. (2024a), a “local” version of a topological Radon theorem (their Lemma 12).

Recall that out of the three definitions of admitting an infinitely supported / finitely supported / simplicial sphere, we picked the finitely supported as the standard one. Putting the finitely supported spheres in the center is justified by the fact that most of the important dimensions, like VC and VC\*, are finitely supported, that is, witnessed by projections to finite subdomains. Moreover, in most of the cases, for example for list replicability (Section 5) or extremal classes (Section 4), we can apply our tools only to finitely supported classes. Note also that the distinction between finitely and infinitely supported spheres vanishes if the class  $\mathcal{H}$  itself is finitely supported; in particular, whenever we upper bound spherical dimension for finitely supported classes, we would often state the bound for  $\text{sd}$ , implying that it trivially extends to a “formally stronger” bound on  $\text{sd}_\infty$ .

Further restricting the spheres to be simplicial is important to our results on disambiguations (Section 3), and can also be considered natural from the standpoint of simplicial geometry. We also

12. That is, for all  $F \in \mathcal{F}$ ,  $x \in F$  implies  $-x \notin F$ .

13. Both Theorem 23 and Theorem 24 are incorrect for  $n = 0$ : The identity map on  $\mathbb{S}^0$  is non-collapsing, and  $\mathbb{S}^0$  can be covered by a non-overlapping open family, in which case the overlap is  $1 < \lceil (0 + 3)/2 \rceil = 2$ . This detail was overlooked in the corresponding statement in Chase et al. (2024b), which leads to minor incorrectness in their Theorem D. We elaborate upon it in Section 5.

note that the spheres witnessing the  $\text{sd} \geq \text{VC} - 1$  and  $\text{sd} \geq \text{VC}^{*a} - 2$  lower bounds in Lemma 2, as well as other lower bounds on  $\text{sd}$  in Lemma 3, are all simplicial.

Simply based on the fact that admitting a finitely supported sphere is more restrictive than admitting an infinitely supported one, and admitting a simplicial sphere is furthermore restrictive, we have  $\text{sd}_{\text{simp}} \leq \text{sd} \leq \text{sd}_{\infty}$ . Although we think that the three spherical dimensions can be separated from each other on the level of individual classes, the difference between them might be insubstantial with respect to Conjectures 14–16. To this end, we say that parameters  $\text{sd}', \text{sd}'' \in \{\text{sd}_{\infty}, \text{sd}, \text{sd}_{\text{simp}}\}$  are *roughly equivalent*, if for any integers  $a$  and  $b$ , there is a class  $\mathcal{H}'$  with  $\text{VC}(\mathcal{H}') \leq a$ ,  $\text{VC}^{*a}(\mathcal{H}') \leq b$ , and  $\text{sd}'(\mathcal{H}') \geq n$  if and only if there is a class  $\mathcal{H}''$  with  $\text{VC}(\mathcal{H}'') \leq a$ ,  $\text{VC}^{*a}(\mathcal{H}'') \leq b$ , and  $\text{sd}''(\mathcal{H}'') \geq n$ . From this perspective, our main result on disambiguations (Theorem 4 in Section 3) establishes rough equivalence between  $\text{sd}_{\text{simp}}$  and yet another parameter called *disambiguation index*  $\text{disamb}$  (although we will not really use this term and will only formally introduce it in the proof of Theorem 4).

We finally note two possible modifications to the notion of spherical dimension which we find potentially worth considering, but which we excluded from the main narrative for the sake of brevity. On one hand, we can restrict these definitions by requiring that the respective equivariant map from  $\mathbb{S}^n$  to  $\Delta^{\text{ant}}$  be an embedding (in which case we call the corresponding sphere *embedded*). But, apart from the psychological comfort of thinking about the respective spheres as “sitting inside”  $\Delta^{\text{ant}}$ , we see no other benefits to this modification. We note, however, that all the spheres in Lemma 2 and Lemma 3 are embedded.

On the other end, we can relax the three notions by letting the spheres be homology spheres or have non-standard antipodalities. The benefit of this would be that it corresponds to the natural domain of applicability of the BU-like theorems (Theorems 21–24) that are used to infer properties of the respective classes. Most of our results would indeed hold with this relaxation. However, it might be considered rather exotic, dragging in heavier topological machinery, while we do not have examples to substantiate the benefit of this approach.

## A.5. Barycentric subdivisions

We refer to Definition 1.7.2 and the subsequent discussion in Matoušek et al. (2003).

For an abstract simplicial complex  $\mathbb{Q}$  and a simplex  $s \in \mathbb{Q}$ , the *relative interior* of  $s$  is a subset of  $\|\mathbb{Q}\|$  defined as  $\|s\| - \bigcup_{t \subsetneq s} \|t\|$  (Definition 1.3.4 in Matoušek et al. (2003)). Relative interiors form a partition of  $\|\mathbb{Q}\|$  and so for every  $x \in \|\mathbb{Q}\|$  there is a unique simplex  $s$  such that  $x$  is in the relative interior of  $s$ ; it is easy to see that this simplex is exactly the support of  $x$ , as defined in Section A.2.

For an abstract simplicial complex  $\mathbb{Q}$ , the *first barycentric subdivision*  $\mathbb{Q}_1$  of  $\mathbb{Q}$  is defined as an abstract simplicial complex whose vertices are non-empty simplices of  $\mathbb{Q}$ ,  $V(\mathbb{Q}_1) = S^{>0}(\mathbb{Q})$ , and whose simplices are chains of simplices of  $S^{>0}(\mathbb{Q})$  (where chain is with respect to the order on  $S^{>0}(\mathbb{Q})$  by inclusion).

For a nonempty simplex  $s \in S(\mathbb{Q})$ , let us define  $v_s \in \|s\|$  as the (geometric) *barycenter* of  $\|s\|$ , that is,  $v_s = \frac{1}{|s|} \sum_{v \in s} v$ . The map  $\chi: \|\mathbb{Q}_1\| \rightarrow \|\mathbb{Q}\|$  defined on  $V(\mathbb{Q}_1) = S^{>0}(\mathbb{Q})$  as  $s \mapsto v_s$  and the extended by linearity to  $\|\mathbb{Q}_1\|$  establishes a homeomorphism between  $\|\mathbb{Q}_1\|$  and  $\|\mathbb{Q}\|$ . In particular, the geometric barycentric subdivision preserves the topology of the space,  $\|\mathbb{Q}_1\| \cong \|\mathbb{Q}\|$ .

Furthermore, if  $(\mathbb{Q}, \nu)$  is an antipodal simplicial complex (as defined in Section 2), then  $\nu$  can be extended from  $V(\mathbb{Q})$  to  $\nu_1$  on  $V(\mathbb{Q}_1)$  by letting  $\nu_1(s) = \nu(s) = \{\nu(v) \mid v \in s\}$ , for any  $s \in V(\mathbb{Q}_1) = S^{>0}(\mathbb{Q})$ . It is well-known and easy to check that  $\nu_1$  is an antipodality map, that is,

satisfies i), ii), and iii), and that  $\nu_1$ , restricted to  $V(\mathbb{Q})$  coincides with  $\nu$  (where we assume that  $V(\mathbb{Q})$  is embedded into  $V(\mathbb{Q}_1)$  by  $v \mapsto \{v\}$ ). Moreover,  $\nu_1$ , extended by linearity to  $\|\mathbb{Q}_1\|$  coincides, via the homeomorphism  $\chi$ , with  $\nu$  extended by linearity to  $\|\mathbb{Q}\|$ , that is,  $\chi \circ \nu_1(x) = \nu \circ \chi(x)$  for all  $x \in \|\mathbb{Q}_1\|$ . In particular,  $(\|\mathbb{Q}_1\|, \nu_1) \cong (\|\mathbb{Q}\|, \nu)$ . Because of this, we will call  $\nu_1$  simply by  $\nu$ . This is illustrated in Figure 5 below.

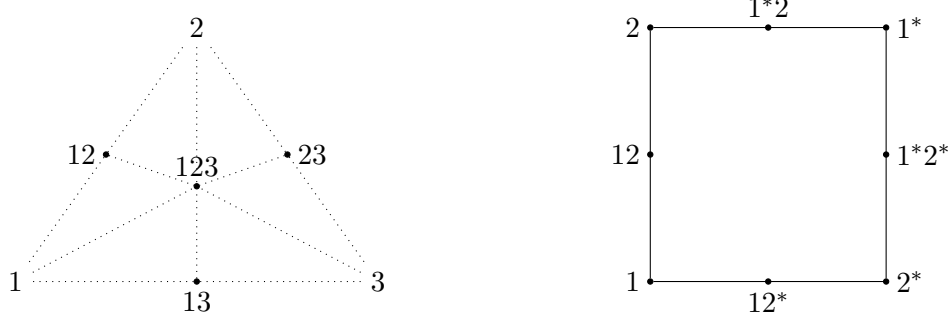


Figure 5: Barycentric subdivision of a 2-dimensional simplex  $\{1, 2, 3\}$  and a barycentric subdivision of a 1-dimensional simplicial complex  $\{12, 1^*2, 1^*2^*, 12^*\}$  that extends the antipodality map  $v \mapsto v^*$ .

One important, in general, as well as for us, example of a barycentric subdivision is the barycentric subdivision of the boundary of a single simplex.

#### Example 1 (Barycentric subdivision of the boundary of a simplex)

The barycentric subdivision  $\mathbb{B}^n$  of the boundary of an  $(n + 1)$ -dimensional simplex  $\Delta_{n+1}$  is a simplicial complex whose vertices are all nontrivial subsets of  $[n + 2]$ , that is all except for  $\emptyset$  and  $[n + 2]$ , and whose simplices are chains of nontrivial subsets of  $[n + 2]$ . Thus defined,  $\|\mathbb{B}^n\|$  is homeomorphic to  $\mathbb{S}^n$ .

Moreover, the complementation map  $S \mapsto [n + 2] - S$  on  $V(\mathbb{B}^n)$  is an antipodality map for  $\mathbb{B}^n$ , and  $(\|\mathbb{B}^n\|, S \mapsto [n + 2] - S) \cong (\mathbb{S}^n, x \mapsto -x)$ .

Finally, let us note that  $\|(\mathbb{Q}, \nu)\| \cong \|(\mathbb{Q}_1, \nu)\|$  enables us to iterate barycentric subdivisions without changing the topology of the space. This is convenient in the following sense: Assuming some metric on  $\|\mathbb{Q}\|$ , let us define the (simplicial) diameter of  $\mathbb{Q}$  as  $\max_{s \in S(\mathbb{Q})} \max_{x, y \in \|s\|} d(x, y)$ , denoted  $\text{diam}(\mathbb{Q})$ .

**Proposition 25** *Let  $\mathbb{Q}_k$ , for  $k \geq 0$ , be  $k$ 'th barycentric subdivision of a finite simplicial complex  $\mathbb{Q}$ , where  $\mathbb{Q} = \mathbb{Q}_0$ . Then the sequence  $\text{diam}(\mathbb{Q}_k)$ , for  $k \geq 0$ , is non-increasing and goes to 0.*

This property is Exercise 3 in Section 1 in [Matoušek et al. \(2003\)](#). Proof of Proposition 2.21 in [Hatcher \(2002\)](#) gives an explicit rate of convergence.

#### A.6. Joins

We follow Section 4.2 in [Matoušek et al. \(2003\)](#) in the development of the respective terminology. In particular, the definitions of joins of simplicial complexes and topological spaces are Definitions 4.2.1 and 4.2.3 there, and Example 5.2.3 defines the join of antipodal spaces.

For simplicial complexes  $K$  and  $L$ , the *join* of  $K$  and  $L$ , denoted  $K * L$ , is the simplicial complex with vertex set  $V(K) \sqcup V(L)$  and with the set of simplices  $\{s \sqcup k \mid s \in S, k \in K\}$ . For topological spaces  $X$  and  $Y$ , their *join*, denoted  $X * Y$  is the topological space  $X \times Y \times [0, 1] / \approx$ , where  $\approx$  is an equivalence relation given by  $(x, y, 0) \approx (x', y, 0)$  and  $(x, y, 1) \approx (x, y', 1)$ , for all  $x, x' \in X$  and  $y, y' \in Y$ . The point  $v = (x, y, \alpha) \in X * Y$  is often written as a formal linear combination  $v = \alpha x + \beta y$ , for  $\beta = 1 - \alpha$ . This notation is consistent with the fact that  $0x + 1y$  and  $0x + 1y'$  define the same point in the join, and we will call  $\alpha, x$  and  $\beta, y$  the *barycentric coordinates* of  $v$  in  $X * Y$ . It is known and easy to check that the definitions of joins for simplicial complexes and for topological spaces are consistent with each other, namely,  $\|K\| * \|L\| \cong \|K * L\|$ .

Joins can also be defined for maps: for  $f_1: X_1 \rightarrow Y_1$  and  $f_2: X_2 \rightarrow Y_2$ , the map  $f_1 * f_2: X_1 * X_2 \rightarrow Y_1 * Y_2$  is defined as  $\alpha x_1 + \beta x_2 \mapsto \alpha f_1(x_1) + \beta f_2(x_2)$ . In a similar way, joins can be defined for simplicial maps and this definition is consistent with the definition for the continuous maps above. This enables us to define the join of antipodal spaces  $(X_1, \nu_1)$  and  $(X_2, \nu_2)$  as  $(X_1 * X_2, \nu_1 * \nu_2)$ . It is easy to check that  $\nu_1 * \nu_2$  will indeed be a fixed-point free involution whenever  $\nu_1$  and  $\nu_2$  are. In the same way, this can also be done for joins of antipodal simplicial complexes. We also note that the join operation is trivially associative, and can thus be easily generalized to joins of any finite number of spaces or simplicial complexes.

The  $n$ -fold joins, also called *join powers*, give us another important example of simplicial spheres. The following example is based on Definition 1.4.1 and Examples 4.2.2 and 5.2.4 from [Matoušek et al. \(2003\)](#).

**Example 2 (Boundary of a crosspolytope)** Let  $\mathbb{D}_0$  be the simplicial complex of a 2-point discrete space, that is,  $V(\mathbb{D}_0) = \{-, +\}$  and  $S(\mathbb{D}_0) = \{\emptyset, -, +\}$ . We also assume  $\mathbb{D}_0$  to be antipodal with the vertex antipodality map  $\nu_0$  exchanging  $-$  and  $+$ .

Let  $\mathbb{D}_n$  be the  $(n + 1)$ -fold join of  $\mathbb{D}_0$ , denoted  $\mathbb{D}_n = \mathbb{D}_0^{*(n+1)}$ . Then  $\mathbb{D}_n$  can be explicitly described as a simplicial complex with vertices  $V(\mathbb{D}_n) = \{(i, -), (i, +) \mid i \in [n + 1]\}$ , whose simplices are precisely the subsets  $S \subseteq [n + 1] \times \{-, +\}$  such that for no  $i \in [n + 1]$  both  $(i, -)$  and  $(i, +)$  are in  $S$ . The vertex antipodality map  $\nu_n$  is defined by sign swap, that is,  $\nu_n(i, -) = (i, +)$  and  $\nu_n(i, +) = (i, -)$ .

This construction is known as the boundary of an  $(n + 1)$ -dimensional crosspolytope. In particular,  $\|\mathbb{D}_n\| \cong (S^n, x \mapsto -x)$ .

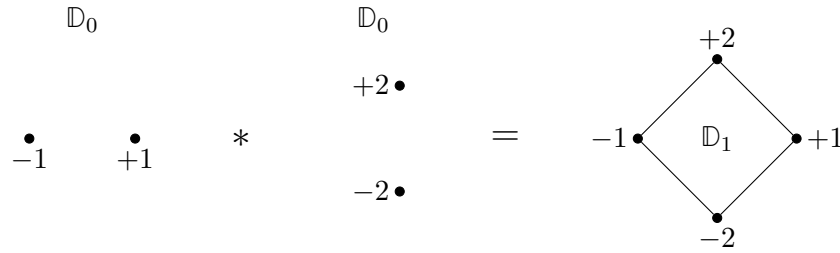


Figure 6: Construction of  $\mathbb{D}_1$  as a join of  $\mathbb{D}_0$  with  $\mathbb{D}_0$ . Note that for the join, we need to make the vertex sets of the two copies of  $\mathbb{D}_0$  disjoint, so we rename  $-$  and  $+$  to  $-1, +1$  for the first copy and  $-2, +2$  for the second one.

A particularly useful property, which can also be considered a corollary of the above Example 2, is the following:

**Proposition 26** *For antipodal spheres  $\mathbb{S}^n$  and  $\mathbb{S}^m$ ,  $\mathbb{S}^n * \mathbb{S}^m \cong \mathbb{S}^{n+m+1}$ .*

## Appendix B. Proofs for basic results

### B.1. Order and equivalence of classes

If a class  $\mathcal{H}_1$  is obtained from  $\mathcal{H}_2$  by restricting to a subdomain or by removing some hypotheses, it is natural to think of  $\mathcal{H}_2$  as being, in a sense, “simpler”. Let us briefly formalize this intuition as follows: For classes  $\mathcal{H}_1$  over  $X_1$  and  $\mathcal{H}_2$  over  $X_2$  we write  $\mathcal{H}_1 \leq \mathcal{H}_2$  if there are functions  $\varphi: X_1 \rightarrow X_2$  and  $\sigma: \mathcal{H}_1 \rightarrow \mathcal{H}_2$  such that  $\sigma(h)(\varphi(x)) = h(x)$  for all  $h \in \mathcal{H}_1$  and  $x \in X_1$ . We write  $\mathcal{H}_1 \equiv \mathcal{H}_2$  if  $\mathcal{H}_1 \leq \mathcal{H}_2$  and  $\mathcal{H}_2 \leq \mathcal{H}_1$ . It is easy to see that  $\leq$  is a preorder and  $\equiv$  is the respective equivalence relation.

It is easy to see that VC, VC\*, and VC<sup>\*a</sup> dimensions are all monotone with respect to this order. While it might be clear on the spot that the same is true for the spherical dimensions, the proof of it is also easy:

**Proposition 27** *If  $\mathcal{H}_1 \leq \mathcal{H}_2$ , then  $\text{sd}_\infty(\mathcal{H}_1) \leq \text{sd}_\infty(\mathcal{H}_2)$ ,  $\text{sd}(\mathcal{H}_1) \leq \text{sd}(\mathcal{H}_2)$ , and  $\text{sd}_{\text{simp}}(\mathcal{H}_1) \leq \text{sd}_{\text{simp}}(\mathcal{H}_2)$ .*

**Proof** Let  $\varphi: X_1 \rightarrow X_2$  and  $\sigma: \mathcal{H}_1 \rightarrow \mathcal{H}_2$  be the functions witnessing  $\mathcal{H}_1 \leq \mathcal{H}_2$ . And let  $f: \mathbb{S}^n \rightarrow \Delta_{\mathcal{H}_1}$  be the equivariant map witnessing  $\text{sd}_\infty(\mathcal{H}_1) \geq n$ . Let  $\bar{\varphi}$  be the lifting of  $\varphi$  to the set of (not necessarily realizable) distributions over  $X_1 \times Y$ . That is, for a distribution  $\varphi$  over  $X_1 \times Y$  and an event  $E \subseteq X_2 \times Y$ , we define  $\mathbb{P}_{\bar{\varphi}(\mu)}(E) = \mathbb{P}_\mu(\varphi^{-1}(E))$ . It is easy to see that thus defined,  $\bar{\varphi}(\mu)$  is indeed a probability distribution over  $X_2 \times Y$ . Moreover, we claim that  $\bar{\varphi}(\mu)$  is realizable by  $\mathcal{H}_2$  whenever  $\mu$  is realizable by  $\mathcal{H}_1$ , and so  $\varphi$  maps  $\Delta_{\mathcal{H}_1}$  into  $\Delta_{\mathcal{H}_2}$ .

