Dual VC Dimension Obstructs Sample Compression by Embeddings

Zachary Chase   ZACHMAN99323@GMAIL.COM
Department of Mathematics, Technion

Bogdan Chornomaz   MARKYZ.KARABAS@GMAIL.COM
Department of Mathematics, Technion

Steve Hanneke   STEVE.HANNEKE@GMAIL.COM
Department of Computer Science, Purdue University

Shay Moran   SMORAN@TECHNION.AC.IL
Departments of Mathematics, Computer Science, and Data and Decision Sciences, Technion
Google Research

Amir Yehudayoff   YEHUDAYOFF@TECHNION.AC.IL
Department of Mathematics, Technion
Department of Computer Science, Copenhagen University

Editors: Shipra Agrawal and Aaron Roth

Abstract

This work studies embedding of arbitrary VC classes in well-behaved VC classes, focusing particularly on extremal classes. Our main result expresses an impossibility: such embeddings necessarily require a significant increase in dimension. In particular, we prove that for every d there is a class with VC dimension d that cannot be embedded in any extremal class of VC dimension smaller than exponential in d.

In addition to its independent interest, this result has an important implication in learning theory, as it reveals a fundamental limitation of one of the most extensively studied approaches to tackling the long-standing sample compression conjecture (Warmuth, 2003). Concretely, the approach proposed by Floyd and Warmuth (1995) entails embedding any given VC class into an extremal class of a comparable dimension, and then applying an optimal sample compression scheme for extremal classes. However, our results imply that this strategy would in some cases result in a sample compression scheme at least exponentially larger than what is predicted by the sample compression conjecture.

The above implications follow from a general result we prove: any extremal class with VC dimension d has dual VC dimension at most $2d + 1$. This bound is exponentially smaller than the classical bound $2^{d+1} - 1$ of Assouad (1983), which applies to general concept classes (and is known to be unimprovable for some classes). We in fact prove a stronger result, establishing that $2d + 1$ upper bounds the dual Radon number of extremal classes. This theorem represents an abstraction of the classical Radon theorem for convex sets, extending its applicability to a wider combinatorial framework, without relying on the specifics of Euclidean convexity. The proof utilizes the topological method and is primarily based on variants of the Topological Radon Theorem (Bajmóczy and Bárány, 1979).

Keywords: Sample Compression Schemes, VC Dimension, Maximum Classes, Extremal Classes, The Topological Method, Radon Numbers, Borsuk-Ulam Numbers

1. Introduction and Main Results

A common idea throughout mathematics is the one of transforming an abstract object of interest into a related "nice" object, possessing useful properties. For example, transforming general functions into smooth functions, transforming arbitrary graphs into connected graphs, or transforming geometric bodies into convex sets. In algebraic geometry, a classical example is the resolution of singularities where the goal is to "make varieties smooth". Quantitatively, it is often important to understand how much we need to "add" to an arbitrary object to make it "nice". For instance, we can ask how many edges we need to add to a graph to make it connected, or how much volume we must add to a body to make it convex.

This general idea has also come up in various contexts within learning theory. For example the use of surrogate or regularized losses to accelerate optimization algorithms. Another example, which arises in the context of sample compressions concerns embedding VC classes within *extremal* classes (a.k.a. lopsided, ample, or shattering-extremal; see e.g. Lawrence (1983); Dress (1996); Moran (2012)). Extremal classes enjoy many properties which are useful in the context of learning theory, for example, for defining sample compression schemes of size equal the VC dimension Moran and Warmuth (2016), or proper optimal PAC learners Bousquet et al. (2020). For this reason, there has been much interest in transforming arbitrary concept classes into extremal classes, and understanding how much one needs to add to an arbitrary class to make it extremal. Concretely, determining how much the VC dimension must increase (defined below).

This question has been studied in many works (e.g., Floyd and Warmuth, 1995; Chepoi et al., 2020, 2022; Rubinstein et al., 2015; Rubinstein and Rubinstein, 2022, and numerous references below), but it has remained open whether it is always possible to avoid a significant increase in VC dimension (e.g., a universal constant factor). The question has been of particular interest in the literature on sample compression schemes, since if it were possible to embed arbitrary concept classes into extremal classes without significantly increasing the VC dimension, it would immediately imply a positive resolution of the long-standing *sample compression conjecture* (Warmuth, 2003).