Indeed, let, for arbitrary  $\varepsilon > 0$ ,  $h_1 \in \mathcal{H}_1$  be a concept with  $L_\mu(h_1) = \mathbb{P}_{(x,y) \sim \mu} [h_1(x) \neq y] \leq \varepsilon$ . Then

$$\begin{aligned} L_{\bar{\varphi}(\mu)}(\sigma(h_1)) &= \mathbb{P}_{(x,y) \sim \bar{\varphi}(\mu)} [\sigma(h_1)(x) \neq y] \\ &= \mathbb{P}_{(x,y) \sim \mu} [\sigma(h_1)(\varphi(x)) \neq y] \\ &= \mathbb{P}_{(x,y) \sim \mu} [h_1(x) \neq y] = L_\mu(h_1) \leq \varepsilon, \end{aligned}$$

as needed. Finally, we are going to show that the map  $\bar{\varphi} \circ \Delta_{\mathcal{H}_1} \rightarrow \Delta_{\mathcal{H}_2}$  is equivariant.. It then easily follows that  $\bar{\varphi} \circ f$  witnesses  $\text{sd}_\infty(\mathcal{H}_2) \geq n$ . Indeed, for every  $E \subseteq X_2 \times Y$  we have:

$$\begin{aligned} \mathbb{P}_{\bar{\varphi}(-\mu)}(E) &= \mathbb{P}_{-\mu}(\varphi^{-1}(E)) \\ &= \mathbb{P}_{(x,y) \sim -\mu} [(\varphi(x), y) \in E] = \mathbb{P}_{(x,y) \sim \mu} [(\varphi(x), -y) \in E] \\ &= \mathbb{P}_{(x,y) \sim \mu} [(\varphi(x), y) \in -E] = \mathbb{P}_{-\bar{\varphi}(\mu)}(E), \end{aligned}$$

and so  $\bar{\varphi}(-\mu) = -\bar{\varphi}(\mu)$ , as needed. This finishes the proof that  $\text{sd}_\infty(\mathcal{H}_1) \leq \text{sd}_\infty(\mathcal{H}_2)$ .

The regular sd, that is, the finitely supported case, follows by observing that for any finitely supported restriction  $\mathcal{H}'_1$  of  $\mathcal{H}_1$ , there is a finitely supported restriction  $\mathcal{H}'_2$  of  $\mathcal{H}_2$ ; if  $\mathcal{H}'_1$  is restricted to  $S$ , then it is enough to take a restriction of  $\mathcal{H}_2$  to  $\varphi(S)$ .

Finally, the simplicial case  $\text{sd}_{\text{simp}}$  follows by observing that the equivariant map  $\bar{\varphi}: \Delta_{\mathcal{H}_1} \rightarrow \Delta_{\mathcal{H}_2}$  is trivially simplicial. ■



**Corollary 28** *For a finite class  $\mathcal{H}$ ,  $\text{sd}_\infty(\mathcal{H}) = \text{sd}(\mathcal{H})$ .*

Note that it resolves some tensions in the definition of  $\text{sd}$ : In Definition 1, we effectively defined it as  $\text{sd}_\infty$ , but only for finite classes. While for finitely supported  $\mathcal{H}$ ,  $\text{sd}_\infty(\mathcal{H}) = \text{sd}(\mathcal{H})$ , this definition aligns with how  $\text{sd}$  was later defined in Definition 5, but only for finitely supported  $\mathcal{H}$ . For finite classes, this equality follows from the present corollary.

**Proof** Let  $\mathcal{H}'$  be a class on the domain of equivalence classes over  $X$  for the equivalence relation  $x \sim y$  if  $x(h) = y(h)$ . It is easy to see that  $\mathcal{H}' \equiv \mathcal{H}$  and that  $\mathcal{H}'$  is finitely supported. But then:

$$\text{sd}(\mathcal{H}) = \text{sd}(\mathcal{H}') = \text{sd}_\infty(\mathcal{H}') = \text{sd}_\infty(\mathcal{H}).$$

■

We further note that such ordering on classes will come in handy in Section C, where we will often see geometrically constructed classes that are naturally defined with a great deal of redundancy.

We find it convenient to explicate some parameters of the minimal classes witnessing  $\text{VC} \geq n$  and  $\text{VC}^* \geq n$ , that is for the classes  $\mathcal{C}_n$  of *binary hypercubes* and *universal classes*  $\mathcal{U}_n = \mathcal{C}_n^*$ . Formally,  $\mathcal{C}_n$  is a class of all  $2^n$  concepts on  $[n]$ , and  $\mathcal{U}_n$  is its dual, but can also be described explicitly as a class of  $n$  indicator functions on the domain of all  $2^n$  subsets of  $[n]$ .

Figure 7 provides some exact values for the families  $\mathcal{C}_n$  and  $\mathcal{U}_n$ , which are sometimes convenient for sanity checks. Most will be substantiated in the full statement of Lemma 2 in Section B.2.

$n$	$ \mathcal{C}^n $ $2^n$	$\text{VC}(\mathcal{C}^n)$ $n$	$\text{VC}^*(\mathcal{C}^n)$ $\lfloor \log n \rfloor$	$\text{sd}(\mathcal{C}_n)$ $n - 1$	$ \mathcal{U}^n $ $n$	$\text{VC}(\mathcal{U}^n)$ $\lfloor \log n \rfloor$	$\text{VC}^*(\mathcal{U}^n)$ $n$	$\text{sd}(\mathcal{U}_n)$ $\geq n - 2$	$\text{sd}(\mathcal{U}_n)$ $\leq$
1	2	1	0	0	1	0	1	-1	-1
2	4	2	1	1	2	1	2	0	1
3	8	3	1	2	3	1	3	1	1
4	16	4	2	3	4	2	4	2	3
5	32	5	2	4	5	2	5	3	3
6	64	6	2	5	6	2	6	4	5

Figure 7: Certain parameters of classes  $\mathcal{C}_n$  and  $\mathcal{U}_n$ .

## B.2. Certain bounds on spherical dimension and related parameters

**Lemma 20** (Assouad’s bounds for antipodal shattering) *For a (partial) class  $\mathcal{H}$ , it holds*

$$\lfloor \log \text{VC}(\mathcal{H}) \rfloor \leq \lfloor \log \text{VC}^{\text{a}}(\mathcal{H}) \rfloor \leq \text{VC}^*(\mathcal{H}) \leq \text{VC}^{*\text{a}}(\mathcal{H}) \leq 2^{\text{VC}(\mathcal{H})+1} - 1,$$

and

$$\text{VC}^*(\mathcal{H}) \leq \text{VC}^{*\text{a}}(\mathcal{H}) \leq 2\text{VC}^*(\mathcal{H}) + 1.$$

Moreover, all inequalities between the neighboring values above are sharp.

We note that the bound  $\text{VC} \leq \text{VC}^{*a} \leq 2\text{VC}^* + 1$  was already proven in [Alon et al. \(2016\)](#) (Appendix E). The refinement of the Assouad’s bounds in terms of antipodal shattering, to our knowledge, is new.

**Proof** In the first chain of inequalities, the first and the third are trivial, and the second and the fourth are equivalent “by duality” by noting that for integers  $a$  and  $b$ ,  $\lfloor \log a \rfloor \leq b$  is equivalent to  $a \leq 2^{b+1} - 1$ . We will thus prove the second inequality.

So, assuming  $\mathcal{H}$  antipodally shatters a set  $S$  of size  $2^d$ , we need to prove that it dually shatters a  $d$ -set of concepts  $H \subseteq \mathcal{H}$ . Let us index the elements of  $S$  by the  $2^d$  patterns on  $[d]$ ,  $S = \{s_t \mid t \in [d] \rightarrow \{-, +\}\}$ . Let  $\tau_i: S \rightarrow \{-, +\}$ , for  $i \in [d]$ , be a pattern on  $S$  defined as  $\tau_i(s_t) = t(i)$ , as illustrated in the table below. Let  $h_i \in \mathcal{H}$  be the concept witnessing either  $\tau_i$  or  $-\tau_i$ , and let  $q: [d] \rightarrow \{-, +\}$  on  $i$  be  $+$  if it is the former, and  $-$  if it is the latter. That is,  $h_i(s_t) = q(i) \cdot \tau_i(s_t)$ , for  $s_t \in S$ .

$S$	---	--+	-+-	-++	+-	++	+++
$\tau_1$	-	-	-	-	+	+	+
$\tau_2$	-	-	+	+	-	-	+
$\tau_3$	-	+	-	+	-	+	+

We claim that  $S$  dually shatters  $H = \{h_i \mid i \in [d]\}$ . Let  $r: [d] \rightarrow \{-, +\}$  be a dual pattern on  $H$ . Then  $s_{q \cdot r}$ , where  $q \cdot r: [d] \rightarrow \{-, +\}$  is a pointwise product, witnesses  $r$ . Indeed,  $h_i(s_{q \cdot r}) = q(i) \cdot \tau_i(s_{q \cdot r}) = q(i) \cdot q(i) \cdot r(i) = r(i)$ , for all  $i \in [d]$ , as needed. The sharpness of the bounds trivially follows from the sharpness of the original bounds by Assouad.

In the second set of inequalities, we only need to prove  $\text{VC}^{*a}(\mathcal{H}) \leq 2\text{VC}^*(\mathcal{H}) + 1$ . We prove it in the form  $\text{VC}^a(\mathcal{H}) \leq 2\text{VC}(\mathcal{H}) + 1$ : these statements are clearly equivalent, just for applications we care about  $\text{VC}^{*a}$  and not  $\text{VC}^a$ . So let  $\text{VC}(\mathcal{H}) = d$  and let  $S$  be a set of size  $n$ , antipodally shattered by  $\mathcal{H}$ . It is clear that then  $S$  is shattered by  $\mathcal{H}' = \mathcal{H} \cup -\mathcal{H}$ , where  $-\mathcal{H} = \{-h : h \in \mathcal{H}\}$ . Denote  $n = \text{VC}^a(\mathcal{H})$ ,  $d = \text{VC}(\mathcal{H})$ . Then, using the Sauer–Shelah bound and some straightforward manipulation with binomial coefficients, we get

$$\sum_{k=0}^n \binom{n}{k} = 2^n = |\mathcal{H}'|_S \leq 2|\mathcal{H}|_S \leq 2 \sum_{k=0}^d \binom{n}{k} = \sum_{k=0}^d \binom{n}{k} + \sum_{k=n-d}^n \binom{n}{k}.$$

But this implies  $n - d \leq d + 1 \sim n \leq 2d + 1$ , implying  $\text{VC}^a(\mathcal{H}) \leq 2\text{VC}(\mathcal{H}) + 1$ , as needed.

The sharpness of the bound is witnessed by a class of all  $(\leq d)$ -subsets of a  $(d + 1)$ -set  $S$ . This class antipodally shatters  $S$ , but shatters only  $(\leq d)$ -sets.  $\blacksquare$

**Lemma 2** For a class  $\mathcal{H}$ ,  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq \text{VC}(\mathcal{H}) - 1$  and  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq \text{VC}^{*a}(\mathcal{H}) - 2$ . Moreover:

- $\text{VC}(\mathcal{C}_n) = n$ ,  $\text{VC}^*(\mathcal{C}_n) = \lfloor \log n \rfloor$ , and  $\text{sd}(\mathcal{C}_n) = n - 1$ ;
- $\text{VC}(\mathcal{U}_n) = \lfloor \log n \rfloor$ ,  $\text{VC}^*(\mathcal{U}_n) = n$ , and  $n - 2 \leq \text{sd}(\mathcal{U}_n) \leq n - 2 + \delta_n$ , where  $\delta_n = 1$  for  $n$  even and 0 for  $n$  odd. In particular,  $\text{sd}(\mathcal{U}_n) = n - 2$  for  $n$  odd.

The statements about  $\mathcal{C}_n$  and  $\mathcal{U}_n$  are true for all three spherical dimensions  $\text{sd}_{\infty}$ ,  $\text{sd}$ , or  $\text{sd}_{\text{simp}}$ .

**Proof** For the class  $\mathcal{C}_n$  the simplices of  $\Delta_{\mathcal{C}_n}$  are precisely the subsets  $F \subseteq X \times Y$  such that for no  $x \in X$  both  $(x, -)$  and  $(x, +)$  are in  $F$ . Note that this is precisely the boundary of  $n$ -dimensional crosspolytope from Example 2. In particular,  $\Delta_{\mathcal{C}_n} = \Delta_{\mathcal{C}_n}^{\text{ant}} \cong (\mathbb{S}^{n-1}, x \mapsto -x)$ , hence  $\text{sd}_{\text{simp}}(\mathcal{C}_n) \geq n - 1$ . This also witnesses that  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq \text{VC}(\mathcal{H}) - 1$  for an arbitrary class  $\mathcal{H}$ . The fact that  $\text{sd}(\mathcal{C}^n) \leq n - 1$  follows by the Borsuk-Ulam theorem (Theorem 21) as, up to homeomorphism,  $\Delta_{\mathcal{C}_n}^{\text{ant}}$  is exactly  $\mathbb{S}^{n-1}$ .

To prove that  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq \text{VC}^{\text{a}}(\mathcal{H}) - 2$ , we can assume, without losing generality, that  $\mathcal{H}$  is a minimal class witnessing  $\text{VC}^{\text{a}}(\mathcal{H}) \geq n$ , for  $n \geq 2$ . Then the domain  $X$  can be associated with the subset of  $2^{[n]}$  such that for every  $S \subseteq [n]$  exactly one of  $S$  and  $-S = [n] - S$  is in  $X$ , and  $\mathcal{H}$  is a set of  $n$  concepts  $h_i$  for  $i = 1, \dots, n$ , where  $h_i = [i \in S]$  is the indicator function of  $i$  in  $S$ , that is,  $h_i(S) = +$  if  $i \in S$  and  $-$  otherwise.

Let  $\mathbb{B}^{n-2}$  be the barycentric subdivision of the boundary of  $(n - 1)$ -dimensional simplex from Example 1. Recall that the vertices of  $\mathbb{B}^{n-2}$  are all subsets of  $[n]$ , except for  $\emptyset$  and  $[n]$ , simplices are chains of such subsets, and the antipodality map is the complementation  $S \mapsto [n] - S$ . Let us define the map  $f: V(\mathbb{B}^{n-2}) \rightarrow V(\Delta_{\mathcal{H}}) \subseteq X \times Y$  as  $f(S) = ([S \in X]S, [S \in X])$ . For example, for  $n = 3$  and the set  $12 = \{1, 2\}$ ,  $f(12) = (12, +)$  if  $12 \in X$  and  $(-12, -) = (3, -)$  otherwise. Note that  $f$  is trivially one-to-one, moreover, it commutes with the antipodality:  $f(-S) = ([-S \in X](-S), [-S \in X]) = ([S \in X]S, -[S \in X]) = -f(S)$ . Thus, for  $f$  to witness an embedded simplicial  $(n - 2)$ -sphere in  $\Delta_{\mathcal{H}}^{\text{ant}}$  it is enough to show that it is simplicial. For that, let  $\sigma = S_1 \subsetneq \dots \subsetneq S_k$  be a chain of subsets of  $[n]$  such that  $S_1 \neq \emptyset$  and  $S_k \neq [n]$ ; in other words,  $\sigma$  is a simplex of  $\mathbb{B}^{n-2}$ . Let  $i \in S_1$ . We claim that  $h_i$  witnesses  $f(\sigma) \in S(\Delta_{\mathcal{H}})$ , that is, that  $h_i(x) = y$  for all  $(x, y) = f(S_j)$ ,  $j = 1, \dots, k$ . But indeed,  $f(S_j) = ([S_j \in X]S_j, [S_j \in X])$ , and  $h_i([S_j \in X]S_j) = [i \in [S_j \in X]S_j] = [S_j \in X]$ , as needed, where we have used the fact that  $i \in S_j$ .

The  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq \text{VC}^{\text{a}}(\mathcal{H}) - 2$  bound implies  $\text{sd}_{\text{simp}}(\mathcal{U}_n) \geq n - 2$ . The upper-bound  $\text{sd}(\mathcal{U}_n) \leq n - 2 + \delta_n$  follows from

$$\lceil (\text{sd}(\mathcal{U}_n) + 3)/2 \rceil \leq \text{LR}(\mathcal{U}_n, \varepsilon) \leq \text{LR}(\mathcal{U}_n) \leq \lceil (n + 1)/2 \rceil,$$

where the first inequality is by Theorem 10 and the last one is by Theorem E in Chase et al. (2024b), where  $\delta_n$  appears from removing the rounding. Note also that Theorem 10 does not apply whenever  $\text{sd}(\mathcal{U}_n) \leq 0$ , which can only happen for  $n = 1$  or  $2$ . For  $n = 1$ ,  $\mathcal{U}_1$  has just one concept and hence, by Lemma 31,  $\text{sd}(\mathcal{U}_1) = -1 = 1 - 2 + \delta_1$ , as needed. And, for  $n = 2$ , the bound is  $\text{sd}(\mathcal{U}_2) \leq 1$  and it is true by the above argument whenever  $\text{sd}(\mathcal{U}_2) \geq 1$ , which trivially implies that it is true in general.  $\blacksquare$

Let us also slightly refine Conjecture 15, by explicating the intuition that we expect that the smallest VC dimension for a class with  $\text{sd} \geq n$  to be achieved by the family  $\mathcal{U}_n$ , and similarly by  $\mathcal{C}_n$  for  $\text{VC}^*$ .

**Conjecture 15** *Spherical dimension is at most exponential in both VC and  $\text{VC}^*$ . Moreover*

- *The optimal upper-bound on sd in terms of VC is witnessed by the family of universal classes  $\mathcal{U}_n$ . That is,  $\text{sd}(\mathcal{H}) \leq \max\{\text{sd}(\mathcal{U}_n) \mid \text{VC}(\mathcal{U}_n) \leq \text{VC}(\mathcal{H})\} = 2^{\text{VC}(\mathcal{H})+1} - 3$ .*
- *Similarly, the optimal upper-bound on sd in terms of  $\text{VC}^*$  is witnessed by the family of cubes  $\mathcal{C}_n$ . That is,  $\text{sd}(\mathcal{H}) \leq \max\{\text{sd}(\mathcal{C}_n) \mid \text{VC}^*(\mathcal{C}_n) \leq \text{VC}^*(\mathcal{H})\} = 2^{\text{VC}^*(\mathcal{H})+1} - 2$ .*

Here we also implicitly conjecture that the lower bound on  $\text{sd}(\mathcal{U}_n) \geq n - 2$  in Lemma 2 is sharp.

### B.3. Classification of low-VC classes via spherical dimension

**Proposition 29** *Let  $\mathcal{T}_d = \{h_i \mid i \in 0, \dots, d\}$ , for  $d \geq 1$ , be the class of thresholds over  $X = [d]$ , that is,  $h_i(j) = +$  if  $j \leq i$  and  $-$  otherwise; in particular,  $h_0$  and  $h_d$  are all  $-$  and all  $+$  concepts respectively. Then  $\text{sd}_{\text{simp}}(\mathcal{T}_d) = \text{sd}(\mathcal{T}_d) = 0$ .*

Look also at Figure 14 in Section D.2 for an illustration for the  $\Delta(\mathcal{T}_3)$  and  $\Delta^{\text{ant}}(\mathcal{T}_3)$ .

**Proof** Note that for all  $x \neq y \in [d]$  either  $\{(x, +), (y, -)\}$  or  $\{(x, -), (y, +)\}$  is not realizable by  $\mathcal{T}_d$ . Hence the only elements in  $\Delta_{\mathcal{T}_d}^{\text{ant}}$  are those distributions that are supported by the all  $+$  or the all  $-$  function. Let

$$\begin{aligned} A &= \{\mu \in \Delta_{\mathcal{T}_d} : L_\mu(h_0) = 0\}, \\ -A &= \{\mu \in \Delta_{\mathcal{T}_d} : L_\mu(h_d) = 0\}. \end{aligned}$$

Then both  $A$  and  $-A$  are nonempty, closed, and disjoint, and  $\Delta_{\mathcal{T}_d}^{\text{ant}} = A \sqcup (-A)$ ; in particular,  $A$  and  $-A$  are the two connected components of  $\Delta_{\mathcal{T}_d}^{\text{ant}}$ . As  $\Delta_{\mathcal{T}_d}$  is nonempty,  $\text{sd}_{\text{simp}}(\mathcal{T}_d) \geq 0$ . At the same time, as for every  $x \in \Delta_{\mathcal{T}_d}^{\text{ant}}$ ,  $-x$  lives in the other connected component of  $\Delta_{\mathcal{T}_d}^{\text{ant}}$  (if  $x \in A$  then  $-x \in -A$  and vice versa),  $\Delta_{\mathcal{T}_d}^{\text{ant}}$  cannot contain an equivariant image of a connected antipodal space, from which  $\text{sd}(\mathcal{T}_d) \leq 0$ , as needed.  $\blacksquare$

**Proposition 30** *Let  $\mathcal{H}$  be a concept class over a finite domain  $X$  such that  $\text{sd}(\mathcal{H}) \leq 0$  and  $\text{VC}(\mathcal{H}) \leq 1$ . Then  $\mathcal{H}$  can be embedded into thresholds up to a bit-flip. That is, there is a bit-flip function  $p : X \rightarrow \{-, +\}$  such that the class  $p \cdot \mathcal{H} = \{p \cdot h \mid h \in \mathcal{H}\}$  can be embedded into a class of thresholds.*

We note that the bit-flip changes neither VC, nor  $\text{VC}^{*a}$ , nor any of the  $\text{sd}$ 's.

**Proof** Assume  $\mathcal{H}$  is such as in the statement. By bit-flipping it, we can assume that the all  $+$  concept is in  $\mathcal{H}$ . We further assume that for every  $x \in X$  there is a concept  $h \in \mathcal{H}$  such that  $h(x) = -$ ; otherwise, we can restrict  $\mathcal{H}$  to a subdomain for which the above is satisfied, embed it into a threshold there, and then it can be easily seen that it can be easily extended to a threshold on a full domain  $X$ . We can also safely assume that for each  $x_1, x_2 \in X$  there is some  $h \in \mathcal{H}$  such that  $h(x_1) \neq h(x_2)$ . Otherwise,  $x_1$  is equivalent to  $x_2$  and we can think of them as of the same point.

Define a partial order on  $X$  by  $x \leq z$  if for all  $h \in \mathcal{H}$  we have that  $h(x) = -$  implies  $h(z) = -$ ; in other words, that the sequence  $\{(x, -), (z, +)\}$  is not realizable by  $\mathcal{H}$ . It is easy to see that this relation is indeed reflexive and transitive, and the assumption that there is always some  $h \in \mathcal{H}$  with  $h(x) \neq h(z)$  guarantees that it will also be antisymmetric. Hence, it is indeed a partial order. Moreover, it is easy to see that  $\mathcal{H}$  can be embedded in a threshold whenever this order is linear, that is, if any two elements are comparable.

Observe that if two elements  $x$  and  $z$  are incomparable with respect to this order, then the sequence  $\{(x, -), (z, -)\}$  is not realizable by  $\mathcal{H}$ . Indeed, as  $\text{VC}(\mathcal{H}) \leq 1$ , one of the four sequences  $\{(x, +), (z, +)\}$ ,  $\{(x, -), (z, +)\}$ ,  $\{(x, +), (z, -)\}$ , or  $\{(x, -), (z, -)\}$  is not realizable. However, the first is realizable by the existence of all  $+$  concept, and the second and the third imply  $x \leq z$  and  $z \leq x$  respectively. So, if  $x$  and  $z$  are incomparable, this leaves out only the last option, as needed. Let us now prove several other properties of this order.

First: There are no two incomparable elements with a common lower bound. Alternatively, if for  $z_1, z_2 \in X$  there is  $x \in X$  such that  $x \leq z_1, z_2$ , then  $z_1$  and  $z_2$  are comparable. Indeed, let us

take  $h \in \mathcal{H}$  such that  $h(x) = -$ . Then, by the definition of the order,  $h(z_1) = h(z_2) = -$ , so the sequence  $\{(z_1, -), (z_2, -)\}$  is realizable and hence, by the above,  $z_1$  and  $z_2$  are comparable.

Second: There are no two incomparable elements with a common upper bound. Note that despite the similarity, the proof of this statement relies on the  $\text{sd}(\mathcal{H}) \leq 0$  assumption, not used in the previous statement. So suppose there are  $z_1, z_2$ , and  $x$  such that  $z_1$  and  $z_2$  are incomparable and  $z_1, z_2 \leq x$ . We claim that then the simplicial 1-dimensional sphere in Figure 8 (left) is realizable by  $\mathcal{H}$  and thus witnesses  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq 1$ . Indeed, the edges  $\{(x, +), (z_1, +)\}$  and  $\{(x, +), (z_2, +)\}$  are realized by all  $+$  concept. Edges  $\{(z_1, +), (z_2, -)\}$  and  $\{(z_1, -), (z_2, +)\}$  are realized by the assumption that  $z_1$  and  $z_2$  are incomparable. And the edge  $\{(x, -), (z_1, -)\}$  is realized by a function  $h$  such that  $h(z_1) = -$ : as  $z_1 \leq x$ , this implies that  $h(x) = -$ ; the edge  $\{(x, -), (z_2, -)\}$  is handled similarly.

Third: there are no three incomparable elements. Indeed, if  $x, z$ , and  $w$  are such incomparable elements then, similarly to the argument above, the simplicial 1-dimensional sphere in Figure 8 (right) is realizable by  $\mathcal{H}$  and thus witnesses  $\text{sd}_{\text{simp}}(\mathcal{H}) \geq 1$ .

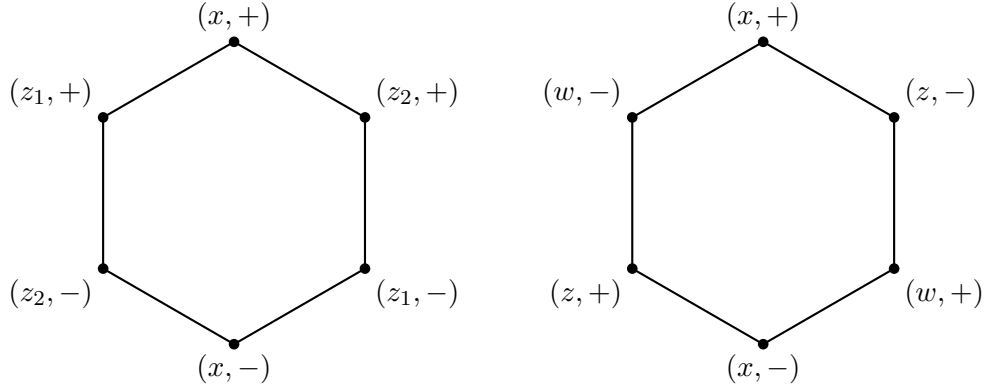


Figure 8: One-dimensional simplicial spheres in  $\Delta_{\mathcal{H}}^{\text{ant}}$ , witnessed by either two incomparable elements  $z_1$  and  $z_2$  with a common upper bound  $x$  (left), or three incomparable elements  $x, z$ , and  $w$  (right).

Now note that the third property implies that there are at most two minimal elements in  $X$ . If there is just one such minimal element, then the first property implies that all elements are comparable, in which case the order is linear, and hence  $\mathcal{H}$  can be embedded into a threshold, as needed. Then the second property implies that each  $x \in X$  is comparable with exactly one of  $x_1$  and  $z_1$ . This implies that our domain consists of two disjoint chains  $X = \{x_i\}_{i=1}^n \sqcup \{z_j\}_{j=1}^m$ , where  $x_i \leq x_{i+1}$ ,  $z_j \leq z_{j+1}$ , and  $x_i$  and  $z_j$  are incomparable for all  $i, j$ . In particular,  $x_n$  and  $z_m$  are incomparable, which implies that the sequence  $\{(x_1, -), (z_m, -)\}$  is not realizable; alternatively,  $h(x_n) = -$  implies  $h(z_m) = +$  for all  $h \in \mathcal{H}$ .

Now, using the contrapositive of the definition of our partial order, it is clear that  $h(z_m) = +$  implies  $h(z_j) = +$  for all  $j$ , hence we can define a bit-flip function  $p : X \rightarrow \{-, +\}$  by  $p(x_i) = +$  and  $p(z_j) = -$  for all  $i$  and  $j$ . Then it is easy to see that after applying this bit-flip, the elements of  $p \cdot \mathcal{H}$  will be linearly ordered as

$$x_1 \leq x_2 \cdots \leq x_n \leq z_m \leq z_{m-1} \cdots z_1.$$

In particular,  $p \cdot \mathcal{H}$  can be embedded into a threshold, as needed. ■

**Lemma 31** *Let  $\mathcal{H}$  be a nonempty class over a finite domain. Then one of the following mutually exclusive cases holds:*

- $\mathcal{H}$  is a class with one hypothesis ( $\text{sd} = -1, \text{VC} = 0$ );
- $\mathcal{H}$  has at least two hypotheses and, up to a bit-flip, is a subclass of thresholds ( $\text{sd} = 0, \text{VC} = 1$ );
- $\mathcal{H}$  is a  $\text{VC} = 1$  class that is not, up to a bit-flip, a subclass of thresholds ( $\text{sd} = 1, \text{VC} = 1$ );
- $\text{VC}(\mathcal{H}) \geq 2$  and  $\text{sd}(\mathcal{H}) \geq 1$ .

Moreover, in the above classification,  $\text{sd}$  is any from  $\text{sd}$  or  $\text{sd}_{\text{simp}}$ .

Moreover, the first bullet holds even without the finite domain assumption, that is,  $\text{sd}_{\infty}(\mathcal{H}) = -1$  if and only if  $\mathcal{H}$  has at most one hypothesis.

**Proof** We start by noting that  $\text{sd} = -1$ , for any  $\text{sd}$ , is equivalent to  $\Delta_{\mathcal{H}}^{\text{ant}}$  being empty, which, in turn, is equivalent to  $\mathcal{H}$  having at most one hypothesis: indeed, if  $\mathcal{H}$  contains  $h_1 \neq h_2$ , then there is  $x \in X$  such that  $h_1(x) \neq h_2(x)$ , in which case  $(x, -), (x, +) \in \Delta_{\mathcal{H}}^{\text{ant}}$  and witness  $\text{sd}_{\text{simp}} \geq 0$ .

So assume from now on that  $\mathcal{H}$  is finite and has at least two hypotheses. Note that if  $\text{VC}(\mathcal{H}) \geq 2$  then, by Lemma 2,  $\text{sd}_{\text{simp}} \geq 1$ , and so this situation falls under the last bullet in the classification. So let us now assume that  $\text{VC}(\mathcal{H}) = 1$ . We now need to check that the second and the third bullets exhaust the possible situations. That is, that  $\text{sd}(\mathcal{H})$  is either 0 or 1, and that  $\text{sd}(\mathcal{H}) = 0$  if and only if  $\mathcal{H}$  can be embedded into thresholds.

Note that  $\text{VC}(\mathcal{H}) = 1$ , again by Lemma 2, implies that  $\text{sd}_{\text{simp}} \geq 0$ . At the same time, it is well-known and easy to prove that  $\text{VC}(\mathcal{H}) = 1$  implies that  $\mathcal{H}$  can be embedded into an extremal class  $\mathcal{E}$  with  $\text{VC}(\mathcal{E}) = 1$ . Then, by (yet to be proven) Theorem 9,  $\text{sd}(\mathcal{H}) \leq 2\text{VC}(\mathcal{E}) - 1 = 1$ . Thus, indeed,  $\text{sd}(\mathcal{H})$  is either 0 or 1.

Finally, Propositions 29 and 30 above establish that, in one direction,  $\text{sd}(\mathcal{H}) \leq 0$  whenever  $\mathcal{H}$  is embeddable into a threshold; here we also use monotonicity of  $\text{sd}$  and the fact that bit-flip does not change  $\text{sd}$  or  $\text{sd}_{\text{simp}}$ . And, in the other direction,  $\mathcal{H}$  can be embedded, up to a bit-flip, into a threshold whenever  $\text{sd}(\mathcal{H}) \leq 0$ . ■

#### B.4. Stability and list-replicability

The proof of Theorem 10 relies on the following statement:

**Proposition 32 (Corollary 3 in Chase et al. (2024b))** *For a finite class  $\mathcal{H}$  over  $X$ ,  $\text{LR}(\mathcal{H}, \varepsilon)$  is the minimal integer  $L$  such that there exist a closed cover  $\mathcal{F} = \{A_h \mid h \in \{-, +\}^X\}$  of  $\Delta_{\mathcal{H}}$  of overlap-degree at most  $L$ , such that  $L_{\mu}(h) \leq \varepsilon$  for all  $h \in \{-, +\}^X$  and  $\mu \in A_h$ .*

**Theorem 10** *Let  $\mathcal{H}$  be a concept class such that  $\text{sd}(\mathcal{H}) \geq 1$ . Then, for any  $0 < \varepsilon < 1/2$ , it holds*

$$\text{LR}(\mathcal{H}, \varepsilon) \geq \lceil (\text{sd}(\mathcal{H}) + 3)/2 \rceil.$$

*In particular,  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1 + \lceil \text{VC}(\mathcal{H})/2 \rceil$  whenever  $\text{VC}(\mathcal{H}) \geq 2$ , and  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1 + \lfloor \text{VC}^{*a}(\mathcal{H})/2 \rfloor$  (with no additional condition).*

*The only exception for the bounds in terms of  $\text{VC}$  and  $\text{sd}$  are the classes of the form  $\mathcal{H} = h_{-}, h_{+}$  such that  $h_{-}(x) = -h_{+}(x)$  for all  $x \in X$ , for example,  $\mathcal{C}_1$ , for which  $\text{LR}(\mathcal{H}, \varepsilon) = 1$ ,  $1 + \lceil \text{VC}(\mathcal{H})/2 \rceil = 1 + \lceil 1/2 \rceil = 2$ , and  $\lceil (\text{sd}(\mathcal{H}) + 3)/2 \rceil = \lceil (0 + 3)/2 \rceil = 2$ .*



**Proof** The main statement of the theorem is a straightforward application of Local LS (Theorem 24) to Proposition 32, which goes as follows. Let  $\beta: \mathbb{S}^d \rightarrow \Delta_{\mathcal{H}'}^{\text{ant}}$ , where  $d = \text{sd}(\mathcal{H}) \geq 1$ ,  $\mathcal{H}'$  is the restriction of  $\mathcal{H}$  to a finite domain, and  $\beta$  an equivariant map, such that  $\mathcal{H}'$  and  $\beta$  witness  $\text{sd}(\mathcal{H}) = d$ . Let  $\mathcal{F}$  be the closed cover from Proposition 32 such that  $\text{LR}(\mathcal{H}', \varepsilon)$  is the overlap degree of  $\mathcal{F}$ . Note that the property  $L_\mu(h) + L_{-\mu}(h) = 1$  (Equation (1)) implies that all sets in  $\mathcal{F}$  are antipodal-free. Let  $\mathcal{G} = \beta^{-1}[\mathcal{F}]$ . Then  $\mathcal{G}$  is a closed antipodal cover of  $\mathbb{S}^n$ , whose overlap equals to the one of  $\mathcal{F}$ . Then, by Theorem 24,  $\text{LR}(\mathcal{H}, \varepsilon) \geq \text{LR}(\mathcal{H}', \varepsilon) \geq \lceil (d+3)/2 \rceil$ , as needed.

The bounds in terms of VC and  $\text{VC}^{*a}$  follow from Lemma 2, where the  $\text{VC}^{*a}$  bound is true with an additional condition  $\text{VC}^{*a}(\mathcal{H}) \geq 3$ , which we will lift in a short while. The remaining claims are easy case studies. Note that, by definition  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1$ . Let us prove the following easy statement: *For a class  $\mathcal{H}$ ,  $\text{LR}(\mathcal{H}, \varepsilon) = 1$  if and only if  $\mathcal{H}$  has just one concept or  $\mathcal{H} = \{h_-, h_+\}$  from the statement of the theorem.* The fact that  $\text{LR}(\mathcal{H}, \varepsilon) = 1$  for the two classes described above is trivial. So assume  $\mathcal{H}$  does not fall under these two cases. In all other situations, there are concepts  $h_1, h_2 \in \mathcal{H}$  and  $x_1, x_2 \in X$  such that  $h_1(x_1) \neq h_2(x_1)$  and  $h_1(x_1) = h_2(x_1)$ . Assuming, for simplicity, that  $h_1(x_1) = h_1(x_2) = h_2(x_2) = +$  and  $h_2(x_1) = -$ , we observe that for  $\mathcal{H}' = \mathcal{H}|_{\{x_1, x_2\}}$ ,  $\Delta_{\mathcal{H}'}$  contains simplices  $\{(x_1, -), (x_2, +)\}$  and  $\{(x_1, +), (x_2, +)\}$ , which means that  $\Delta_{\mathcal{H}'}$  contains a path between antipodal vertices  $(x_1, -)$  and  $(x_1, +)$ . If  $\mathcal{F}$  is the covering of  $\Delta_{\mathcal{H}'}$  from Proposition 32, then no  $F$  in  $\mathcal{F}$  can cover both  $(x_1, -)$  and  $(x_1, +)$ , and so, essentially by the definition of connectedness, the overlap of  $\mathcal{F}$  is at least 2. Hence,  $\text{LR}(\mathcal{H}, \varepsilon) \geq \text{LR}(\mathcal{H}', \varepsilon) \geq 2$ , as needed.

For the VC bound  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1 + \lceil \text{VC}(\mathcal{H})/2 \rceil$ , it is true for  $\text{VC} \geq 2$ . It is also easy to see that, in the remaining cases, it can only be invalidated if  $\text{VC}(\mathcal{H}) = 1$  and  $\text{LR}(\mathcal{H}, \varepsilon) = 1$ . By the above statement, it can only be the case when  $\mathcal{H} = \{h_-, h_+\}$ , and indeed, this case is an exception.

The  $\text{VC}^{*a}$  bound  $\text{LR}(\mathcal{H}, \varepsilon) \geq 1 + \lceil \text{VC}(\mathcal{H})/2 \rceil$  is true for  $\text{VC}^{*a} \geq 3$ , and again, can only be invalidated if  $\text{VC}^{*a}(\mathcal{H}) = 2$  and  $\text{LR}(\mathcal{H}, \varepsilon) = 1$ . However, this is impossible, as in both two cases  $\mathcal{H} = \{h\}$  or  $\mathcal{H} = \{h_-, h_+\}$  for which  $\text{LR}(\mathcal{H}, \varepsilon) = 1$ , the corresponding class does not dually antipodally shatter two concepts.

Finally, the main bound  $\text{LR}(\mathcal{H}, \varepsilon) \geq \lceil (\text{sd}(\mathcal{H}) + 3)/2 \rceil$  is true for  $\text{sd} > 0$  and can only be invalidated if  $\text{sd}(\mathcal{H}) = 0$  and  $\text{LR}(\mathcal{H}, \varepsilon) = 1$ . Again, just as in the VC bound, it can only happen when  $\mathcal{H} = \{h_-, h_+\}$ , and this case is an exception.  $\blacksquare$

## B.5. SCO reductions

Prior to proving Theorem 13, let us define and discuss the *sign-rank*, introduced in its formulation in Section 6. In the definition below we follow Definition 11 in Chornomaz et al. (2024), where the sign-rank was studied as a special kind of reduction to a class defined by homogenous hyperplanes in  $\mathbb{R}^d$ ; this, in turn, follows Definition 1 in Kamath et al. (2020).

**Definition 33 (Sign-rank)** *For a class  $\mathcal{H} \subseteq \{\pm 1\}^X$ , the sign-rank of  $\mathcal{H}$  is the smallest  $d$  for which there are maps  $\varphi: X \rightarrow \mathbb{R}^d$  and  $w: \mathcal{H} \rightarrow \mathbb{R}^d$  such that for all  $h \in \mathcal{H}$  and  $x \in X$ , it holds  $\text{sign}\langle w(h), \varphi(x) \rangle = h(x)$ .*

We also note that the canonical definition of the sign-rank comes from Paturi and Simon (1986), where it is defined as a minimal rank of a real matrix, whose sign pattern coincides with the class. The fact that it coincides with the above definition is well-known. Note also that the definition in

terms of the sign matrix can be easily interpreted as an embedding into extremal class (see, for example, Section 3.1 in [Chase et al. \(2024a\)](#)), yielding

$$\text{VC}^{\text{extr}}(\mathcal{H}) \leq \text{sign-rank}(\mathcal{H}).$$

The proof of Theorem 12 closely follows the proof of Theorem 1 in [Chornomaz et al. \(2024\)](#), so we will only do a brief walkthrough. It relies on the following version of the Borsuk-Ulam theorem:

**Theorem 34 (Theorem 9 in [Chornomaz et al. \(2024\)](#))** *Let  $W$  be a convex set in  $\mathbb{R}^d$  and let  $G \subseteq \mathbb{S}^n \times W$ . Additionally, suppose that:*

- *For any  $u \in \mathbb{S}^n$ , the set  $G_u = \{w \in W \mid (u, w) \in G\}$  is nonempty;*
- *For any  $u \in \mathbb{S}^n$ , the convex hulls of  $G_u$  and  $G_{-u}$  are disjoint;*
- *For any  $w \in W$ , the set  $G^w = \{u \in \mathbb{S}^n \mid (u, w) \in G\}$  is open.*

*Then  $d \geq n + 1$ .*

Also, as a technical statement, we use the following:

**Lemma 35 (Lemma 21 in [Chornomaz et al. \(2024\)](#))** *Let  $I \subseteq \mathbb{R}^d$  be a nonempty collection of pointwise positive vectors in  $\mathbb{R}^d$ . Let  $X = \{x \in [0, 1]^d : \sum_{i=1}^d x_i = 1\}$  and define  $F : X \rightarrow \mathbb{R}$  by*

$$F(x) = \inf_{\alpha \in I} \langle x, \alpha \rangle.$$

*Then  $F$  is continuous.*

**Theorem 12** *For any class  $\mathcal{H}$  we have*

$$\text{sd}(\mathcal{H}) + 1 \leq \text{SCO}(\mathcal{H}) \leq \text{sign-rank}(\mathcal{H}).$$

**Proof** The fact that  $\text{SCO}(\mathcal{H}) \leq \text{sign-rank}(\mathcal{H})$  was shown in [Chornomaz et al. \(2024\)](#), so we now concentrate on proving the first inequality.