In the present work, we prove that for some concept classes $\mathcal{C}$, it is necessary to add *quite a lot* to make it extremal; that is, any extremal class containing $\mathcal{C}$ must have *exponentially larger* VC dimension. More generally, for any class $\mathcal{C}$, we argue that the increase in VC dimension must be at least proportional to the *Radon number* of version spaces (which is never smaller than the dual VC dimension). In particular, our result implies that the commonly-studied approach aiming to resolve the sample compression conjecture by embedding in extremal classes cannot significantly improve over the size of known sample compression schemes (Moran and Yehudayoff, 2016).

1.1. The Sauer-Shelah-Perles and Pajor Inequalities

The Sauer-Shelah-Perles[1] (SSP) inequality is one of the basic results in the Vapnik-Chervonenkis (VC) theory. It is a key technical lemma in the proof of the fundamental equivalence between finite VC dimension, uniform laws of large numbers, and PAC learnability. The SSP inequality also plays a significant role in geometry, combinatorics, and model theory.

---

1. The result was independently proven by Vapnik and Chervonenkis (1968); Sauer (1972); Shelah (1972), with Vapnik and Chervonenkis (1968) having proven a slightly weaker version. Shelah (1972) gives credit also to Micha Perles. Word has it that, amusingly, Perles proved the result twice, ten years apart, and with different proofs! This unique contribution has humorously led some to refer to it as the Perles-Sauer-Shelah-Perles Lemma.

Definition 1 (VC Dimension) *A concept class $\mathcal{C} \subseteq \{0,1\}^n$ is said to* shatter *a set $\{x_1, \ldots, x_k\} \subseteq [n]$ if*

$$\{(c(x_1), \ldots, c(x_k)) : c \in \mathcal{C}\} = \{0,1\}^k$$

(by convention, the empty set is always shattered). The VC dimension of $\mathcal{C}$, denoted by $\mathtt{vc}(\mathcal{C})$, is the largest size of a set that is shattered by $\mathcal{C}$.

Theorem 2 (Sauer-Shelah-Perles Inequality (Sauer, 1972; Shelah, 1972)) *Let $\mathcal{C} \subseteq \{0,1\}^n$ be a concept class, and let $d = \mathtt{vc}(\mathcal{C})$. Then,*

$$|\mathcal{C}| \le \binom{n}{\le d} := \sum_{i=0}^{d} \binom{n}{i}.$$

The SSP inequality is tight: for every n and d, there exist classes that achieve equality in it. These classes are interesting and have a relatively rigid structure. They also arise naturally in geometrically defined classes such as halfspaces in $\mathbb{R}^n$ (Gärtner and Welzl, 1994).

Definition 3 (Maximum Classes) *A class $\mathcal{C} \subseteq \{0,1\}^n$ is called a* maximum *class if $|\mathcal{C}| = \binom{n}{\le d}$, where $d = \mathtt{vc}(\mathcal{C})$.*

A notable generalization of the SSP inequality was proven by Pajor (1985):

Theorem 4 (Pajor Inequality (Pajor, 1985)) *Let $\mathcal{C} \subseteq \{0,1\}^n$ be a concept class, and let*

$$\mathtt{shatter}(\mathcal{C}) = \{A \subseteq [n] : A \text{ is shattered by } \mathcal{C}\}.$$

Then,

$$|\mathcal{C}| \le |\mathtt{shatter}(\mathcal{C})|.$$

Pajor's inequality indeed implies the SSP inequality, because the size of $A \in \mathtt{shatter}(\mathcal{C})$ is $|A| \le \mathtt{vc}(\mathcal{C})$. Pajor's inequality is also tight. Classes that meet Pajor's inequality with equality are known as lopsided (Lawrence, 1983), ample (Dress, 1996), or shattering-extremal (Moran, 2012). In this text, we use the term *extremal* (Moran, 2012; Moran and Warmuth, 2016).

Definition 5 (Extremal Classes) *A class $\mathcal{C} \subseteq \{0,1\}^n$ is called* extremal *if $|\mathcal{C}| = |\mathtt{shatter}(\mathcal{C})|$.*

While every maximum class is extremal, the converse does not hold.

Extremal classes are notable for their characterization through a range of equivalent definitions, which may initially seem quite distinct. For example, Lawrence (1983) discovered extremal classes in his investigation of sign patterns of convex sets: $\{\mathtt{sign}(v) : v \in K\}$, where $K \subseteq \mathbb{R}^n$ is convex. He defined these classes in a manner that is unrelated to Theorem 4 and does not even rely on the notion of *shattering*. Consequently, these classes have been independently identified in various contexts, including discrete geometry (Lawrence, 1983), functional analysis (Pajor, 1985), extremal combinatorics (Bollobás, Leader, and Radcliffe, 1989; Bollobás and Radcliffe, 1995), and computational biology (Dress, 1996).

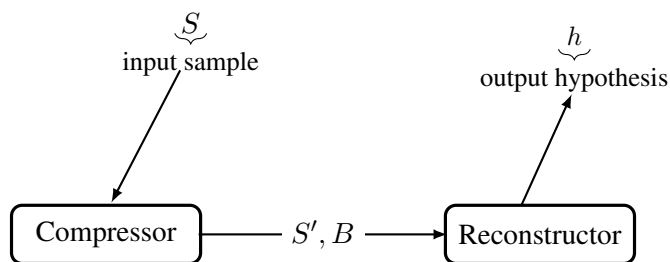**A pictorial definition of a sample compression scheme**

Figure 1: S' is a subsample of S and B is a binary string of additional information.

1.2. Sample Compression Schemes

Sample compression schemes were introduced by Littlestone and Warmuth (1986) in their seminal paper as a tool for proving generalization bounds in learning theory. More generally, this concept can be understood as a mathematical model for data simplification. It is akin to a scientist who collects a large number of experimental observations and then selects a small, representative subset. The scientist's aim is to develop a hypothesis from this subset that can accurately predict the outcomes of the remaining, unselected observations.

A sample compression scheme comprises a *compressor* and a *reconstructor* (see Figure 1). The compressor gets an input sequence of labeled examples S. From this sequence, the compressor selects a subsequence S' and sends it to the reconstructor, along with a binary string B of additional information. The reconstructor then uses S' and B to generate a hypothesis $h = h(S', B)$. The aim is that the hypothesis h correctly classifies the entire input sequence S, including on those examples in S that are not transmitted to the reconstructor.

Definition 6 *A sample compression scheme for a concept class $\mathcal{C}$ is defined by the following criterion. For any input sequence $S = \{(x_i, y_i)\}_{i=1}^m$ that is realizable[2] by $\mathcal{C}$, the hypothesis $h = h(S', B)$ generated by the reconstructor must be consistent with the entire input sample S; i.e. $h(x_i) = y_i$ for all $i \in [m]$. We say that the size of the compression scheme is s if the number of bits in B plus the size of S' is at most s.*

A classical example of a sample compression scheme for the class of d-dimensional linear classifiers is the *Support Vector Machine* algorithm in $\mathbb{R}^d$. Here, the compressor sends to the reconstructor the $d + 1$ support vectors which determine the maximum margin separating hyperplane.

Many well-studied classes admit a sample compression of size equal to their VC dimension. Littlestone and Warmuth (1986) asked about a general connection between the VC dimension and the size of the sample compression scheme. This became one of the most extensively studied and long-standing problems in the theory of learning. Moran and Yehudayoff (2016) affirmatively demonstrated the existence of a sample compression scheme with size exponential in the VC dimension. Yet, it remains an open question whether every class has a compression scheme of size linear or polynomial in its VC dimension, as explicitly asked in Floyd and Warmuth (1995). Warmuth (2003) has offered a $600 reward for solving this problem. For a broader discussion, see e.g. Chapter 17 in Wigderson (2019) or Chapter 30 in Shalev-Shwartz and Ben-David (2014).

---

2. I.e. there exists $c \in \mathcal{C}$ such that $c(x_i) = y_i$ for all $i \leq m$.

Conjecture 7 (Floyd and Warmuth (1995); Warmuth (2003)) *For every concept class $\mathcal{C}$ there exists a sample compression scheme of size $O(\text{vc}(\mathcal{C}))$.*

In their important work, Floyd and Warmuth (1995) layed an approach aimed at resolving the sample compression conjecture. This approach is based on a simple two-step program.

Floyd and Warmuth Program

1. Prove that any maximum class $\mathcal{C}$ with VC dimension d has a sample compression scheme of size $O(d)$.

2. Prove that any class $\mathcal{C}$ with VC dimension d is contained within a maximum class with VC dimension $O(d)$.

The Floyd and Warmuth program has been a prominent method in attempts to resolve Conjecture 7. Several studies have explored this approach, including works by Ben-David and Litman (1998); Warmuth (2003); Kuzmin and Warmuth (2007); Rubinstein, Bartlett, and Rubinstein (2009); Rubinstein and Rubinstein (2012); Rubinstein, Rubinstein, and Bartlett (2015); Moran and Warmuth (2016); Chepoi, Knauer, and Philibert (2020, 2021, 2022); Chalopin, Chepoi, Moran, and Warmuth (2022); Rubinstein and Rubinstein (2022); Chalopin, Chepoi, Inerney, Ratel, and Vaxès (2023).