Let  $\text{SCO}(\mathcal{H}) = d$ . Recall that it means that there are abstract set  $Z$ , convex set  $W \subseteq \mathbb{R}^d$ , and a collection of convex loss functions  $\{\ell_z\}_{z \in Z}$  from  $W$  to  $\mathbb{R}^+$ , and a pair of functions  $r_{\text{in}} : X \times \{-, +\} \rightarrow Z$  and  $r_{\text{out}} : W \rightarrow \{-, +\}^X$ , defining an  $(\alpha, \beta)$  reduction, for  $\alpha > 0$  and  $0 < \beta < 1/2$ . That is,  $r_{\text{in}}$  and  $r_{\text{out}}$  are such that and for every realizable distribution  $\mu \in \Delta_{\mathcal{H}}$  we have

$$L_{r_{\text{in}}(\mu)}(w) \leq \inf_{w' \in W} L_{r_{\text{in}}(\mu)}(w') + \alpha \implies L_{\mu}(r_{\text{out}}(w)) \leq \beta.$$

Let also  $\text{sd}(\mathcal{H}) = n$  and  $f : \mathbb{S}^n \rightarrow \Delta_{\mathcal{H}}^{\text{ant}}$  be a finitely supported  $n$ -sphere witnessing it. Note that the fact that the sphere is finitely supported enables us to assume, without losing generality, that the sets  $X$ , and, consequently,  $Z$ , are finite. Let us now define  $G \subseteq \mathbb{S}^n \times W$  as:

$$\begin{aligned} G &= \{(u, w) \mid w \text{ is } \alpha/2\text{-optimal for } r_{\text{in}}(f(u))\} \\ &= \left\{ (u, w) \mid L_{r_{\text{in}}(f(u))}(w) < \inf_{w' \in W} L_{r_{\text{in}}(f(u))}(w') + \alpha/2 \right\}. \end{aligned}$$

Note that if we prove that  $G$  satisfies the conditions in Theorem 34, the application of this theorem yields the desired bound  $n + 1 \leq d$ . Also, out of the three conditions on  $G$ , the first one is trivially satisfied by construction, and the second one follows by observing that each  $G_u$  is convex and if  $w \in G_u \cap G_{-u}$  then the hypothesis  $r_{\text{out}}(w)$  is  $\beta$ -optimal for both distributions  $f(v)$  and  $-f(v)$ , contradicting Equation (1).

Finally, for  $w \in W$ ,  $G^w$  is defined as  $G^w = \{u \in \mathbb{S}^n \mid F_w(f(u)) < \alpha/2\}$ , where

$$F_w(\mu) = L_{r_{\text{in}}(\mu)}(w) - \inf_{w' \in W} L_{r_{\text{in}}(\mu)}(w').$$

Thus,  $G^w$  is open whenever  $F_w$  is continuous, for which, in turn, it is enough to check that the map  $\text{OPT}(\eta) = \inf_{w' \in W} L_\eta(w')$  is continuous; here  $\eta = r_{\text{in}}(\mu)$  is a probability distribution over  $Z$ . The latter follows by applying Lemma 35 to  $\text{OPT}$  rewritten as

$$\text{OPT}(\eta) = \inf_{w' \in W} \mathbb{E}_{z \sim \eta} l_z(w') = \inf_{w' \in W} \sum_{z \in Z} \mathbb{P}_\eta(z) \cdot l_z(w').$$

■

**Theorem 13** *For any two concept classes  $\mathcal{H}_1, \mathcal{H}_2$  we have*

$$\text{sign-rank}(\mathcal{H}_1 \times \mathcal{H}_2) \leq \text{sign-rank}(\mathcal{H}_1) + \text{sign-rank}(\mathcal{H}_2).$$

**Proof** For  $i \in \{1, 2\}$ , let  $n_i$  be the sign-ranks of  $\mathcal{H}_i$ , and let  $\varphi_i : X_i \rightarrow \mathbb{R}^{n_i}$  and  $w_i : \mathcal{H}_i \rightarrow \mathbb{R}^{n_i}$  be the respective representations maps, witnessing it (as per Definition 33). Thus, for all  $x_i \in X_i$  and  $h_i \in \mathcal{H}_i$  we have

$$h_i(x_i) = \text{sign}\langle w_i(h_i), \varphi_i(x_i) \rangle.$$

Now define the representation maps  $\varphi : X_1 \sqcup X_2 \rightarrow \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$ ,  $w : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$  by

$$\varphi(x) = \begin{cases} (\varphi_1(x), 0) & x \in X_1, \\ (0, \varphi_2(x)) & x \in X_2, \end{cases}$$

and

$$w(h_1 \times h_2) = (w(h_1), w(h_2)).$$

A simple calculation shows that we indeed have

$$h(x) = \text{sign}\langle w(h), \varphi(x) \rangle.$$

for all  $h \in \mathcal{H}_1 \times \mathcal{H}_2$ ,  $x \in X_1 \sqcup X_2$ , which implies the desired result. ■

### B.6. Products of classes

For classes  $\mathcal{H}_1$  and  $\mathcal{H}_2$  on disjoint domains  $X_1$  and  $X_2$ , the *product*  $\mathcal{H}_1 \times \mathcal{H}_2$  is defined as:

$$\mathcal{H}_1 \times \mathcal{H}_2 = \{h_1 \times h_2 \mid h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\},$$

where  $h = h_1 \times h_2$  is a function on  $X_1 \sqcup X_2$  such that  $h(x) = h_1(x)$  for  $x \in X_1$  and  $h_2(x)$  for  $x \in X_2$ . In simplicial geometry, joins naturally correspond to direct products, see Section A.6. In particular, this extends to our setup:

**Lemma 36** *For classes  $\mathcal{H}_1$  and  $\mathcal{H}_2$ ,  $\Delta_{\mathcal{H}_1 \times \mathcal{H}_2} = \Delta_{\mathcal{H}_1} * \Delta_{\mathcal{H}_2}$ , where the join is a geometric join of antipodal spaces. Moreover, if the domains of  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are finite, in which case  $\Delta_{\mathcal{H}_1}$  and  $\Delta_{\mathcal{H}_2}$  have an additional structure of simplicial complexes, then the above equality is also in the sense of the join of antipodal simplicial complexes.*

**Proof** The correspondence is established by representing  $\mu$  on  $X_1 \sqcup X_2$ , realizable by  $\mathcal{H}_1 \times \mathcal{H}_2$ , in the barycentric coordinates of the join as

$$\mu = \mu(X_1) \cdot \frac{\mu|_{X_1}}{\mu(X_1)} + \mu(X_2) \cdot \frac{\mu|_{X_2}}{\mu(X_2)}.$$

Note that, unless  $\mu(X_1) = 0$ , the measure  $\mu|_{X_1}/\mu(X_1)$  is a distribution, realizable by  $X_1$ . And the case  $\mu(X_1) = 0$ , and, similarly,  $\mu(X_2) = 0$ , aligns with the definition of the join, as in this case the corresponding coefficient  $\mu(X_1)$  is 0, and hence the term  $\mu|_{X_1}/\mu(X_1)$  does not matter. The check for the simplicial case is trivial.  $\blacksquare$

**Lemma 3** *For classes  $\mathcal{H}_1$  and  $\mathcal{H}_2$ ,  $\text{VC}(\mathcal{H}_1 \times \mathcal{H}_2) = \text{VC}(\mathcal{H}_1) + \text{VC}(\mathcal{H}_2)$  and  $\text{sd}(\mathcal{H}_1 \times \mathcal{H}_2) \geq \text{sd}(\mathcal{H}_1) + \text{sd}(\mathcal{H}_2) + 1$ , where  $\text{sd}$  can be any from  $\text{sd}_\infty$ ,  $\text{sd}$ , or  $\text{sd}_{\text{simp}}$ .*

*In particular, for an  $m$ 'th direct power  $\mathcal{U}_n^m$  of the universal class  $\mathcal{U}_n$ , we have:*

- $\text{VC}(\mathcal{U}_n^m) = m \lfloor \log n \rfloor$ ;
- $n \leq \text{VC}^*(\mathcal{U}_n^m) \leq n + \lfloor \log m \rfloor$ . Moreover, for  $n \geq 3$ , the lower bound can be improved to  $n - 2 + \lfloor \log m \rfloor \leq \text{VC}^*(\mathcal{U}_n^m)$ ;
- $m(n - 1) - 1 \leq \text{sd}_{\text{simp}}(\mathcal{U}_n^m) \leq \text{sd}(\mathcal{U}_n^m) \leq mn$ .

*In particular, for the class  $\mathcal{U}_n^n$ ,  $\text{VC}(\mathcal{U}_n^n) \sim \text{VC}^*(\mathcal{U}_n^n) \sim n$  (up to a log factor) and  $\text{sd}_{\text{simp}}(\mathcal{U}_n^n) \gtrsim n^2$ . More generally, for  $\alpha, \beta \in \mathbb{R}^+$ , we can define a family of classes  $\mathcal{U}_{\beta n}^{\alpha n}$  (with some rounding), that would witness the asymptotical behavior  $\text{VC} \sim n^\alpha$ ,  $\text{VC}^* \sim n^\beta$ , and  $\text{sd}_{\text{simp}} \sim n^{\alpha+\beta}$ .*

**Proof** The equality  $\text{VC}(\mathcal{H}_1 \times \mathcal{H}_2) = \text{VC}(\mathcal{H}_1) + \text{VC}(\mathcal{H}_2)$  is trivial, and the inequality  $\text{sd}_\infty(\mathcal{H}_1 \times \mathcal{H}_2) \geq \text{sd}_\infty(\mathcal{H}_1) + \text{sd}_\infty(\mathcal{H}_2) + 1$  follows from the statement  $\Delta_{\mathcal{H}_1 \times \mathcal{H}_2} = \Delta_{\mathcal{H}_1} * \Delta_{\mathcal{H}_2}$  in Lemma 36, together with the discussion in Section A.6, in particular, from  $\mathbb{S}^n * \mathbb{S}^m \cong \mathbb{S}^{n+m+1}$  in Proposition 26. Indeed, if  $\beta_1: \mathbb{S}^{d_1} \rightarrow \Delta_{\mathcal{H}_1}$  and  $\beta_2: \mathbb{S}^{d_2} \rightarrow \Delta_{\mathcal{H}_2}$  witness  $\text{sd}_\infty(\mathcal{H}_1) \geq d_1$  and  $\text{sd}_\infty(\mathcal{H}_2) \geq d_2$  respectively, then  $\beta_1 * \beta_2: \mathbb{S}^{d_1+d_2+1} \rightarrow \Delta_{\mathcal{H}_1 \times \mathcal{H}_2}$  witness  $\text{sd}_\infty(\mathcal{H}_1 \times \mathcal{H}_2) \geq d_1 + d_2 + 1$ .

The fact that it applies to all three  $\text{sd}$ 's also requires only some straightforward checks. The  $\text{sd}$  case is by taking some finitely supported witnesses  $\mathcal{H}'_1$  and  $\mathcal{H}'_2$ , and the  $\text{sd}_{\text{simp}}$  is by the discussion in Section A.6 that the geometric join is consistent with the simplicial one.

The equality  $\text{VC}(\mathcal{U}_n^m) = m \lfloor \log n \rfloor$  and the inequality  $\text{sd}_{\text{simp}}(\mathcal{U}_n^m) \geq m(n-2) + (m-1) = m(n-1) - 1$  follow from the above and from the bounds  $\text{VC}(\mathcal{U}_n) = \lfloor \log n \rfloor$  and  $\text{sd}_{\text{simp}}(\mathcal{U}_n) \geq n-2$  in Lemma 2. The upper bound  $\text{sd}(\mathcal{U}_n^m) \leq mn$  follows from Theorem 13 as

$$\text{sd}(\mathcal{U}_n^m) \leq \text{sign-rank}(\mathcal{U}_n^m) \leq m \cdot \text{sign-rank}(\mathcal{U}_n) = mn.$$

The upper bound  $\text{VC}^*(\mathcal{U}_n^m) \leq n + \lfloor \log m \rfloor$  follows by observing that in order to dually shatter  $d$  functions, the class should have at least  $2^d$  elements in the domain, so  $2^{\text{VC}^*(\mathcal{U}_n^m)} \leq m2^n$ . The  $n \leq \text{VC}^*(\mathcal{U}_n^m)$  lower bound holds because  $\mathcal{U}_n$  is a restriction of  $\mathcal{U}_n^m$ .

The  $n-2 + \lfloor \log m \rfloor \leq \text{VC}^*(\mathcal{U}_n^m)$  lower bound follows from the following construction. Recall that the natural parametrization of  $\mathcal{U}_n$  is to consider its domain to be all subsets of  $[n]$ , and its functions  $h_i$  to be characteristic functions of an element,  $h_i(S) = [i \in S]$ , for  $i \in [n]$ . Let  $\mathcal{U}'_n$  be the same class, but additionally containing all  $-$  and all  $+$  concepts. Trivially,  $\mathcal{U}'_{n-2}$  is a subclass of  $\mathcal{U}_n$ , so assume that we are dealing with  $\mathcal{U}'_{n-2}$  instead of  $\mathcal{U}_n$ . Let  $d$  be maximal such that  $2^d \leq m$ ; assume, without losing generality, that  $m = 2^d$ , and associate  $m$  with the set of all subsets of  $[d]$ , making the domain of  $(\mathcal{U}'_{n-2})^m$  to be  $2^{[n-2]} \times 2^{[d]}$ . Now, for  $i \in [n-2+d]$ , let us define the function  $g_i$  as:

- For  $i \in [n-2]$ ,  $g_i = h_i^m$ , that is,  $g_i(S, P) = [i \in S]$ , for all  $S \subseteq [n-2]$  and  $P \subseteq [d]$ ;
- For  $j \in [d]$ ,  $g_{n-2+j}(S, P) = [j \in P]$ . That is,  $g_{n-2+j}(\cdot, P) \equiv -$  whenever  $j \notin P$  and  $\equiv +$  whenever  $j \in P$ . As both  $\equiv -$  and  $\equiv +$  are in  $\mathcal{U}'_n$ ,  $g_{n-2+j} \in (\mathcal{U}'_n)^m$ .

It is then easy to see that  $(\mathcal{U}'_n)^m$  dually shatter  $g_i$ , for  $i \in [n-2+d]$ , as needed.  $\blacksquare$

## Appendix C. Proofs for disambiguations

Although we typically denote a generic class by  $\mathcal{H}$ , in the context of disambiguations our usual notation for it would be  $\mathcal{D}$ . In particular, we use  $\mathcal{H}$  in Section C.1, which does not deal specifically with disambiguation and mostly uses  $\mathcal{D}$  through the rest of the section.

### C.1. Antipodal domains

We say that a set  $X$  is an *antipodal domain* if it is equipped with a fixed-point free involution  $\nu$ , which we call *antipodal map*; if the  $\nu$  is clear from the context, we will denote  $\nu(x)$  by  $-x$ . This is an obvious relaxation of antipodal space introduced in Section 2, in particular, we consider any antipodal space as an antipodal domain with the same antipodality. However, in this definition, we do not assume any topology on  $X$  and, naturally, do not require  $\nu$  to be continuous.

For an antipodal domain  $X$ , we say that a total hypothesis  $h$  on  $X$  is *antipodal* if  $h(-x) = -h(x)$ , for all  $x \in X$ ; a class is antipodal if all hypotheses in it are antipodal. We say that  $r: X \rightarrow X$  is a *representation map* if for every  $x \in X$ , either  $r(x) = r(-x) = x$ , or  $r(x) = r(-x) = -x$ . For a total hypothesis  $h$  on  $X$  and a representation map  $r$  we define the *symmetrization* of  $h$  by  $r$ , denoted  $h^r: X \rightarrow \{-, +\}$ , as

$$h^r(x) = \begin{cases} h(x) & \text{if } h(x) = -h(-x); \\ h(x) & \text{if } h(x) \neq -h(-x) \text{ and } x = r(x); \\ -h(x) & \text{if } h(x) \neq -h(-x) \text{ and } x = -r(x). \end{cases}$$

The point of symmetrization is to make a hypothesis antipodal, so it would make sense to call it antipodalization; however, this would sound ugly. Furthermore, for a total class  $\mathcal{H}$  and a representation map  $r$  over  $X$ , we define the *symmetrization* of  $\mathcal{H}$  by  $r$ , denoted  $\mathcal{H}^r$ , as  $\mathcal{H}^r = \{h^r \mid h \in \mathcal{H}\}$ ; crucially, all concepts of  $\mathcal{H}$  are symmetrized via the same representation map. Finally, we say that  $\mathcal{H}'$  is a symmetrization of  $\mathcal{H}$  if  $\mathcal{H}' = \mathcal{H}^r$  for some symmetrization map  $r$ .

It is easy to see that a symmetrization of  $\mathcal{H}$  is antipodal and that antipodal classes are invariant under symmetrization. What is also important for us is that  $\text{VC}$  and  $\text{VC}^{*a}$  behave well under symmetrization.

**Proposition 37 (Symmetrization does not increase VC and  $\text{VC}^{*a}$ )**

1. An antipodal class  $\mathcal{H}$  dually shatters  $H \subseteq \mathcal{H}$  if and only if it dually antipodally shatters  $H$ . In particular,  $\text{VC}^*(\mathcal{H}) = \text{VC}^{*a}(\mathcal{H})$ ;
2. Let  $\mathcal{H}'$  be a symmetrization of  $\mathcal{H}$ . Then  $\text{VC}(\mathcal{H}') \leq \text{VC}(\mathcal{H})$  and  $\text{VC}^{*a}(\mathcal{H}') \leq \text{VC}^{*a}(\mathcal{H})$ .

*Note.* It is not, in general, true that  $\text{VC}^*(\mathcal{H}') \leq \text{VC}^*(\mathcal{H})$ . To an extent, we introduced dual antipodal shattering as a way of circumventing this problem.

**Proof** The first statement is trivial: Indeed, for any  $x \in X$  that witnesses pattern  $t$  on  $H \subseteq \mathcal{H}$ , by the definition of antipodal class,  $-x$  witnesses pattern  $-t$ .

The rest of the proof is for the second statement. So let  $\mathcal{H}'$  be a symmetrization of  $\mathcal{H}$  with a representation map  $r$ . We will use the following statement, which easily follows from the definition of symmetrization: *If  $r(x) = x$ , then  $h^r(x) = h(x)$ , for any  $h \in \{-, +\}^X$  and  $x \in X$ .*

Let  $S \subseteq X$  be a set witnessing  $\text{VC}(\mathcal{H}') \geq d$ , that is,  $S = \{s_1, \dots, s_d\}$  is a  $d$ -set shattered by  $\mathcal{H}'$ . Let  $S' = \{s'_1, \dots, s'_d\}$  be such that  $s'_i = s_i$  or  $-s_i$ , for all  $i \in [d]$ . We claim that  $\mathcal{H}'$  shatters any such  $S'$ . Indeed, let  $t \in \{-, +\}^d$  be an arbitrary pattern on  $S$ , and let  $t' \in \{-, +\}^d$  be defined as  $t'(i) = t(i)$  if  $s'_i = s_i$  and  $-t(i)$  otherwise. Now, if  $h'_t \in \mathcal{H}'$  is a concept witnessing  $t$  on  $S$ , that is, such that  $h'_t(s_i) = t(i)$ , for all  $i \in [d]$ , then, by construction,  $h'_{t'}(s'_i) = [s_i = s'_i] \cdot h'_t(s_i) = [s_i = s'_i]t(i) = [s_i = s'_i]^2 t(i) = t(i)$ , where  $[s_i = s'_i]$  is  $+$  if  $s_i = s'_i$  and  $-$  otherwise. Thus,  $h'_{t'}$  witnesses the pattern  $t$  on  $S'$ . Trivially, the map  $t \mapsto t'$  is a bijection on  $\{-, +\}^d$ , and so  $\mathcal{H}'$  witnesses all the patterns on  $S'$ , that is, shatters  $S'$ .