In their work, Floyd and Warmuth (1995) addressed the first step of their program by demonstrating that every maximum class $\mathcal{C}$ admits a sample compression scheme of size equal to its VC dimension. Following this, a series of subsequent works expanded on the scheme proposed by Floyd and Warmuth (1995). Specifically, Moran and Warmuth (2016) extended it to extremal classes, thereby affirming Conjecture 7 for this category of classes. This extension not only confirmed the conjecture for extremal classes but also simplified Floyd and Warmuth's program. It demonstrated that the second step of their program could be relaxed, showing that it is sufficient to embed any class in an extremal class while maintaining a comparable VC dimension.

For the second step, Floyd and Warmuth (1995) exhibited a class $\mathcal{C}$ that is maximal (not maximum), meaning that any addition of a concept to this class would increase its VC dimension, but is not maximum. This finding implies that not every class $\mathcal{C}$ can be extended to a maximum class without increasing the VC dimension. In the other direction, a recent study by Rubinstein and Rubinstein (2022) proved that any intersection-closed class is contained in an extremal class with a VC dimension at most 11 times larger.

Our first main result places a barrier on Floyd and Warmuth's program, proving that it cannot prove Conjecture 7.

Theorem A (Main Result I) *For every integer $d > 0$, there exists a concept class $\mathcal{C}$ with $\text{vc}(\mathcal{C}) = d$ such that every extremal class $\mathcal{C}'$ such that $\mathcal{C} \subseteq \mathcal{C}'$ has $\text{vc}(\mathcal{C}') \geq 2^d - 1$.*

Theorem A suggests that the best outcome achievable via Floyd and Warmuth's program would be a sample compression scheme of size $2^d - 1$. This would be on par with the best known scheme by Moran and Yehudayoff (2016), which is of size $2^{O(d)}$. However, even reaching this benchmark appears non-trivial, which brings us back to the first paragraph of this work. Can we embed arbitrary classes in "nice" classes?

Open Question 1 *Is there a function $f : \mathbb{N} \to \mathbb{N}$ such that every concept class with VC dimension d is contained in a maximum (extremal) class with VC dimension $f(d)$?*

A special case of maximum classes is hyperplane arrangements (see Example 5). The smallest dimension of a hyperplane arrangement that contains a class $\mathcal{C}$ is equivalent to the sign rank of the matrix $(M_{c,x})$ defined by $M_{c,x} = c(x)$. Alon, Moran, and Yehudayoff (2017) proved that if we replace "maximum class" by "hyperplane arrangment" in the question, then the answer is negative. For every m, there is a class of VC dimension two that cannot be embedded in an m-dimensional hyperplane arrangement.

1.3. Dual VC Dimension

We establish Theorem A by distinguishing extremal classes from general classes with the same VC dimension, focusing on the interplay between VC and dual VC dimensions.

Definition 8 (Dual Class and VC dimension) *Let $\mathcal{C} \subseteq \{0,1\}^n$ be a class. Each $x \in [n]$ defines a function $f_x : \mathcal{C} \to \{0,1\}$ by $f_x(c) = c(x)$. The dual class of $\mathcal{C}$ is $\mathcal{C}^{\star} = \{f_x : x \in [n]\}$. The dual VC dimension of $\mathcal{C}$ is $\mathtt{vc}^{\star}(\mathcal{C}) = \mathtt{vc}(\mathcal{C}^{\star})$.*

The VC and dual VC dimensions are exponentially related:

Theorem 9 (Assouad (1983)) *Every concept class $\mathcal{C}$ satisfies*

$$\lfloor \log \mathtt{vc}(\mathcal{C}) \rfloor \leq \mathtt{vc}^{\star}(\mathcal{C}) \leq 2^{\mathtt{vc}(\mathcal{C})+1} - 1.$$

Both inequalities are sharp (in some cases). While the relationship between VC and dual VC dimensions is fully understood for general classes, our result provides an exponential improvement for extremal classes (in the right inequality).

> **Theorem B (Main Result II)** *Every extremal concept class $\mathcal{C}$ satisfies*
>
> $$\lfloor \log \mathtt{vc}(\mathcal{C}) \rfloor \leq \mathtt{vc}^{\star}(\mathcal{C}) \leq 2\mathtt{vc}(\mathcal{C}) + 1.$$

Theorem A is a corollary of Theorem B. Indeed, choose a concept class $\mathcal{C}$, with $\mathtt{vc}(\mathcal{C}) = d$ and $\mathtt{vc}^{\star}(\mathcal{C}) = 2^{d+1} - 1$. This class can be chosen as a dual of a class $\mathcal{C}_1$, witnessing the sharpness of the right bound in Theorem 9, that is, of $\mathcal{C}_1$ for which $\mathtt{vc}^{\star}(\mathcal{C}_1) = d$ and $\mathtt{vc}(\mathcal{C}_1) = 2^{d+1} - 1$. Let $\mathcal{C}'$ be an extremal class such that $\mathcal{C}' \supseteq \mathcal{C}$. Theorem B implies that

$$2^{d+1} - 1 = \mathtt{vc}^{\star}(\mathcal{C}) \leq \mathtt{vc}^{\star}(\mathcal{C}') \leq 2\mathtt{vc}(\mathcal{C}') + 1.$$

The left bound in Theorem B follows from Theorem 9; it is tight for all VC dimensions as demonstrated by cubes (see Example 2 in Appendix A). We leave it as an open question whether the upper bound in Theorem B is tight for all VC dimensions. In Appendix A we demonstrate that it is tight for $d = 1$ (Example 4) and notice that hyperplane arrangements yield extremal classes (in fact, maximum classes) with $\mathtt{vc}(\mathcal{C}) = d$ and $\mathtt{vc}^{\star}(\mathcal{C}) = d + 1$ (Example 5).

Open Question 2 *Is it true that for every $d \in \mathbb{N}$ there exists an extremal class $\mathcal{C}$ with $\mathtt{vc}(\mathcal{C}) = d$ and $\mathtt{vc}^{\star}(\mathcal{C}) = 2d + 1$? Is there such a maximum class?*

1.4. Abstract Convexity

To state our most general result we use the language of abstract convexity theory. An (abstract) convexity space offers a simplified yet profound abstraction of Euclidean convexity. This concept originated in the work of Levi (1951) and was later given its current formalization by Kay and Womble (1971). For a detailed introduction to this subject, readers may refer to the survey by Danzer, Grünbaum, and Klee (1963), or the more recent book van de Vel (1993).

A *convexity space* is a pair $(\Omega, \mathcal{F})$ where $\mathcal{F} \subseteq 2^{\Omega}$ is a family of subsets that satisfies the following conditions:

- $\emptyset, \Omega \in \mathcal{F}$.

- $\mathcal{F}$ is closed under intersections.

Sets within $\mathcal{F}$ are termed *convex sets*. Convexity spaces are prevalent in mathematics, manifesting in forms such as closed sets in topological spaces, subgroups of groups, subtrees of graphs, and more. We note that in van de Vel (1993), the convexity space is also required to be closed under nested unions. However, all the convexity spaces that we use in this paper can be assumed to be finite, in which case this requirement becomes redundant.

Many basic concepts associated with convexity are still definable in this abstract setting. For instance, a convex set $F \in \mathcal{F}$ is termed a *half-space* if its complement is also convex.[3] The *convex hull* of a set $P \subseteq \Omega$, denoted by $\text{conv}(P) = \text{conv}_{\mathcal{F}}(P)$, is the intersection of all convex sets $F \in \mathcal{F}$ that contain P. These definitions enable the abstraction of classical results, such as Radon's Theorem (Radon, 1921), Helly's Theorem (Helly, 1923), and the Weak Epsilon-Net Theorem (Bárány, Füredi, and Lovász, 1990; Alon, Bárány, Füredi, and Kleitman, 1992; Moran and Yehudayoff, 2020; Holmsen and Lee, 2021).

Definition 10 (Radon Numbers) *Let $(\Omega, \mathcal{F})$ be a convexity space. We say that $p_1, \ldots, p_n \in \Omega$ are Radon-independent if $\text{conv}(\{p_i : i \in I\}) \cap \text{conv}(\{p_j : j \in J\}) = \emptyset$ for every non-trivial partition $I \cup J = [n]$. The Radon number of $\mathcal{F}$ is the largest size of a Radon-independent set.*

The classical Radon Theorem asserts that the Radon number of the convex sets in $\mathbb{R}^d$ is equal to $d + 1$. We now explain how to associate a convexity space to every concept class. These convexity spaces were implicitly studied in learning theory (more details follow). We note that there are also other ways to associate a convexity space to a concept class (for example, version spaces[4] are convex).

Definition 11 (Convexity Space of a Class) *The convexity space of a class $\mathcal{C} \subseteq \{0,1\}^n$ is defined as follows. The domain is $\Omega = \mathcal{C}$. The half-spaces are sets of the form*

$$\mathcal{C}_{x,y} = \{c \in \mathcal{C} : c(x) = y\}$$

where $x \in [n]$ and $y \in \{0,1\}$. The convex sets in $\mathcal{F} = \mathcal{F}_{\mathcal{C}}$ are arbitrary intersections of these half-spaces.