By the above, by taking an appropriate  $S'$  instead of  $S$ , we can, without losing generality, assume that for every  $i \in [d]$ ,  $r(s_i) = s_i$ . We claim that then  $\mathcal{H}$  shatters  $S$ . Indeed, if  $h'_t \in \mathcal{H}'$  is the concept witnessing  $t$  on  $S$ , and let  $h_t \in \mathcal{H}$  be such that  $h'_t = h_t^r$ . As  $r(s_i) = s_i$ , by the outlined statement above, it follows that  $h'_t(s_i) = h_t(s_i) = t(i)$ , for all  $i \in [d]$ . Thus,  $h_t \in \mathcal{H}$  witnesses  $t$  on  $S$ , and so  $\mathcal{H}$  shatters  $S$ , implying  $\text{VC}(\mathcal{H}) \geq d$ . As this is true for any  $d \leq \text{VC}(\mathcal{H}')$ , we have  $\text{VC}(\mathcal{H}) \geq \text{VC}(\mathcal{H}')$ , as needed.

Regarding the dual dimensions, note first that  $\text{VC}^{*a}(\mathcal{H}') = \text{VC}^*(\mathcal{H}')$  by Proposition 20. Now, suppose  $\mathcal{H}'$  dually shatters some  $H' \in \mathcal{H}'$ . Note that  $\mathcal{H}$  as a set can be identified with  $\mathcal{H}$ , and we will denote by  $H$  the set  $H'$ , treated as a subset of  $\mathcal{H}$ ; in particular,  $H' = \{h^r \mid h \in H\}$ . Let  $S \subseteq X$  be the set witnessing all dual patterns for  $\mathcal{H}'$  on  $H'$  and let  $S' = \{r(s) \mid s \in S\}$ . Then for every dual pattern  $t$  on  $H'$ , there is  $s' \in S'$  witnessing for  $\mathcal{H}'$  either  $t$  or  $-t$  on it. However, every  $s \in S'$  satisfies  $s = r(s)$ , and so, by the outlined statement,  $h \equiv h' = h^r$  on  $S'$ , for every  $h \in H$ . Thus,  $S'$  witnesses the dual antipodal shattering of  $H$  by  $\mathcal{H}_n$ . This immediately entails  $\text{VC}^{*a}(\mathcal{H}) \geq \text{VC}^{*a}(\mathcal{H}')$ , as needed.  $\blacksquare$

For a set  $X$ , we define the *antipodal extension*  $X^a$  of  $X$  as a set  $X^a = X \times \{-, +\}$  together with a fixed-point free involution  $(x, y) \mapsto (x, -y)$ . We also often consider  $X^a = X^- \sqcup X^+$ , where

$X^y = X \times \{y\}$ , and in this construction we typically identify  $X^+$  with  $X$ . We also usually assume that  $X^a$  is equipped with a standard representation map  $r(x, y) = (x, +)$ . For a hypothesis  $h$  on  $X$ , the *antipodal extension*  $h^a$  of  $h$  is a function  $h^a: X^a \rightarrow Y$  defined as  $h^a(x, +) = h(x)$  and  $h^a(x, -) = -h(x)$ ; in other words,  $h^a(x, y) = y \cdot h(x)$ . Finally, an antipodal extension of a class  $\mathcal{H}$  over  $X$  is an antipodal class  $\mathcal{H}^a = \{h^a \mid h \in \mathcal{H}\}$ .

In the other direction, for an antipodal class  $\mathcal{H}$  over antipodal domain  $X$  with a representation map  $r$ , we define an *restriction to representatives*  $\mathcal{H}^r$  of  $\mathcal{H}$  as a restriction (in a usual sense) of  $\mathcal{H}$  to  $\text{Im}(r)$ .

While symmetrization enforces antipodality on a class on an already antipodal domain, extensions are used to introduce antipodality to both the domain and the class. Unlike symmetrizations, they do not lose information, and pairing them with restrictions enables us to consider usual classes as antipodal ones and vice versa. All in all, these two operations are rather straightforward and used for technical convenience.

**Proposition 38 (Antipodal extension preserves VC and VC<sup>\*a</sup>)**

1. For a class  $\mathcal{H}$ ,  $(\mathcal{H}^a)^r = \mathcal{H}$ ;
2. For an antipodal class  $\mathcal{H}$ ,  $(\mathcal{H}^r)^a = \mathcal{H}$ ;
3. For a class  $\mathcal{H}$ ,  $\text{VC}(\mathcal{H}) = \text{VC}(\mathcal{H}^a)$  and  $\text{VC}^{*a}(\mathcal{H}) = \text{VC}^{*a}(\mathcal{H}^a)$ .

Note that restrictions with different representation maps can produce different classes. However, the equivalence of classes that arise this way can be easily characterized and, all in all does not produce anything deep.

**Proof** (1) and (2) are trivial. In (3)  $\text{VC}(\mathcal{H}) \leq \text{VC}(\mathcal{H}^a)$  and  $\text{VC}^{*a}(\mathcal{H}) \leq \text{VC}^{*a}(\mathcal{H}^a)$  are also trivial, and we only need to prove the other direction. Let  $S^a \subseteq X^a$  be shattered by  $\mathcal{H}^a$ . Then, obviously,  $S^a$  does not contain a pair of antipodal points. Let  $S^+ = S^a \cap X^+$  and  $S^- = S^a \cap X^-$ . It is now easy to see that  $S^a$  is shattered by  $\mathcal{H}^a$  if and only if  $S^+ \sqcup -S^- \subseteq X$  is shattered by  $\mathcal{H}$ . That proves that  $\text{VC}(\mathcal{H}) = \text{VC}(\mathcal{H}^a)$ .

Similarly, if  $H \subseteq \mathcal{H}^a$  is dually antipodally shattered and it is witnessed by  $S^a \subseteq X^a$ , then  $S = S^a \cup -S^a$  witnesses dual shattering of  $H$  by  $\mathcal{H}^a$  and  $S \cap X$  witnesses dual antipodal shattering of  $H$  by  $\mathcal{H}$ . Hence,  $\text{VC}^{*a}(\mathcal{H}) = \text{VC}^{*a}(\mathcal{H}^a)$ . ■

## C.2. Disambiguations and covers

For a partial function  $h$  on  $\mathbb{S}^n$ , let  $h^+ = h^{-1}(+) \subseteq \mathbb{S}^n$ , and similarly for  $h^-$  and  $h^*$ . We define the *interior*  $U_h$  of  $h$  as the (topological) interior of  $h^+ \cap (-h^-)$ . It is easy to see that, by construction,  $U_h$  is open and antipodal-free. Note that  $U_h$  can also be explicitly defined as

$$U_h = \{x \in \mathbb{S}^n \mid \text{there is } \varepsilon > 0 \text{ s.t. } h \equiv + \text{ on } U_\varepsilon(x) \\ \text{and } h \equiv - \text{ on } U_\varepsilon(-x)\},$$

where, as usual,  $U_\varepsilon(x)$  stands for an open  $\varepsilon$ -ball around  $x$ . Finally, for a class  $\mathcal{D}$  over  $\mathbb{S}^n$ , we define  $\mathcal{U}(\mathcal{D})$  as  $\mathcal{U}(\mathcal{D}) = \{U_h \mid h \in \mathcal{D}\}$ . Although formally this definition makes sense for partial classes, we will only use it for total.



**Proposition 39** *For a total class  $\mathcal{D}$  on  $\mathbb{S}^n$*

1.  *$\mathcal{D}$  disambiguates  $\mathbb{S}^n$  if and only if  $\mathcal{U}(\mathcal{D})$  is an (open, antipodal-free) cover;*
2. *If  $\mathcal{D}$  disambiguates  $\mathbb{S}^n$  then there is a finite subclass  $\mathcal{D}'$  of  $\mathcal{D}$  that disambiguates  $\mathbb{S}^n$ ;*
3. *If  $\mathcal{D}$  disambiguates  $\mathbb{S}^n$  then  $|\mathcal{D}| \geq n + 2$ .*
4. *If  $\mathcal{D}$  disambiguates  $\mathbb{S}^n$  then any symmetrization of  $\mathcal{D}$  disambiguates  $\mathbb{S}^n$ .*

In view of (4) and also because, by Lemma 37, symmetrization does not increase VC and  $\text{VC}^{*a}$ , we can concentrate on antipodal classes as long as we are interested in those two parameters.

**Proof** Note that for any total class  $\mathcal{D}$  on  $\mathbb{S}^n$ , by construction,  $\mathcal{U}(\mathcal{D})$  is a family of open antipodal-free sets. If, additionally,  $\mathcal{D}$  disambiguates  $\mathbb{S}^n$ , then it disambiguates  $\mathcal{L}_{n,\varepsilon}$  for some  $\varepsilon > 0$ . For  $u \in \mathbb{S}^n$ , let  $\bar{u}$  be the concept of  $\mathcal{D}$  extending  $u_\varepsilon$ . Then, trivially,  $u \in U_\varepsilon(u) \subseteq U(\bar{u})$ , and hence  $\mathcal{U}(\mathcal{D})$  is a cover.

In the other direction, if  $\mathcal{U}(\mathcal{D})$  is a cover, let us, by compactness of  $\mathbb{S}^n$ , pick a finite subclass  $\mathcal{D}'$  of  $\mathcal{D}$  such that  $\mathcal{U}(\mathcal{D}')$  is a cover. Again by compactness, there is  $\varepsilon > 0$  such that for every  $u \in \mathbb{S}^n$  there is  $U \in \mathcal{U}(\mathcal{D}')$ ,  $U = U_h$  for  $h \in \mathcal{D}'$ , such that  $U_\varepsilon(u) \subseteq U_h$ . Thus,  $\mathcal{D}'$  disambiguates  $\mathcal{L}_{n,\varepsilon'}$  for any  $0 < \varepsilon' < \varepsilon$ .

The second statement directly follows from the first one, and the third is by a direct application of the Lusternik–Schnirelmann theorem (Theorem 22). Finally, the fourth statement follows from an easy observation that for any symmetrization  $h'$  of a concept  $h$ ,  $U(h') \supseteq U(h)$ .  $\blacksquare$

### C.3. Disambiguations of simplicial spheres

Let  $(\mathbb{Q}^n, v \mapsto -v)$  be a simplicial  $n$ -sphere (introduced in Section 2). We say that a class  $\mathcal{D}$  on the antipodal domain  $V(\mathbb{Q}^n)$  *disambiguates*  $\mathbb{Q}^n$  if for every  $s \in S(\mathbb{Q}^n)$  there is  $h \in \mathcal{D}$  such that  $h(v) = +$  for all  $v \in s$  and  $h(v) = -$  for all  $v \in -s$ . Note that if  $\mathcal{D}$  is antipodal, then it suffices to require that the respective  $h \equiv +$  only on  $s$ .

Similarly to how we defined simplicial diameter of  $\mathbb{Q}$  in Section A, for a simplicial  $n$ -sphere  $\mathbb{Q}^n$  and an equivariant homeomorphism  $\iota: \|\mathbb{Q}^n\| \rightarrow \mathbb{S}^n$  we define the (*simplicial*) *diameter* of  $\iota$ , denoted  $\text{diam}_\iota(\mathbb{Q}^n)$ , as

$$\text{diam}_\iota(\mathbb{Q}^n) = \max_{s \in S(\mathbb{Q}^n)} \max_{x,y \in \|s\|} d(\iota(x), \iota(y)).$$

#### Lemma 40 (Equivalence of simplicial and regular disambiguations)

1. *For every disambiguation  $\mathcal{D}$  of  $\mathbb{S}^n$  there is  $\varepsilon > 0$  such that for every simplicial  $n$ -sphere  $\mathbb{Q}^n$  together with an equivariant homeomorphism  $\iota: \|\mathbb{Q}^n\| \rightarrow \mathbb{S}^n$  such that  $\text{diam}_\iota(\mathbb{Q}^n) < \varepsilon$  there is a disambiguation  $\mathcal{D}'$  of  $\mathbb{Q}^n$  such that  $\mathcal{D}' \leq \mathcal{D}$  and  $\mathcal{D}'$  is antipodal whenever  $\mathcal{D}$  is;*
2. *Let  $\mathcal{D}$  be a disambiguation of a simplicial  $n$ -sphere  $\mathbb{Q}^n$ . Then for the barycentric subdivision  $\mathbb{Q}_1^n$  of  $\mathbb{Q}^n$ , there is a disambiguation  $\mathcal{D}'$  of  $\|\mathbb{Q}_1^n\|$  such that  $\mathcal{D}' \equiv \mathcal{D}$  and  $\mathcal{D}'$  is antipodal whenever  $\mathcal{D}$  is.*

Here the relations  $\leq$  and  $\equiv$  between classes are as defined at the beginning of the section. Note that Proposition 25 in Section A plus compactness imply that  $\text{diam}_\iota(\mathbb{Q}_k^n) \xrightarrow{k} 0$ , for any simplicial  $n$ -sphere  $\mathbb{Q}^n$ . This implies that for any simplicial  $n$ -sphere  $\mathbb{Q}^n$ , for example for the barycentric subdivision of the boundary of  $(n+1)$ -simplex  $\mathbb{B}^n$  (Example 1), a sufficiently high barycentric subdivision of  $\mathbb{Q}^n$  would satisfy the condition from (1). We also note that several other equivalences can be formulated along the lines of Theorem 40. In particular, the disambiguation  $\mathcal{D}'$  in (2) turns out to be simplicial, that is, constant on the relative interiors of simplices of  $\mathbb{Q}_1^n$ . Here we stick to the bare minimum needed for us to proceed.

**Proof** (I). Let  $\mathcal{U} = \mathcal{U}(\mathcal{D})$ ; as  $\mathcal{D}$  disambiguates  $\mathbb{S}^n$ ,  $\mathcal{U}$  is a cover of  $\mathbb{S}^n$ . By a standard compactness argument (also used in the proof of Proposition 39), there is  $\varepsilon > 0$  such that for every  $x \in \mathbb{S}^n$ ,  $U_\varepsilon(x) \subseteq U$  for some  $U \in \mathcal{U}$ . For this  $\varepsilon$ , if  $\text{diam}(\iota) < \varepsilon$ , let  $\mathcal{D}'$  be the restriction of  $\mathcal{D}$  to  $V(\mathbb{Q}^n)$ , that is,  $\mathcal{D}' = \{h' \mid h \in \mathcal{D}\}$ , where  $h'(v) = h(\iota(v))$ . By construction,  $\mathcal{D}' \leq \mathcal{D}$  and it is antipodal whenever  $\mathcal{F}$  is. We only now need to check that  $\mathcal{D}'$  disambiguates  $\mathbb{Q}^n$ .

Let  $s$  be an arbitrary simplex of  $\mathbb{Q}^n$ , and let  $v \in s$ . By our choice of  $\varepsilon$ , there is  $h \in \mathcal{H}$  such that  $U_\varepsilon(\iota(v)) \subseteq U(h)$ . As  $\text{diam}(\iota) < \varepsilon$ , for every  $u \in s$ ,  $d(\iota(v), \iota(u)) < \varepsilon$ , so  $\iota(u) \in U_\varepsilon(\iota(v)) \subseteq U(h)$ , for every  $u \in s$ . By the definition of  $U(f)$ , this means that  $h'(u) = h(\iota(u)) = +$  for every  $u \in s$  and  $h'(u) = h(\iota(u)) = -$  for every  $u \in -s$ . Thus,  $h'$  witnesses the disambiguation on the simplex  $s$ . As  $s$  is arbitrary,  $\mathcal{D}'$  disambiguates  $\mathbb{Q}^n$ , as needed.

(2). Recall that, by the definition of (abstract) barycentric subdivision, vertices of  $\mathbb{Q}_1^n$  are nonempty simplices of  $\mathbb{Q}^n$ ,  $V(\mathbb{Q}_1^n) = S^{>0}(\mathbb{Q}^n)$ , and simplices of  $\mathbb{Q}_1^n$  are chains of nonempty simplices of  $\mathbb{Q}^n$ . As  $V(\mathbb{Q}_1^n)$  are treated both like vertices and like simplices, this can be confusing, so we illustrate the construction and the proof in Figure 9 below. Let us define  $\rho: S^{>0}(\mathbb{Q}^n) \rightarrow V(\mathbb{Q}^n)$  in an arbitrary way, satisfying two properties:

- For a simplex  $s$ ,  $\rho(s) \in s$ ;
- For a simplex  $s$ ,  $\rho(-s) = -\rho(s)$ .

For  $h \in \mathcal{D}$ , let us define  $h'$  on  $\|\mathbb{Q}_1^n\|$  as follows. For  $x \in \|\mathbb{Q}_1^n\|$ , let  $s \in S(\mathbb{Q}_1^n)$  be the support of  $x$ ,  $s = \text{supp}(x)$ , that is, the smallest simplex  $s \in S(\mathbb{Q}_1^n)$  such that  $x \in \|s\|$ . Let  $v$  be the minimal, as a simplex in  $S^{>0}(\mathbb{Q}^n)$ , vertex of  $s$ . Finally, for  $h \in \mathcal{D}$ , we put  $h'(x) = h(\rho(v))$ . It is easy to see that for  $\mathcal{D}' = \{h' \mid h \in \mathcal{D}\}$ ,  $\mathcal{D}' \equiv \mathcal{D}$  and  $\mathcal{D}'$  is antipodal whenever  $\mathcal{D}$  is. Again, we need to check that  $\mathcal{D}'$  disambiguates  $\|\mathbb{Q}_1^n\|$ .

Let  $x \in \|\mathbb{Q}_1^n\|$ ,  $s = \text{supp}(x)$ ,  $v$  the minimal vertex of  $s$ ,  $v \in V(\mathbb{Q}_1^n) = S^{>0}(\mathbb{Q}^n)$ , and  $h \in \mathcal{D}$  be a hypothesis that is constant  $+$  on  $v$  and constant  $-$  on  $-v$ . It is enough to check that for every maximal simplex  $s'$ , extending  $s$ , and for every  $x'$  in the relative interior of  $s'$  (that is, for those  $x'$  for which  $s' = \text{supp}(x')$ ), it holds  $h'(x') = +$  and  $h'(-x') = -$ . As  $s$  is a subchain of  $s'$ , we have  $v' \subseteq v$ , where again we treat  $v'$  and  $v$  as simplices in  $S^{>0}(\mathbb{Q}^n)$ . So  $\rho(v') \in v' \subseteq v$  and, by the choice of  $h$ ,  $h'(x') = h(\rho(v')) = +$ , and similarly,  $h'(-x') = h(\rho(-v')) = -$ , as needed. ■

#### C.4. Equivalence with spherical dimension

Let us now draw the main connection between the spherical dimension and disambiguations of spheres.

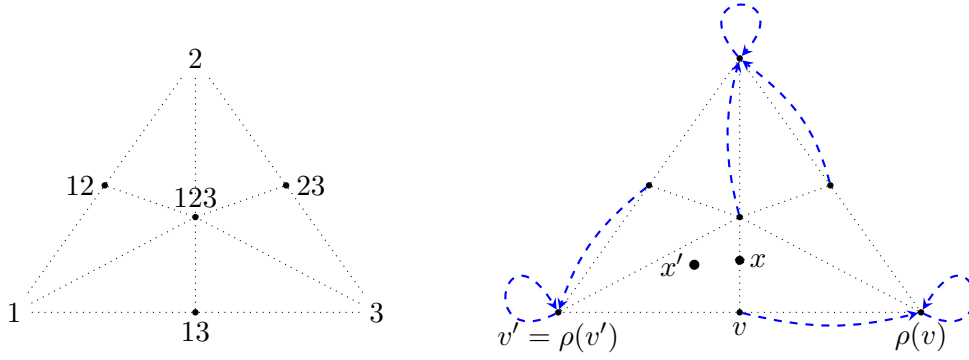


Figure 9: A barycentric subdivision of a simplex  $\{1, 2, 3\}$  and an illustration for the proof. In the second picture, the blue lines indicate  $\rho$ . For a point  $x$ ,  $s = \{13, 123\}$ , so  $v = 13$ ; for  $x'$ ,  $s' = \{1, 13, 123\}$  and  $v' = 1 \subseteq 13 = v$ .

**Theorem 4 (Rough equivalence of disambiguations and spherical dimension)**

1. Let  $\mathcal{D}$  be an antipodal disambiguation of a simplicial  $n$ -sphere  $\mathbb{Q}^n$ . Then both  $\mathcal{D}$  and an arbitrary restriction to representatives  $\mathcal{D}'$  of  $\mathcal{D}$  admit an embedded simplicial  $n$ -sphere.
2. Let  $\mathcal{H}$  be a class over  $X$  admitting a simplicial  $n$ -sphere  $\beta: V(\mathbb{Q}^n) \rightarrow V(\Delta_{\mathcal{H}}^{\text{ant}}) \subseteq X \times Y$ , and let  $\mathcal{H}^a$  be its antipodal extension to  $X^a = X \times Y$ . Then  $\beta$ -pullback  $\mathcal{H}_\beta$  of  $\mathcal{H}^a$  is an antipodal disambiguation of  $\mathbb{Q}^n$ .