---

3. In the standard Euclidean context, not all half-spaces are open or closed.
4. Given a concept class $\mathcal{C}$ and a sequence of labeled examples $\{(x_i, y_i)\}_{i=1}^m$, the associated version space is $\{c \in \mathcal{C} : c(x_i) = y_i \text{ for all } i\}$.

For $X \subseteq [n]$ and $t \colon X \to \{0, 1\}$, let us define

$$\mathcal{C}_{X,t} = \{c \in \mathcal{C} : c(x) = t(x) \text{ for } x \in X\}.$$

Because $\mathcal{C}_{X,t} = \bigcap_{x \in X} \mathcal{C}_{x,t(x)}$, the set $\mathcal{C}_{X,t}$ is convex. The other way around, every non-empty convex set has this form.

Definition 12 (Radon Number of a Class) *The* Radon number of a concept class $\mathcal{C}$, *denoted* $\mathtt{r}(\mathcal{C})$, *is the Radon number of the associated convexity space $\mathcal{F}_{\mathcal{C}}$.*

We relate the dual Radon number to the VC dimension of extremal classes.

Theorem C (Main Result III) *Let $\mathcal{C} \subseteq \{0,1\}^n$ be an extremal concept class. Then,*

$$\mathtt{r}(\mathcal{C}) \leq 2\mathtt{vc}(\mathcal{C}) + 1.$$

Moreover, the following lower bounds on $\mathtt{r}(\mathcal{C})$ hold

$$\lfloor \log(2\mathtt{vc}(\mathcal{C}) + 2) \rfloor \leq \mathtt{r}(\mathcal{C}) \qquad \textit{for any class } \mathcal{C};$$
$$\mathtt{vc}(\mathcal{C}) + 1 \leq \mathtt{r}(\mathcal{C}) \qquad \textit{for any maximum class } \mathcal{C} \neq \{0,1\}^n;$$
$$\lfloor \log(2\mathtt{vc}(\mathcal{C}) + 2) \rfloor = \mathtt{r}(\mathcal{C}) \qquad \textit{for } \mathcal{C} = \{0,1\}^n.$$

The right bound in Theorem B follows from the first bound in Theorem C, because every $c_1, \ldots, c_k \in \mathcal{C}$ that is shattered by $\mathcal{C}^\star$ is Radon independent; in particular, $\mathtt{vc}^*(\mathcal{C}) \leq \mathtt{r}(\mathcal{C})$. Indeed if $I \cup J = [k]$ is a non-trivial partition then there exists x such that $c_i(x) = 1$ for all $i \in I$ and $c_j(x) = 0$ for all $j \in J$. The complementing pair of half-spaces $\mathcal{C}_{x,1}$ and $\mathcal{C}_{x,0}$ separate $\{c_i : i \in I\}$ and $\{c_j : j \in J\}$ and hence their convex hulls are disjoint.

The upper bound on $\mathtt{r}(\mathcal{C})$ in Theorem C is the main technical result of the paper and is proven in Section 2. Example 5 in Appendix A shows that it is also tight up to a factor of two. The two lower bounds on $\mathtt{r}(\mathcal{C})$ are proven in Propositions 21 and 23 in Appendix B. Additionally, Examples 2 (cube) and 3 (dented cube) in Appendix A show that these lower bounds are tight; the proofs that the values in these examples are as claimed are in Propositions 22 and 25 in Appendix B. These lower bounds demonstrate that the dual Radon numbers can abruptly drop when adding a single concept. Determining the tightness of the upper bound is left as an open question for future research.

Open Question 3 *Is it the case that for every $d \in \mathbb{N}$ there exists an extremal class $\mathcal{C}$ with $\mathtt{vc}(\mathcal{C}) = d$ and $\mathtt{r}(\mathcal{C}) = 2d + 1$? Is there such a maximum class?*

Our final result summarizes the relationship between the dual Radon number and dual VC dimension. This result, while less central, completes the triangular relationship between VC, dual VC, and dual Radon.

Theorem D

1. *Every concept class $\mathcal{C}$ satisfies $\mathtt{vc}^\star(\mathcal{C}) \le \mathtt{r}(\mathcal{C})$.*

2. *For general classes, $\mathtt{r}(\mathcal{C})$ is not upper-bounded by $\mathtt{vc}(\mathcal{C})$ or $\mathtt{vc}^*(\mathcal{C})$.*

3. *If $\mathcal{C}$ is extremal, then $\mathtt{r}(\mathcal{C}) \le 2^{\mathtt{vc}^\star(\mathcal{C})+2} - 1$, and this bound is tight up to a multiplicative factor.*

Theorem D is proven in Proposition 27 in Appendix B. The fact that the dual Radon number of a class is not upper-bounded by its VC and dual VC dimensions is demonstrated by the class of singletons (Example 1 in Appendix A). Example 3 (dented cube) in Appendix A demonstrates that for extremal classes, the dual Radon number can indeed be exponential with respect to the dual VC dimension.