Here  $\mathcal{H}_\beta$  is a class on  $V(\mathbb{Q}^n)$  defined as  $\mathcal{H}_\beta = \{h_\beta \mid h \in \mathcal{H}^a\}$ , where  $h_\beta(v) = h(\beta(v))$ , for  $v \in V(\mathbb{Q}^n)$ . In particular,  $\mathcal{H}_\beta \leq \mathcal{H}^a$ .

In particular, for integers  $a$ ,  $b$ , and  $n$ , the following are equivalent. Below,  $\mathbb{Q}^n$  denotes a simplicial  $n$ -sphere:

- There is a class with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ , and  $\text{sd}_{\text{simp}} \geq n$ ;
- There is a class with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ , and  $\text{sd}_{\text{simp}} \geq n$ , witnessed by embedded spheres;
- There is an antipodal disambiguation of  $\mathbb{Q}^n$  with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ ;
- There is a disambiguation of  $\mathbb{Q}^n$  with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ ;
- There is an antipodal disambiguation of  $\mathbb{S}^n$  with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ ;
- There is a disambiguation of  $\mathbb{S}^n$  with  $\text{VC} \leq a$ ,  $\text{VC}^{*a} \leq b$ .

In particular, if for a class  $\mathcal{H}$  we define the disambiguation index  $\text{disamb}(\mathcal{H})$  as the maximal  $n$  for which there is a disambiguation  $\mathcal{D}$  of  $\mathbb{S}^n$  such that  $\mathcal{D} \leq \mathcal{H}$ , then the parameters  $\text{sd}_{\text{simp}}$  and  $\text{disamb}$  are roughly equivalent (and also roughly equivalent to  $\text{sd}_{\text{simp}}$ , witnessed by embedded spheres).

We also note that in the “in particular” case, the respective statements are applicable not only to  $\text{VC}$  and  $\text{VC}^{*a}$ , but to all class characteristics that are non-increasing under  $\leq$ , symmetrization, and antipodal extension.

**Proof** Let us start by recalling that a finite class  $\mathcal{H}$  admits an embedded simplicial  $n$ -sphere if there is a simplicial  $n$ -sphere  $\mathbb{Q}^n$  and a one-to-one equivariant map  $\beta: V(\mathbb{Q}^n) \rightarrow V(\Delta_{\mathcal{H}}^{\text{ant}}) \subseteq X \times Y$  such that  $\beta$  maps simplices of  $\mathbb{Q}^n$  to simplices of  $\Delta_{\mathcal{H}}^{\text{ant}}$ .

(1.) For  $\mathcal{D}$  and  $\mathcal{D}'$  as in the statement, let  $X = r[V(\mathbb{Q}^n)] \subseteq V(\mathbb{Q}^n)$ , where  $r$  is the respective representation map for which  $\mathcal{D}' = \mathcal{D}^r$ . Note that  $\mathcal{D}'$  is defined on  $X$  and is not antipodal. Let us now define  $\beta: V(\mathbb{Q}^n) \rightarrow X \times Y$  as follows: For  $v \in V(\mathbb{Q}^n)$ ,  $\beta(v) = (v, +)$  if  $v = r(v)$  and  $(-v, -)$  if  $v = -r(v)$ . Alternatively,  $\beta(v) = (r(v), [r(v) = v])$ , where, similarly to Lemma 37,  $[r(v) = v] = +$  if  $r(v) = v$  and  $-$  if  $r(v) = -v$ . Trivially,  $\beta$  is one-to-one. Moreover,  $\beta(-v) = (r(-v), [r(-v) = -v]) = (r(v), [r(v) = -v]) = (r(v), -[r(v) = v]) = -\beta(v)$ , and so  $\beta$  is also equivariant.

We also claim that  $\beta$  is simplicial to  $\Delta_{\mathcal{D}'}$ . Indeed, let us take an arbitrary  $s \in S(\mathbb{Q}^n)$  and let  $h_s \in \mathcal{D}$  be such that  $h_s \equiv +$  on  $s$ ; note that this automatically implies that  $h(-s) \equiv -$ , as  $\mathcal{D}$  is antipodal. But then for any  $v \in s$ ,  $h_s(r(v)) = [r(v) = v]$ , and so  $\beta(v) = (r(v), h_s(r(v)))$ . Thus,  $\beta(s) = \{(r(v), h_s(r(v))) \mid v \in s\}$  is realizable by  $h_s|_X \in \mathcal{D}'$ , and hence belongs to  $\Delta_{\mathcal{D}'}$ . The fact that  $\beta(s)$  is actually in  $\Delta_{\mathcal{D}'}^{\text{ant}}$  follows by noting that  $-\beta(s) = \beta(-s) \in \Delta_{\mathcal{D}'}$ , where the first equality is by equivariance of  $\beta$ , and the second inclusion is witnessed by some other function  $h_{-s}$ . This concludes the proof for  $\mathcal{D}'$ ; the statement for  $\mathcal{D}$  follows from that because  $\Delta_{\mathcal{D}'}$  is (simplicially) embedded into  $\Delta_{\mathcal{D}}$  with the same antipodality.

This proof bears mixed features of both triviality and evil magic. To rectify this, we illustrate how it works (and fails) in Figure 10 below.

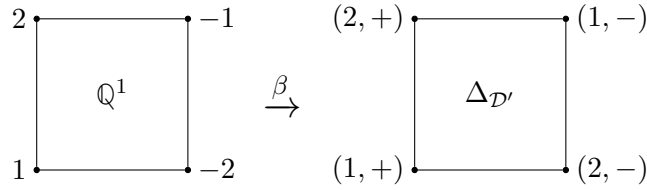


Figure 10: The map  $\beta$  from the above proof. Note that for the function, say,  $h_{2,-1} \in \mathcal{D}$ , positive on the vertices of the top simplex in  $\mathbb{Q}^1$ ,  $h_{2,-1}(1) = -$  and  $h_{2,-1}(2) = +$ , and so it witnesses the existence of the same top simplex in  $\Delta_{\mathcal{D}'}$ . And, in the absence of  $h_{2,-1}$  in  $\mathcal{D}$ , the respective simplex would be absent in  $\Delta_{\mathcal{D}'}$ , breaking the simpliciality of the map.

(2.) We only need to check that  $\mathcal{H}_\beta$  is antipodal and that it is a disambiguation. For antipodality, quite trivially,  $h_\beta(-v) = h(\beta(-v)) = h(-\beta(v)) = -h(\beta(v)) = -h_\beta(v)$ , for all  $h_\beta \in \mathcal{H}$  and  $v \in V(\mathbb{Q}^n)$ ; here we use antipodality of  $\beta$  and antipodality of  $h \in \mathcal{H}^a$ .

For disambiguation, let  $s \in S(\mathbb{Q}^n)$ , and let  $h_s \in \mathcal{H}$  be the concept witnessing  $s \in \Delta_{\mathcal{H}}$ , that is,  $h_s(x) = y$  for all  $(x, y) \in \beta(s)$ . But then  $h_s^a(x, y) = +$  for all  $(x, y) \in \beta(s)$ , which is equivalent to  $(h_s^a)_\beta(v) = +$  for all  $v \in s$ , and hence  $(h_s^a)_\beta(s) \equiv +$ , as needed; note that  $(h_s^a)_\beta(-s) \equiv -$  follows by antipodality.

The final list of equivalences is just tallying up the results from this section, shown in Figure 11 below. ■

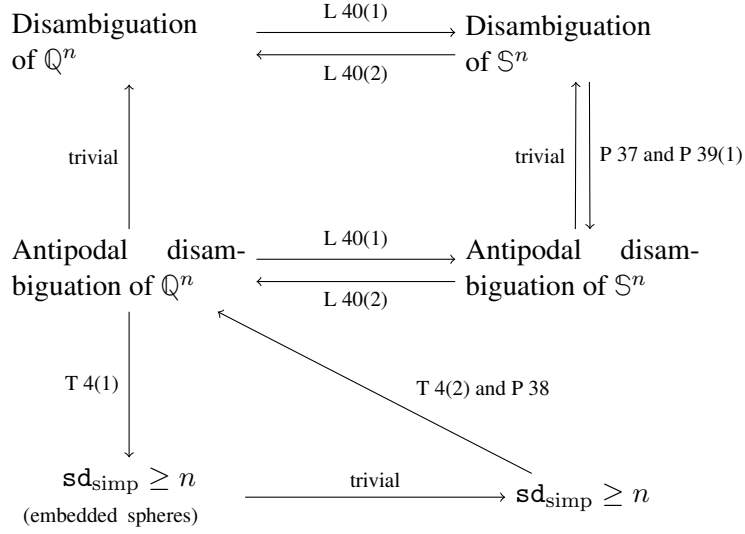


Figure 11: Implications between statements about classes with  $VC \leq a$  and  $VC^{*a} \leq b$  from the conclusion of Theorem 4.

## Appendix D. Proofs for extremal classes

As some proofs in this section are rather involved, here we will drop our convention for not using geometrization brackets  $\|\cdot\|$  for the spaces  $\Delta$  and  $\Delta^{\text{ant}}$ . That is, we will write  $\Delta$  and  $\Delta^{\text{ant}}$  for simplicial complexes and  $\|\Delta\|$  and  $\|\Delta^{\text{ant}}\|$  for their geometric realizations.

### D.1. Cubical complex of a class

Our proof will crucially rely on the notion of a *cubical complex* of a class. Similarly to simplicial complexes, cubical complexes can be defined abstractly; See, for example, Definition 7.32 in [Bridson and Haefliger \(2013\)](#), or [Sageev \(2014\)](#) for a more informal walkthrough. Such formal approach, however, is unnecessary meticulous and, following Section 2 in [Chase et al. \(2024a\)](#), we will make a shortcut and will construct our cubical complex as a subspace of the usual Euclidean cube (composed out of faces of this cube).

Our first technical step is to introduce *strong shattering*. It will be convenient for us to use partial hypotheses, so we will use some basic notation and terminology from the beginning of Section 3. For a partial hypothesis  $h$  over a finite domain  $X$ , we define the *dimension* of  $h$ , denoted  $d(h)$ , as  $d(h) = |\{x \in X \mid h(x) = *\}| = |X| - |\text{supp}(h)|$ . We define the *discrete cube* of  $h$ , denoted  $C(h)$ , as a set of all  $2^{d(h)}$  total hypotheses, extending  $h$ . For a class  $\mathcal{H}$  over  $X$  and  $S \subseteq X$ , we say that  $S$  is *strongly shattered* by  $\mathcal{H}$  if there is a partial hypotheses  $h$  such that  $S = X - \text{supp}(h)$  and  $C(h) \subseteq \mathcal{H}$ . Trivially, every set strongly shattered by a class is shattered by it. It is well-known that for extremal classes strong shattering is equivalent to shattering, see, for example, [Bollobás and Radcliffe \(1995\)](#); moreover, this correspondence goes both ways, that is, a class is extremal if and only if it strongly shatters all its shattered sets. This will be crucial for us in the construction of the cubical complex, which revolves around strong shattering. So, although the same definition can be considered for an arbitrary class  $\mathcal{H}$ , we will now restrict our attention to a finite extremal class  $\mathcal{E}$ .

Recall that in Section 3 we defined the order on partial hypotheses by extension by saying that  $h_1 \leq h_2$  if  $\text{supp}(h_1) \subseteq \text{supp}(h_2)$  and  $h_2|_{\text{supp}(h_1)} = h_1|_{\text{supp}(h_1)}$ . It is easy to see that this can be equivalently formulated as  $h_1 \leq h_2$  if  $C(h_1) \supseteq C(h_2)$ . From this perspective, it would be natural to flip the order on partial hypotheses to make it coincide with the order on cubes by inclusion; however, we stick with the present one for consistency with Section 3. It is also easy to see that discrete cubes and partial hypotheses over  $X$  are in bijective correspondence to each other, so with some ambiguity, we would at times refer to one as to another. In particular, we define the dimension of a cube as the dimension of the corresponding partial hypothesis and refer to cubes of dimension  $d$  as to  $d$ -cubes.

For an extremal class  $\mathcal{E}$ , we say that a discrete cube  $C$  is a *cube of  $\mathcal{E}$*  if  $C \subseteq E$ , and we define an (abstract) cubical complex of  $\mathcal{E}$ , denoted  $\mathbb{C}_{\mathcal{E}}$ , as a collection of all discrete cubes of  $\mathcal{E}$ . Note that here we do not allow our cubes to be empty. Let us note several basic facts about the cubical complex  $\mathbb{C}_{\mathcal{E}}$  that we are going to utilize:

- The dimension of  $\mathbb{C}_{\mathcal{E}}$ , denoted  $\dim(\mathbb{C}_{\mathcal{E}})$  is defined as a maximal dimension of a cube in  $\mathbb{C}_{\mathcal{E}}$ . By the abovementioned correspondence between shattering and strong shattering for extremal classes,  $\dim(\mathbb{C}_{\mathcal{E}}) = \text{VC}(\mathcal{E})$ ;
- 0-cubes of  $\mathbb{C}_{\mathcal{E}}$ , also called *vertices* of  $\mathbb{C}_{\mathcal{E}}$ , are precisely the concepts of  $\mathcal{E}$ . We also at times call 1-cubes by *edges* and 2-cubes by *squares*;
- We say that  $C_1$  is a *subcube* of  $C_2$  if  $C_1 \subseteq C_2$ . As argued, for the respective partial hypotheses  $h_1$  and  $h_2$ , this is equivalent to  $h_2$  extending  $h_1$ , that is,  $C_1 \subseteq C_2 \sim h_1 \geq h_2$ ;
- Any subcube of a cube of  $\mathbb{C}_{\mathcal{E}}$  is a cube of  $\mathbb{C}_{\mathcal{E}}$ . Moreover, any two cubes of  $\mathbb{C}_{\mathcal{E}}$  are either disjoint or intersect by a cube of  $\mathbb{C}_{\mathcal{E}}$ . This property enables us to indeed call  $\mathbb{C}_{\mathcal{E}}$  a *cube complex*, as opposed to it being just a collection of cubes.

Let  $I = [-1, 1]$  and, for a finite domain  $X$ , let  $I^X$  be a  $|X|$ -dimensional Euclidean cube with coordinates parametrized by  $X$ . For a discrete cube  $C = C(h)$  over  $X$  we define the *geometric realization* of  $C$ , denoted  $\|C\| \subseteq I^X$ , as

$$\|C\| = \{v \in I^X \mid v(x) = h(x) \text{ for } x \in \text{supp}(h)\}.$$

Finally, we define the geometric cubical complex of  $\mathcal{E}$ , denoted  $\|\mathbb{C}_{\mathcal{E}}\|$ , or sometimes simply  $\|\mathcal{E}\|$ , as  $\|\mathcal{E}\| = \bigcup_{C \in \mathbb{C}_{\mathcal{E}}} \|C\|$ . This construction is illustrated in Figure 12 below.

The geometric realization of cubes and complexes enjoy nice properties, similar to those listed for abstract cubical complexes:

- $\dim(\|\mathcal{E}\|) = \dim(\mathbb{C}_{\mathcal{E}}) = \text{VC}(\mathcal{E})$ ;
- $C_1 \subseteq C_2$  if and only if  $\|C_1\| \subseteq \|C_2\|$ ;  $C_1$  and  $C_2$  are disjoint if and only if  $\|C_1\|$  and  $\|C_2\|$  are disjoint; and, if  $C_1$  and  $C_2$  are not disjoint, then  $\|C_1\| \cap \|C_2\| = \|C_1 \cap C_2\|$ .

## D.2. Barycentric subdivisions of the cubical complex and the space of realizable distributions of a class

For a finite extremal class  $\mathcal{E}$ , we define the *barycentric subdivision* of the cubical complex  $\mathbb{C}_{\mathcal{E}}$ , denoted  $\mathbb{C}_{\mathcal{E},1}$ , as a simplicial complex whose vertices are cubes of  $\mathbb{C}_{\mathcal{E}}$  and whose simplices are chains of cubes of  $\mathbb{C}_{\mathcal{E}}$ .

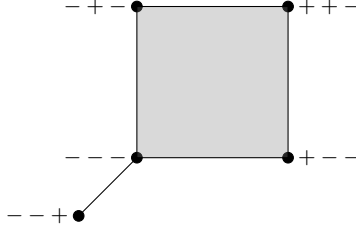


Figure 12: A 2-dimensional illustration of the cubical complex for the extremal class  $\mathcal{E} = \{- - -, - + -, + + -, + - -, - - +\}$ . It has 5 vertices, 5 edges, and 1 square. The square corresponds to a cube  $** -$ , and the unique maximal edge, connecting  $- - -$  and  $- - +$ , corresponds to a cube  $- - *$ .

The respective geometrization of  $\mathbb{C}_{\mathcal{E},1}$  is then defined as follows. Recall that the cube complex  $\|\mathcal{E}\|$  is naturally embedded into  $I^X$ , where  $X$  is the domain of  $\mathcal{E}$ . Let  $\iota: \{-, +, *\} \rightarrow I$  be defined as  $\iota(-) = -1$ ,  $\iota(+)=1$ , and  $\iota(*)=0$ . We then extend  $\iota$  to the map  $\iota: \{-, +, *\}^X \rightarrow I^X$ , whose domain are all partial hypotheses over  $X$ , by letting  $\iota(h) = (\iota(h(x)))_{x \in X}$ .

Note that  $\iota$  maps the discrete cubes (interpreted as partial hypotheses), in a way, consistent with the definition of the geometric realization of these cubes. In particular, for any discrete cube  $C$ ,  $\|C\|$  is a convex hull of  $\iota(C) = \{\iota(h) \mid h \in C\}$ ; and for any partial hypothesis  $h$ ,  $\|h\| = \iota(h)$ . Thus,  $\iota$  provides a natural embedding of the vertices of  $\mathbb{C}_{\mathcal{E},1}$ , which are cubes of  $\mathcal{E}$ , into  $\|\mathcal{E}\|$ . It can now be easily checked that, extended by linearity,  $\iota$  establishes an isomorphism between the geometric realization  $\|\mathbb{C}_{\mathcal{E},1}\|$  (or simply  $\|\mathcal{E}_1\|$ ) of the barycentric subdivision of  $\mathbb{C}_{\mathcal{E}}$ , and the geometric realization  $\|\mathcal{E}\|$  of  $\mathbb{C}_{\mathcal{E}}$ . As before, we usually will interpret this as if  $\|\mathcal{E}_1\|$  is itself embedded into  $I^X$ . This is illustrated in Figure 13 below. Once more:  $\|\mathcal{E}\|$  and  $\|\mathcal{E}_1\|$  are the same topological space and we only use different names for it when we want to emphasize the underlying simplicial complex from which it arises.

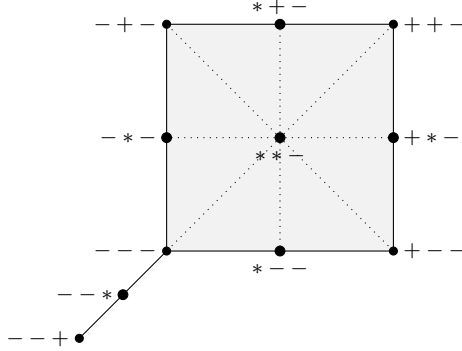


Figure 13: A barycentric subdivision of the cubical complex in Figure 9. The corresponding simplicial complex has 11 vertices, 18 edges, and 6 2-dimensional simplices.

To define the barycentric subdivision of  $\Delta_{\mathcal{E}}$ , denoted  $\Delta_{\mathcal{E},1}$ , all we need is to apply the standard definition of the barycentric subdivision from Section A to the definition of  $\Delta_{\mathcal{E}}$  in Section A.2. Thus, the vertices of  $\Delta_{\mathcal{E},1}$  are partial hypotheses  $h$  with nonempty support that can be extended to a total hypothesis in  $\mathcal{E}$  (we say that  $h$  is realizable by  $\mathcal{E}$ ), and the simplices are chains of such



partial hypotheses. Geometrically, a partial hypothesis  $h \in V(\Delta_{\mathcal{E}})$  is interpreted as a uniform distribution over its (nonempty) domain, and points in simplices are convex combinations of uniform distributions over nested domains. We also note that because  $\Delta_{\mathcal{E}}^{\text{ant}}$  is simplicially embedded into  $\Delta_{\mathcal{E}}$ , it follows that  $\Delta_{\mathcal{E},1}^{\text{ant}}$  is an antipodal simplicial complex, simplicially embedded into  $\Delta_{\mathcal{E},1}$ , with antipodality extended in a standard way to the barycentric subdivision.