Organization. The rest of this manuscript is organized as follows. Section 2 provides a proof of our main result, that is, $\mathtt{r}(\mathcal{C}) \le 2\mathtt{vc}(\mathcal{C}) + 1$ for extremal class $\mathcal{C}$; it also introduces the relevant topological notions and theorems. The remaining sections are moved to the appendices. In Appendix A we present several, rather simple, examples that demonstrate the sharpness of our bounds. The supplementary proofs, all of which are mostly elementary, are provided in Appendix B; this includes showing that the values of the relevant parameters in examples are indeed what they are claimed to be. Finally, in Appendix C we give a broader discussion on the topological notions introduced in Section 2.

2. Main Proof

We now present the proof for the upper bound in Theorem C. As previously discussed, Theorem A and Theorem B are derived as corollaries. Consider an extremal class $\mathcal{C} \subseteq \{0,1\}^n$; our goal is to demonstrate that $\mathtt{r}(\mathcal{C}) \le 2\mathtt{vc}(\mathcal{C}) + 1$. The proof employs a topological approach. For this purpose, we define a cube complex (see Figure 2). We think of the cube complex as a topological subspace of $[0,1]^n$. It comprises cubes of various dimensions that are glued together according to the structure of $\mathcal{C}$.

Definition 13 (The Cube Complex) *The* cube complex $Q = Q(\mathcal{C}) \subseteq [0,1]^n$ *of* $\mathcal{C} \subseteq \{0,1\}^n$ *is defined as follows. The vertices of Q are the concepts of $\mathcal{C}$. The cubes of Q are the "filled cubes of $\mathcal{C}$".*

Formally, a d-cube in $\mathcal{C}$ is a pair (Y, f), where $Y \subseteq [n]$ is of size $|Y| = d$ and f is a function $f \colon [n] \setminus Y \to \{0,1\}$ such that for every $g \colon Y \to \{0,1\}$ there is a concept $c \in \mathcal{C}$ such that $c(x) = g(x)$ for every $x \in Y$ and $c(x) = f(x)$ for every $x \in [n] \setminus Y$.

For a cube (Y, f) of $\mathcal{C}$, we define the corresponding solid cube $(Y, f)_Q \subseteq [0,1]^n$ *as*

$$(Y, f)_Q = \{v \in [0,1]^n : v(x) = f(x) \text{ for } x \in Y\}.$$

We now define the cubes of Q as the collection $(Y, f)_Q$, over all cubes (Y, f) of $\mathcal{C}$.

Let us note several basic facts about the cube complex $Q(\mathcal{C})$ that we are going to utilize:

- For a d-cube (Y, f) of $\mathcal{C}$, $d = |Y|$ is called the *dimension* of this cube, and it coincides with the topological dimension of $(Y, f)_Q$. We define the *dimension* of $Q = Q(\mathcal{C})$, denoted $\dim(Q)$, as the maximum dimension of its cube;

- 0-cubes of V, also called *vertices* of V, are precisely the concepts of $\mathcal{C}$. Here we assume that a concept $c \in \mathcal{C} \subseteq \{0, 1\}^n$ is naturally identified with the point in $[0, 1]^n$ with the same coordinates. Moreover, every point of $Q(\mathcal{C})$ with all coordinates in $\{0, 1\}$ corresponds to a concept of $\mathcal{C}$. This essentially means that $\mathcal{C}$ is "embedded" into $Q(\mathcal{C})$;

- We say that (Y_1, f_1) is a *subcube* of (Y_2, f_2) if $Y_1 \subseteq Y_2$ and f_2 is a restriction of f_1 to $[n] - Y_2$. Then $(Y_1, f_1)_Q \subseteq (Y_2, f_2)_Q$ if and only if (Y_1, f_1) is a subcube of (Y_2, f_2). In this case we also say that $(Y_1, f_1)_Q$ is a subcube of $(Y_2, f_2)_Q$;