With this elaboration, the parallel between the barycentric subdivisions of  $\mathbb{C}_{\mathcal{E}}$  and of  $\Delta_{\mathcal{E}}$  becomes quite pronounced:

- The vertex sets in both cases are subsets of the set of all partial hypotheses over  $X$ . The simplices are also formed in a similar way, namely, as chains of partial hypotheses;
- For  $\Delta_{\mathcal{E},1}$ , we pick partial hypotheses that are *nonempty* and realizable by  $\mathcal{E}$ , and for  $\mathbb{C}_{\mathcal{E}}$  those that correspond to cubes of  $\mathcal{E}$ . Trivially, any  $h$  corresponding to a cube is also realizable, and so the only partial hypotheses that can be in  $V(\mathbb{C}_{\mathcal{E},1}) - V(\Delta_{\mathcal{E},1})$  is the completely undefined hypothesis  $h_*$ , which only happens when  $\mathcal{E}$  is the binary hypercube  $\mathcal{C}_n$ ;
- If  $\mathcal{E} = \mathcal{C}_n$ , then  $\|\mathcal{E}\| = I^n$  and  $\|\Delta_{\mathcal{E}}\| = \|\Delta_{\mathcal{E}}^{\text{ant}}\| \cong \mathbb{S}^{n-1}$  is its boundary. Moreover,  $\Delta_{\mathcal{E},1}$  is simplicially embedded into  $\mathbb{C}_{\mathcal{E},1}$ ;
- In all other cases, that is, when  $\mathcal{E} \neq \mathcal{C}_n$ ,  $\mathbb{C}_{\mathcal{E},1}$  is simplicially embedded into  $\Delta_{\mathcal{E},1}$ . Moreover,  $\mathbb{C}_{\mathcal{E},1}$  is a *full* subcomplex of  $\Delta_{\mathcal{E},1}$ , that is, for any simplex  $s \in S(\Delta_{\mathcal{E},1})$ , if  $v \in V(\mathbb{C}_{\mathcal{E},1})$  for all  $v \in s$ , then  $s \in S(\mathbb{C}_{\mathcal{E},1})$ .

Figure 14 below illustrates the embedding of  $\mathbb{C}_{\mathcal{E},1}$  into  $\Delta_{\mathcal{E},1}$  for a particular extremal class.

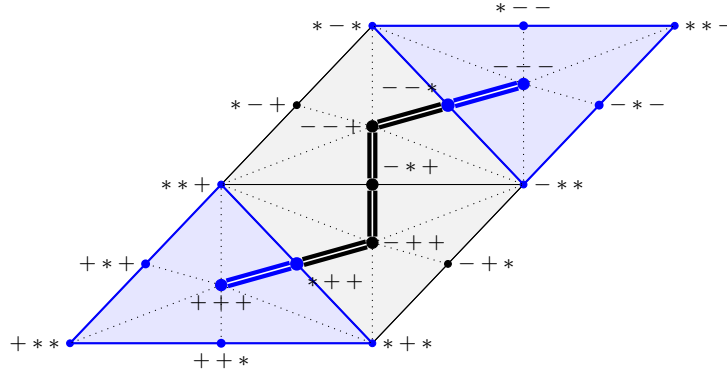


Figure 14: The barycentric subdivision of the cubical complex  $\mathbb{C}_{\mathcal{E},1}$  simplicially embedded into the barycentric subdivision of the space of realizable distributions  $\Delta_{\mathcal{E},1}$  for the extremal class  $\mathcal{E} = \{---, --+, -++, +++)\}$ . Antipodal regions are indicated in blue. Note that the point, say,  $(1, -) \in X \times Y$  from the original definition of  $\Delta_{\mathcal{E}}$  (see Section A.2 and Figure 1), when restated in terms of partial hypotheses, becomes  $-**$ .

### D.3. Deformation retraction

Having defined a simplicial embedding of  $\mathbb{C}_{\mathcal{E},1}$  into  $\Delta_{\mathcal{E},1}$ , we are now going to prove that the former is a deformation retract of the latter. So let us briefly recall the respective topological definitions.

For a topological space  $B$  and a subspace  $A \subseteq B$ , a map  $f: B \rightarrow A$  is called a *retraction* if  $f$ , restricted to  $A$ , is an identity, that is, for all  $a \in A$ ,  $f(a) = a$ . A *deformation retraction* in this setup is a map  $F: [0, 1] \times B \rightarrow B$  such that  $F(0, \cdot)$  is an identity on  $B$ ,  $F(1, B) = A$ , and  $F(t, \cdot)$  is an identity on  $A$  for all  $t \in [0, 1]$ , see Section 1.2 in [Matoušek et al. \(2003\)](#). The subspace  $A$  is called a (*deformation*) *retract* of  $B$  if there is a (deformation) retraction from  $B$  to  $A$ . Note that for a deformation retraction  $F$  from  $B$  to  $A$ , the map  $F(1, \cdot)$  is a retraction from  $B$  to  $A$ , and so every deformation retract is a retract (the opposite is not true in general). Being a deformation retract is a rather strong property, in particular, this is a special case of homotopy equivalence.

It is quite clear how to define a *simplicial retraction*: For a subcomplex  $K$  of  $L$  this is simply a simplicial map  $f: V(L) \rightarrow V(K)$  which is an identity on  $V(K)$ . At the same time, we do not know of a standard definition of a simplicial deformation retraction. A good reason for that is that  $L \times [0, 1]$ , which would be a domain of the respective  $F$ , does not naturally inherit the simplicial structure from  $L$ . This, however, is a typical obstacle in defining simplicial homotopy, which is usually resolved via prism operators (see proof of Theorem 2.10 in [Hatcher \(2002\)](#)), so probably a reasonable definition can be given along these lines. In any case, this goes far beyond the aim of the present paper, so we will not try to enforce simpliciality on the deformation retraction that we are going to construct.

Finally, a topological space  $B$  is called *contractible* if there exists a *contraction map*  $F: [0, 1] \times B \rightarrow B$  to a point  $x \in B$ , that is,  $F(0, \cdot)$  is an identity on  $B$  and  $F(1, \cdot) \equiv x$ .

#### D.4. Restrictions

For a finite extremal class  $\mathcal{E}$  on  $X$  and a partial hypothesis  $h$  on  $X$ , we define an  *$h$ -restriction* of  $\mathcal{E}$ , denoted  $\mathcal{E}_h$ , as a class on  $X$  defined as the class of all total hypotheses of  $\mathcal{E}$ , extending  $h$ , that is

$$\mathcal{E}_h = \{g \in \mathcal{E} \mid g \geq h\}.$$

We note that this is closely related (but not identical) to how restrictions are defined in [Moran \(2012\)](#) (Definition 3.5). In particular, it is known and easy to check that the following properties hold for restrictions:

- $\mathcal{E}_h \subseteq \mathcal{E}$  and it is nonempty whenever  $h$  is realizable by  $\mathcal{E}$ . Also, for the completely undefined hypothesis  $h_*$ ,  $\mathcal{E}_{h_*} = \mathcal{E}$ ;
- For any  $h$ , realizable by  $\mathcal{E}$ ,  $\mathbb{C}_{\mathcal{E}_h,1}$  is a full subcomplex of  $\mathbb{C}_{\mathcal{E},1}$ , spanned by the vertices  $\{g \in V(\mathbb{C}_{\mathcal{E},1}) \mid g \geq h\}$ ;
- For any  $h \neq h_*$ , there is no  $g \geq h, -h$ ; In particular,  $\|\mathcal{E}_{h(x),1}\|$  and  $\|\mathcal{E}_{-h(x),1}\|$  are disjoint. As an easy corollary, for  $x \in \|\Delta_{\mathcal{E},1}\|$ , let us define  $h(x) \in V(\Delta_{\mathcal{E},1})$  as the minimal vertex of the support of  $x$ . Then  $\|\mathcal{E}_{h(x),1}\|$  and  $\|\mathcal{E}_{h(-x),1}\|$  are disjoint for all  $x \in \|\Delta_{\mathcal{E},1}^{\text{ant}}\|$ .

The usefulness of restrictions comes from the fact that they enable us to associate vertices of  $\Delta_{\mathcal{E},1}$ , parametrized by realizable partial hypotheses of  $\mathcal{E}$ , with nonempty subcomplexes of  $\mathbb{C}_{\mathcal{E},1}$ . This is illustrated in Figure 15 below.

Let us also elaborate on the second bullet. If we define  $\mathbb{C}_{\mathcal{E},1}^h$  as a full subcomplex of  $\mathbb{C}_{\mathcal{E},1}$ , spanned by the vertices  $\{g \in V(\mathbb{C}_{\mathcal{E},1}) \mid g \geq h\}$ , then the claim was that  $\mathbb{C}_{\mathcal{E},1}^h = \mathbb{C}_{\mathcal{E}_h,1}$ ; its proof is by a straightforward consideration of what are the discrete cubes of  $\mathcal{E}_h$ . At the same time, let

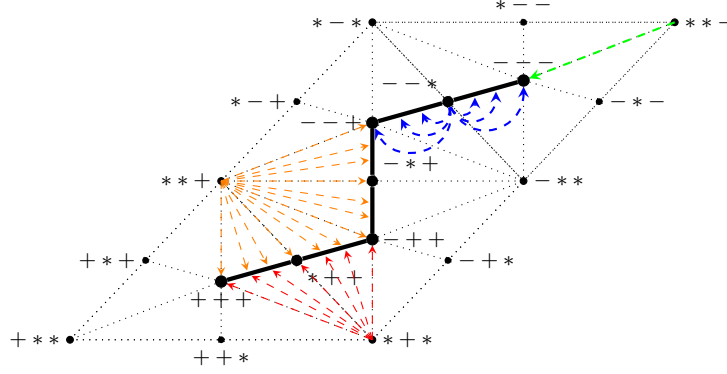


Figure 15: Subcomplexes of  $\mathbb{C}_{\mathcal{E},1}$  for  $\mathcal{E}$  from Figure 14, that correspond to restrictions to partial hypotheses  $*+*$ ,  $**+$ ,  $--*$ , and  $**-$ . Note that, as argued above, for antipodal vertices  $**+$  and  $**-$ , the geometric cube complexes for the corresponding restrictions are disjoint.

us similarly define  $\Delta_{\mathcal{E},1}^h$  as a full subcomplex of  $\Delta_{\mathcal{E},1}$  spanned by  $\{g \in V(\Delta_{\mathcal{E},1}) \mid g \geq h\}$ . One can notice that  $\Delta_{\mathcal{E},1}^h$  is, in fact, also a full subcomplex of  $\Delta_{\mathcal{E},1}$ , but spanned by a larger vertex set  $\{g \in V(\Delta_{\mathcal{E},1}) \mid \text{there is } w \in V(\Delta_{\mathcal{E},1}) \text{ s.t. } g, h \leq w\}$ . Hence,  $\Delta_{\mathcal{E},1}^h$  is a full subcomplex of  $\Delta_{\mathcal{E},1}$ , but the two are not, in general, equivalent. Note also that from the above it is easy to see that  $\|\mathcal{E}_{h,1}\| = \|\Delta_{\mathcal{E},1}^h\| \cap \|\mathcal{E}_1\|$ . This is illustrated in Figure 16 below.

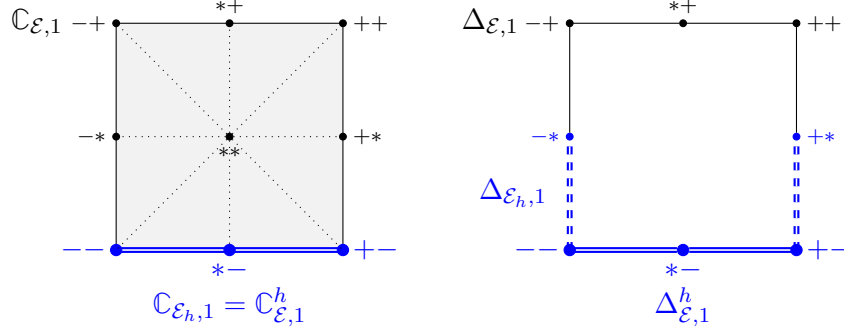


Figure 16: Complexes  $\mathbb{C}_{\mathcal{E},1} = \mathbb{C}_{\mathcal{E},1}^h$ ,  $\Delta_{\mathcal{E},1}$ , and  $\Delta_{\mathcal{E},1}^h$  for  $\mathcal{E} = \mathcal{C}_2 = \{--, -, +, ++\}$ ,  $h = *-$ , and  $\mathcal{E}_h = \{-+, ++\}$ .

Although most of the theory developed so far can, in principle, be applicable to arbitrary classes, we now give a crucial fact about cubical complexes of specifically extremal classes.

**Proposition 41** *For an extremal class  $\mathcal{E}$ :*

- The cubical complex  $\|\mathcal{E}\|$  is contractible, that is there is  $x \in \|\mathcal{E}\|$  and a contraction map  $F: [0, 1] \times \|\mathcal{E}\| \rightarrow \|\mathcal{E}\|$  such that  $F(0, \cdot)$  is the identity map and  $F(1, \cdot) \equiv x$ ;
- For any  $h$ , realizable by  $\mathcal{E}$ , the class  $\mathcal{E}_h$  is extremal. In particular, the cubical complex  $\|\mathcal{E}_h\|$  is contractible.

The above properties are Proposition 4.12 and Theorem 3.1(6) in Chalopin et al. (2022). They were also crucial ingredients for the abovementioned  $\text{VC}^* \geq 2\text{VC} + 1$  bound for extremal classes from Chase et al. (2024a), see Proposition 13 there. We also note that in Chalopin et al. (2022) they establish *collapsibility*, which is a stronger version of contractibility.

### D.5. Deformation retraction from $\|\Delta_{\mathcal{E}}\|$ to $\|\mathbb{C}_{\mathcal{E}}\|$

We are now ready to prove the central claim for this section. Note that we prove it in a much stronger way than originally stated in Section 4.

**Theorem 9** *Let  $\mathcal{E}$  be a finite extremal class that is not a binary hypercube  $\mathcal{C}_n$ , for any  $n$ .*

*Then  $\mathbb{C}_{\mathcal{E},1}$  is a full subcomplex of  $\Delta_{\mathcal{E},1}$  and there is a deformation retraction  $F$  from  $\|\Delta_{\mathcal{E},1}\|$  to  $\|\mathbb{C}_{\mathcal{E},1}\|$  such that for any  $x \in \|\Delta_{\mathcal{E},1}\|$  and  $t \in [0, 1]$ ,  $F(t, x) \in \|\Delta_{\mathcal{E},1}^{h(x)}\|$ , where  $h(x) \in V(\Delta_{\mathcal{E},1})$  is the minimal vertex of the support of  $x$ .*

*In particular, for the final map  $f_1$  of  $F$ ,  $f_1(x) = F(1, x)$ , it holds  $f_1(x) \in \|\mathbb{C}_{\mathcal{E},1}\|$ . In particular,  $f_1$  does not collapse antipodal points, that is,  $f_1(x) \neq f_1(-x)$  for any  $x \in \|\Delta_{\mathcal{E},1}^{\text{ant}}\|$ .*

*And, if  $\mathcal{E}$  is a binary hypercube  $\mathcal{C}_n$ , then  $\|\mathbb{C}_{\mathcal{E},1}\|$  is an  $n$ -dimensional cube, and  $\|\Delta_{\mathcal{E},1}\| = \|\Delta_{\mathcal{E},1}^{\text{ant}}\|$  is its boundary,  $\|\Delta_{\mathcal{E},1}^{\text{ant}}\| \cong (\mathbb{S}^{n-1}, x \mapsto -x)$ . In particular, the natural embedding of  $\|\Delta_{\mathcal{E},1}^{\text{ant}}\|$  into  $\|\mathbb{C}_{\mathcal{E},1}\|$  does not collapse antipodal points.*

*In both cases,  $\|\Delta_{\mathcal{H}}\|$  can be mapped into the cubical complex  $\|\mathbb{C}_{\mathcal{E}}\|$  of dimension  $\text{VC}(\mathcal{E})$  without collapsing any antipodal points in  $\|\Delta_{\mathcal{E}}^{\text{ant}}\|$ . In particular,  $\text{sd}(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) - 1$  and  $\text{VC}^*(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) + 1$ . Consequently, for any class  $\mathcal{H}$ ,  $\text{sd}(\mathcal{H}) \leq 2\text{VC}^{\text{extr}}(\mathcal{H}) - 1$ , or, equivalently,  $\text{VC}^{\text{extr}}(\mathcal{H}) \geq \lceil (\text{sd}(\mathcal{H}) + 1)/2 \rceil$ .*

Before the proof, let us give two cautionary examples. First of all, the situation in Figure 14, where  $\Delta^{\text{ant}} \cap \mathbb{C}$  is nonempty and closed under antipodality, is not, in general, true. Figure 17 below shows a simple extremal class for which  $\Delta^{\text{ant}}$  is nonempty, but  $\Delta^{\text{ant}} \cap \mathbb{C}$  is. In particular, this means that we cannot hope to make the final map  $f_1$  above antipodal. We also note that there are easy examples where  $\Delta^{\text{ant}} \cap \mathbb{C}$  is nonempty, but not closed under antipodality (for example, for  $\mathcal{E} = \{++++, -++-, --++\}$ ).

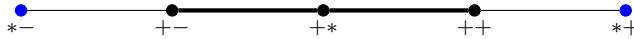


Figure 17: The embedding of  $\mathbb{C}_{\mathcal{E},1}$  into  $\Delta_{\mathcal{E},1}$  for the extremal class  $\mathcal{E} = \{+-, ++\}$ . The antipodal part  $\Delta_{\mathcal{E},1}^{\text{ant}}$  consists of two vertices  $*-$  and  $*+$  and is disjoint from  $\mathbb{C}_{\mathcal{E},1}$ .

As a second warning, recall that in Section D.3 we argued that although defining simplicial deformation retraction is problematic, the definition of simplicial retraction is straightforward. We thus might ask if the deformation retraction  $F$  above can be made such that the final map  $f_1$  is simplicial; note that  $f_1$  itself is a retraction from  $\|\Delta_{\mathcal{E},1}\|$  to  $\|\mathbb{C}_{\mathcal{E},1}\|$ . However, the answer to this question is no. This can easily be seen so for the vertex  $**+$  for the class in Figure 15. Even without the restriction  $f_1(**+) \in V(\Delta_{**+,1})$  (but assuming that  $f_1$  is a retraction, that is, that it keeps  $\mathbb{C}$  fixed), one can easily see that mapping  $**+$  to any vertex of  $\mathbb{C}$  would break simpliciality.

However, although for technical reasons we will not prove it, we note that the deformation retraction that we will construct will be *piecewise-linear*, which is the next best thing after simpliciality, see Bryant (2002) for some background. Moreover, from the construction, it would be

obvious that  $f_1$  can be made simplicial from a subdivision of  $\Delta_{\mathcal{E},1}$ . See Figure 18 below for an example of how this will look for the particular abovementioned vertex  $**+$ .

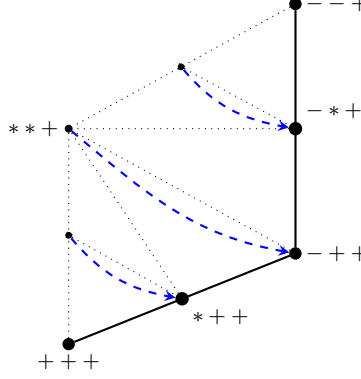


Figure 18: Piecewise-linear, but not simplicial, part of the retraction of  $\Delta_{\mathcal{E},1}$  to  $\mathbb{C}_{\mathcal{E},1}$  for  $\mathcal{E}$  from Figure 14.

**Proof** The case  $\mathcal{E} = \mathcal{C}_n$  has been discussed in Section D.2. In the same section, we also argued that  $\mathbb{C}_{\mathcal{E},1}$  is a full subcomplex of  $\Delta_{\mathcal{E},1}$  whenever  $\mathcal{E} \neq \mathcal{C}_n$ . Note that  $F(t, x) \in \|\Delta_{\mathcal{E},1}^{h(x)}\|$  and  $F(1, x) \in \|\mathcal{E}_1\|$  trivially imply  $f_1(x) = F(1, x) \in \|\Delta_{\mathcal{E},1}^{h(x)}\| \cap \|\mathcal{E}_1\| = \|\mathcal{E}_{h(x),1}\|$ . Also, the statement  $f_1(x) \neq f_1(-x)$  for any  $x \in \|\Delta_{\mathcal{E},1}^{\text{ant}}\|$  follows from the fact that, as argued in Section D.4,  $\|\mathcal{E}_{h(x),1}\|$  and  $\|\mathcal{E}_{h(-x),1}\|$  are disjoint.

The concluding statement  $\text{sd}(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) - 1$  follows by a straightforward application of Theorem 23 (Local BU): The dimension of  $\mathbb{C}_{\mathcal{E}}$  is  $\text{VC}(\mathcal{E})$  and, as the map does not collapse antipodal points,  $2\text{VC}(\mathcal{E}) > \text{sd}(\mathcal{E}) \sim 2\text{VC}(\mathcal{E}) \geq \text{sd}(\mathcal{E}) + 1$ . Again, Theorem 23 is not applicable whenever  $\text{sd} \leq 0$ , but in this case the claimed bound can only be invalidated whenever  $\text{sd} = \text{VC} = 0$ . But  $\text{VC} = 0$  implies the class has only one concept, hence  $\text{sd} = -1$ , and the bound stands. Finally,  $\text{VC}^*(\mathcal{E}) \leq 2\text{VC}(\mathcal{E}) + 1$  follows from it as, by Lemma 2,  $\text{sd} \geq \text{VC}^* - 2$ .

So it only remains to prove the existence of the deformation retraction  $F$  with the prescribed properties in case  $\mathcal{E} \neq \mathcal{C}_n$ .

Let  $W$  be a subset of vertices of  $V(\Delta_{\mathcal{E},1})$  such that  $W$  is upwards-closed with respect to the order on partial hypotheses and contains  $V(\mathbb{C}_{\mathcal{E},1})$ . And let  $\mathbb{Q}_W$  be the full subcomplex of  $\Delta_{\mathcal{E},1}$ , spanned by  $W$ . We are going to prove, by induction on the poset of such sets  $W$ , that there is the corresponding deformation retraction  $F_W$  from  $\|\mathbb{Q}_W\|$  to  $\|\mathcal{E}_1\|$ . The statement of the lemma then corresponds to  $W = V(\Delta_{\mathcal{E},1})$ ; and the base of induction,  $W = V(\mathbb{C}_{\mathcal{E},1})$ , is given by a constant deformation retraction map  $F_{V(\mathbb{C}_{\mathcal{E},1})}: \|\mathbb{Q}_{V(\mathbb{C}_{\mathcal{E},1})}\| \times [0, 1] \rightarrow \|\mathcal{E}_1\|$  defined by  $F(x, t) = x$ , where we use the fact that  $\mathbb{Q}_{V(\mathbb{C}_{\mathcal{E},1})} = \mathbb{C}_{\mathcal{E},1}$ .

For the induction step, let us take such  $W$  and any minimal  $w \in W$  such that  $w \notin V(\mathbb{C}_{\mathcal{E},1})$ . Now assuming, by induction, the existence of  $F_{W'}$ , for  $W' = W - w$ , we want to construct the corresponding  $F_W$ .

For this  $w$ , let us define the *star* and the *link* of  $w$  in  $W$  as the subcomplexes  $\text{St}_W(w) = \{s \in S(\mathbb{Q}_W) \mid \text{there is } t \in S(\mathbb{Q}_W) \text{ s.t. } s, w \leq t\}$  and  $\text{Lk}_W(w) = \{s \in S(\text{St}_W(w)) \mid w \notin s\}$ . Those notions are quite standard, see Bryant (2002). Recall that joins of simplicial complexes and maps between them were defined in Section A.6. We utilize the usual fact that  $\text{St}_W(w)$  is the join of  $w$

and  $\text{Lk}_W(w), \text{St}_W(w) = w * \text{Lk}_W(w)$ , that is,

$$S(\text{St}_W(w)) = \{s_w \sqcup s_l \mid s_w \in \{\emptyset, w\}, s_l \in S(\text{Lk}_W(w))\}.$$

We note that, by construction,  $\text{Lk}_W(w)$  is a subcomplex of  $\mathbb{Q}_{W'}$ , moreover,  $\text{Lk}_W(w) = \mathbb{Q}_{W'} \cap \text{St}_W(w)$  and  $\mathbb{Q}_W = \mathbb{Q}_{W'} \cup \text{St}_W(w)$ . Finally, by the choice of  $w$  as minimal in  $V(W)$ , it easily follows that  $\text{St}_W(w) = \Delta_{\mathcal{E},1}^w$ .

We now define an intermediate deformation retraction  $G$  from  $\|\mathbb{Q}_W\| = \|\mathbb{Q}_{W'}\| \cup w * \|\text{Lk}_W(w)\|$  to  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$  as  $G(t, x) = F_{W'}(t, x)$  for  $x \in \|\mathbb{Q}_{W'}\|$  and  $G(t, x) = (I_w * F_{W'})(t, x)$  for  $x \in w * \|\text{Lk}_W(w)\| = \|\text{St}_W(w)\|$ , where  $I_w$  is the identity deformation retraction of  $w$  to itself. Here  $I_w * F_{W'}$  is the join of the respective retractions. The construction of  $G$  is illustrated in Figure 19 below.

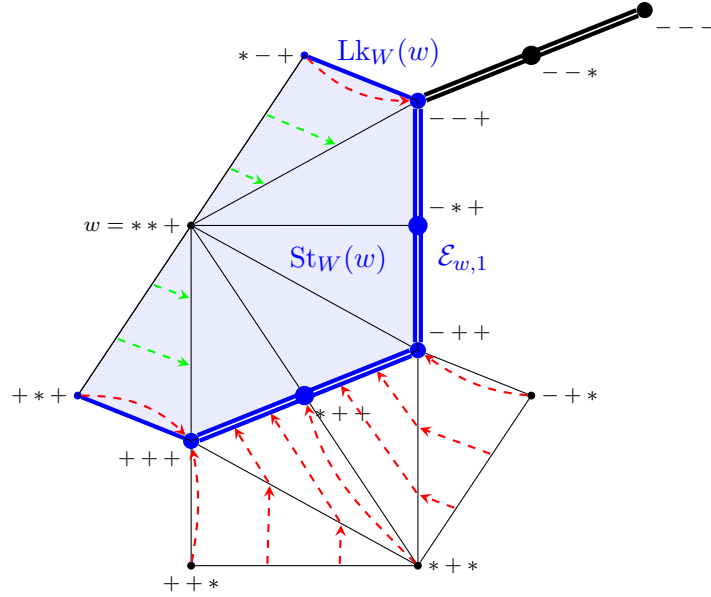


Figure 19: Construction of the deformation retraction  $G$  from  $\|\mathbb{Q}_W\| = \|\mathbb{Q}_{W'}\| \cup w * \|\text{Lk}_W(w)\|$  to  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$ . Here  $\mathcal{E}$  is the extremal class from Figure 15. We assume that, apart from vertices of  $\mathbb{C}_{\mathcal{E}}$ ,  $W'$  also contains vertices  $+ + *$ ,  $- + *$ ,  $* + *$ ,  $+ * +$ , and  $* - +$ , and  $w = * * +$ . Filled blue region represents  $\|\text{St}_W(w)\|$ , the thick blue lines are  $\|\text{Lk}_W(w)\| = \|\text{St}_W(w)\| \cap \|\mathbb{Q}_{W'}\|$ , and double thick blue lines are  $\|\mathcal{E}_{w,1}\| = \|\text{St}_W(w)\| \cap \|\mathcal{E}\|$ . Red dotted lines represent the retraction  $F_{W'}$ , and the green ones are the extension of  $F_{W'}$  to  $I_w * F_{W'}$ .

Let us, however, define  $G$  explicitly:

- For  $x = w$ ,  $G(t, w) \equiv w$ ;
- For  $x \in \|\mathbb{Q}_{W'}\|$ ,  $G(t, x) = F_{W'}(t, x)$ ;
- Finally, for  $x \in \|\text{St}_W(w)\| = w * \|\text{Lk}_W(w)\|$ ,  $x \neq w$ , let  $x = \alpha w + \beta x'$ , for  $x' \in \|\text{Lk}_W(w)\|$ , be the representation of  $x$  in the barycentric coordinates of the join (see Section A.6). Note that  $x \neq w$  implies that  $\beta > 0$  and that  $x'$  is uniquely defined. Then  $G(t, x) = \alpha w + \beta F_{W'}(t, x') \in w * \|\text{Lk}_W(w)\|$ .

Finally, we need to check a few things about  $G$ . First, note that  $G$  is defined separately on  $\|\mathbb{Q}_{W'}\|$  as  $F_{W'}$  and on  $w * \|\text{Lk}_W(w)\|$  as  $(I_w * F_{W'})$ . The fact that this definition is consistent on  $\|\text{Lk}_W(w)\| = \|\mathbb{Q}_{W'}\| \cap w * \|\text{Lk}_W(w)\|$  follows from the construction of  $(I_w * F_{W'})$ . This also implies that  $G$  is continuous, as it is defined in a continuous way on two closed subspaces and is consistent on their intersection. The fact that the range of the final map  $G(1, \cdot)$  is  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$  follows from the inductive statement that  $F_{W'}$  is a deformation retraction to  $\|\mathcal{E}_1\|$  and that  $F_{W'}(t, \|\Delta_{\mathcal{E},1}^w\|) \subseteq \|\Delta_{\mathcal{E},1}^w\|$ , together with the fact that  $\|\mathcal{E}_{w,1}\| = \|\mathcal{E}_1\| \cap \|\Delta_{\mathcal{E},1}^w\|$ . And the fact that  $G$  is a retraction, that is, an identity on  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$ , follows from  $F_{W'}$  being identity on  $\|\mathcal{E}_1\|$  and, consequently, on  $\|\mathcal{E}_{w,1}\|$ .

Let us now define a deformation retraction  $H$  from  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$  to  $\|\mathcal{E}_1\|$  as follows. Let  $C_w$  be the contraction map from  $\|\mathcal{E}_{w,1}\|$  to some  $x_w \in \|\mathcal{E}_{w,1}\|$  from Proposition 41. Recall that this means that  $C_w(0, \cdot)$  is the identity on  $\|\mathcal{E}_{w,1}\|$  and  $C_w(1, \cdot) \equiv x_w \in \|\mathcal{E}_{w,1}\|$ . Now, we define  $H: [0, 1] \times (\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|) \rightarrow (\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|)$  as identity on  $\|\mathcal{E}_1\|$ , and on  $w * \|\mathcal{E}_{w,1}\|$  we define it as

$$H(t, x) = \alpha(1 - t)w + (\beta + \alpha t) C_w(\alpha t / (\beta + \alpha t), x'),$$

where  $x = \alpha w + \beta x'$ , for  $x' \in \|\mathcal{E}_{w,1}\|$ . We note that the term  $\alpha t / (\beta + \alpha t)$  is undefined for  $\alpha = 1$  (alternatively,  $\beta = 0$ ) and  $t = 0$ , in which case we put it to be 1. Note, however, that thus extended, this term becomes discontinuous at  $\alpha = 1, t = 0$ , which is something we will need to deal with later. This is illustrated in Figure 20 below.

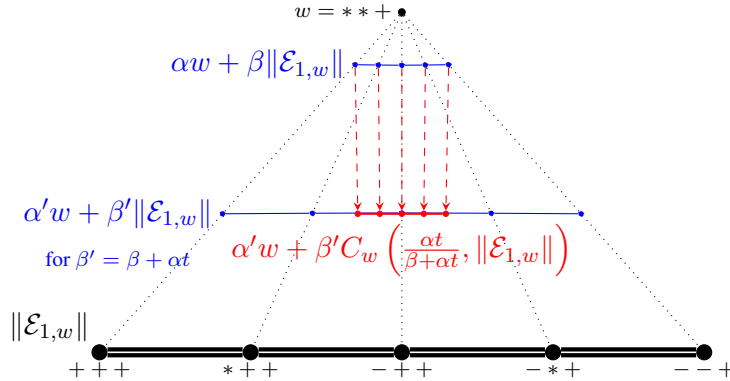


Figure 20: Construction of the deformation retraction  $H$  on  $w * \|\mathcal{E}_{w,1}\|$  for the same parameters as in Figure 19, compare also to Figure 18. The set  $L_\beta = \alpha w + \beta \|\mathcal{E}_{1,w}\|$  is mapped by  $H(t, \cdot)$  into  $L_{\beta'}$  for  $\beta' = \beta + \alpha t$ . Note that  $\text{diam}(L_{\beta'}) = \beta' \text{diam}(\|\mathcal{E}_{1,w}\|) = \beta / \beta' \text{diam}(L_\beta)$  and so, for  $\beta$  small and  $t$  not close to 0,  $\text{diam}(L_{\beta'}) \gg \text{diam}(L_\beta)$ . However, we need to keep the diameter of  $H(t, L_\beta)$  comparable to  $\text{diam}(L_\beta)$ . This motivates the somewhat peculiar choice of parameter  $\alpha t / (\beta + \alpha t)$ ; in particular, taking it, more naively, to be  $\alpha t$  would break condition i) below.

Let us now check the properties of  $H$  on  $w * \|\mathcal{E}_{w,1}\|$ . First of all, note that the barycentric coordinates of the image are  $\alpha(1 - t)$  and  $\beta + \alpha t$ , and their sum is  $\beta + \alpha t + \alpha(1 - t) = \beta + \alpha = 1$ , as needed. We now need to check four things about  $H$ : i) Recall that for  $\alpha = 1, x = w$ , which does not depend on the choice of  $x'$  in the decomposition  $x = 1w + 0x'$ . Thus, for the definition of  $H$  to be consistent, we need to check that  $H(t, x)$  in this case also does not depend on the particular choice of  $x'$ ; ii) For  $\alpha = 0$ , by the definition of retraction,  $H(t, x) = x$ . In particular, this is



consistent with the definition of  $H$  as the identity on  $\|\mathcal{E}_1\|$ , and hence makes  $H$  to be consistently defined on the whole domain; iii) For  $t = 0$ ,  $H(0, x) = x$ ; and iv) For  $t = 1$ ,  $H(1, x) \in \|\mathcal{E}_{w,1}\|$ . Those conditions are summarized in Figure 21 below. Apart from that, we need to separately argue that  $H$  is continuous.

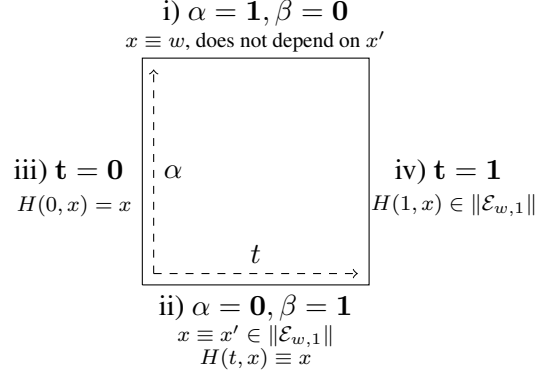


Figure 21: Diagram summarizing the conditions on the intermediate deformation retraction  $H$ .

Let us proceed with checking i)–iv):

i) For  $\alpha = 1$ ,  $H(t, x)$  does not depend on the choice of  $x'$ .

$$H(t, x) = (1 - t)w + tC_w(t/t, x') = (1 - t)w + tC_w(1, x') = (1 - t)w + tc_w,$$

as needed. Note also that changing  $t/t$  into 1 is consistent with our extension of the term  $\alpha t/(\beta + \alpha t)$ ;

ii) For  $\alpha = 0$ ,  $H(t, x) = x$ .

$$H(t, x) = C_w(0, x') = x' = x,$$

as needed.

iii) For  $t = 0$ ,  $H(0, x) = x$ . If  $\alpha \neq 1$  then

$$H(0, x) = \alpha w + \beta C_w(0, x') = \alpha w + \beta x' = x,$$

as needed. Otherwise, if  $\alpha = 1$ , then  $x = w$  and

$$H(0, x) = 1w + 0C_w(1, x') = 1w + 0c_w = w,$$

as needed.

iv) For  $t = 1$ ,  $H(1, x) \in \|\mathcal{E}_{w,1}\|$ .

$$H(1, x) = (\beta + \alpha) C_w(\alpha, x') = C_w(\alpha/(\beta + \alpha), x') = C_w(\alpha, x') \in \|\mathcal{E}_{w,1}\|,$$

as needed.

The only thing left to be checked is that  $H$  is continuous and, trivially, the continuity can only break at the discontinuity of the term  $\alpha t/(\beta + \alpha t)$ , that is, at  $\alpha = 1$  and  $t = 0$ . So let  $x_i \in w * \|\mathcal{E}_{w,1}\|$  and  $t_i \in [0, 1]$  be the sequences converging to  $w$  and 0 respectively. Without losing generality, we

can assume that  $x_i \neq w$  for all  $i$ , which means that each  $x_i$  has a unique representation in barycentric coordinates of the join,  $x_i = \alpha_i w + \beta_i x'_i$ ,  $\alpha_i \neq 1$ , with  $\alpha_i \rightarrow 1$  and  $\beta_i = 1 - \alpha_i \rightarrow 0$ , but with no control over  $x'_i$ . Our goal is to show that  $H(t_i, x_i) \rightarrow H(0, w) = w$ . By the definition of  $H$ ,

$$\begin{aligned} H(t_i, x_i) &= \alpha_i(1 - t_i)w + (\beta_i + \alpha_i t_i) C_w(\alpha_i t_i / (\beta_i + \alpha_i t_i), x'_i) \\ &= \alpha_i(1 - t_i)w + (\beta_i + \alpha_i t_i) x''_i, \end{aligned}$$

for some  $x''_i \in \|\mathcal{E}_{w,1}\|$ . Again, we do not have much control over  $x''_i$ , but, as  $\alpha_i(1 - t_i) \rightarrow 1$  and  $(\beta_i + \alpha_i t_i) \rightarrow 0$ , this obviously converges to  $w$ , as needed.

Finally, the required deformation retraction  $F_W$  is obtained by chaining  $H$  after  $G$ . Formally,  $F_W(t, x) = G(2t, x)$  for  $t \in [0, 1/2]$  and  $F_W(t, x) = H(2(t - 1/2), G(1, x))$  for  $t \in [1/2, 1]$ . As  $G$  is the deformation retraction from  $\|\mathbb{Q}_W\|$  to  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$  and  $H$  is from  $\|\mathcal{E}_1\| \cup w * \|\mathcal{E}_{w,1}\|$  to  $\|\mathcal{E}_1\|$ ,  $F_W$  is a deformation retraction from  $\|\mathbb{Q}_W\|$  to  $\|\mathcal{E}_1\|$ . The only thing left to check is that for any  $x \in \|\mathbb{Q}_W\|$  and  $t \in [0, 1]$ ,  $F_W(t, x) \in \|\Delta_{\mathcal{E},1}^{h(x)}\|$ . This, however, is rather easy: if  $x \in \|\mathbb{Q}_W\|$  is such that  $h(x) \neq w$ , then  $x \in \|\mathbb{Q}_{W'}\|$ , and so, by the induction hypothesis,  $G$ , which coincides with  $F_{W'}$  on  $\|\mathbb{Q}_{W'}\|$ , would keep the image of  $x$  in  $\|\Delta_{\mathcal{E},1}^{h(x)}\|$  and map it there into  $\|\mathcal{E}_1\|$ , where it will stay unaffected by further application of  $H$ . And if  $h(x) = w$  then  $x \in \|\text{St}_W(w)\| = \|\Delta_{\mathcal{E},1}^w\| = w * \|\text{Lk}_W(w)\|$ , where it will first be moved by  $G$  to  $w * \|\mathcal{E}_{w,1}\|$ , and then by  $H$  to  $\|\mathcal{E}_{w,1}\|$ , all without leaving  $\|\Delta_{\mathcal{E},1}^w\|$ .  $\blacksquare$

**Theorem 17** *If  $\text{VC}(\mathcal{H}) \leq 1$  then  $\text{sd}(\mathcal{H}) \leq 1$ . Alternatively, if  $\text{sd}(\mathcal{H}) \geq 2$  then  $\text{VC}(\mathcal{H}) \geq 2$ .*

**Proof** Let  $\mathcal{H}$  be a class with  $\text{VC}(\mathcal{H}) \leq 1$ . It is known (Ben-David (2015)) that in this case  $\mathcal{H}$  can be embedded into an extremal class of  $\text{VC} \leq 1$ , hence  $\text{VC}^{\text{extr}}(\mathcal{H}) \leq 1$ . But then, by Theorem 9,  $\text{sd}(\mathcal{H}) \leq 2\text{VC}^{\text{extr}}(\mathcal{H}) - 1 \leq 1$ , as needed.  $\blacksquare$