- Any subcube of a cube $(Y, f)_Q$ of $Q(\mathcal{C})$ is a cube of $Q(\mathcal{C})$. Also, any cubes (Y_1, f_1) and (Y_2, f_2) of $Q(\mathcal{C})$ either disjoint, or intersect by a cube of $Q(\mathcal{C})$. This property enables us to indeed call $Q(\mathcal{C})$ a cube *complex*, as opposed to it being just a collection of cubes.
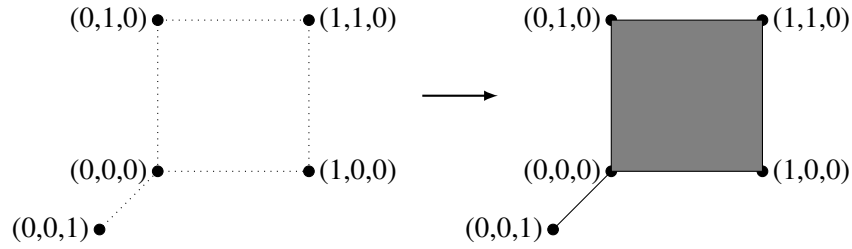

Figure 2: A 2-dimensional illustration of the cube complex for the extremal class $\mathcal{C} = \{000, 010, 110, 100, 001\}$. It has 5 vertices (0-dimensional cubes), 5 edges (1-dimensional cubes), and 1 square (2-dimensional cube). The square corresponds to a cube (Y, f) of $\mathcal{C}$ with $Y = \{1, 2\}$ and $f: 3 \mapsto 0$. A unique maximal edge, connecting $(0, 0, 0)$ and $(0, 0, 1)$, corresponds to a cube with $Y = \{3\}$ and $f: 1 \mapsto 0, 2 \mapsto 0$.

For $Y \subseteq [n]$, we say that Y is *strongly shattered* by $\mathcal{C}$ if there is a cube (Y, f) in $\mathcal{C}$. It is well-known that for extremal classes strong shattering is equivalent to shattering; see, for example, Bollobás and Radcliffe (1995).[5] It then follows that for an extremal $\mathcal{C}$, $\dim(Q) = \texttt{vc}(\mathcal{C})$.

A key concept that we utilize in the proof is the *Radon number* of a topological space. Denote by Δ^d the standard d-dimensional simplex (with $d + 1$ vertices), and by $\partial\Delta^d$ its boundary, which is homeomorphic to the $(d - 1)$-dimensional sphere $S^{d-1} \subset \mathbb{R}^d$.

Definition 14 (Radon number of a topological space) *For a topological space X, the* Radon number $\texttt{r}(X)$ *is the largest integer d such that there is a continuous mapping $f: \partial\Delta^{d+1} \to X$ such that the images of disjoint faces are disjoint. The map f is called a* Radon map.

Alternatively, $\texttt{r}(X) + 1$ is the minimal integer d such that for every continuous map $f: \partial\Delta^{d+1} \to X$, there is a pair of disjoint faces whose images intersect. The name "Radon number" is motivated by the famous Topological Radon Theorem.

---

5. The authors use the terms *strongly traced* and *traced*.

Theorem 15 (Topological Radon Theorem (Bajmóczy and Bárány, 1979))

$$\mathbf{r}(\mathbb{R}^d) = d - 1.$$

Our proof consists of two main steps, whose statements are now easy to formulate. First, we prove that for an extremal class $\mathcal{C}$ the "topological" Radon number $\mathbf{r}(Q(\mathcal{C}))$ can be lower-bounded by the "discrete" Radon number $\mathbf{r}(\mathcal{C})$.

Proposition 16 *Let $\mathcal{C}$ be an extremal class, and let $c_1, \ldots, c_k \in \mathcal{C}$ be Radon independent. Then, there exists a Radon map $f \colon \partial\Delta^{k-1} \to Q(\mathcal{C})$. In particular, $\mathbf{r}(Q(\mathcal{C})) \geq \mathbf{r}(\mathcal{C}) - 2$.*

The second step upper bounds the Radon number of an arbitrary simplicial complex in terms of its dimension.

Lemma 17 *For a finite d-dimensional simplicial complex K, we have $d - 1 \leq \mathbf{r}(K) \leq 2d - 1$.*

The slightly weaker upper bound $\mathbf{r}(K) \leq 2d$ can be derived as a consequence of the Topological Radon Theorem. Indeed, by the geometric realization theorem, K can be embedded in $\mathbb{R}^{2d+1}$. Thus, any Radon map into K yields a Radon map into $\mathbb{R}^{2d+1}$, and so $\mathbf{r}(K) \leq \mathbf{r}(\mathbb{R}^{2d+1}) = 2d$.

The required bound can be obtained by combining Proposition 16 and Lemma 17 as follows. Let $Q^t(\mathcal{C})$ be a *triangulation* of Q. That is, Q^t is a finite simplicial complex such a) every simplex of Q^t is contained in a cube of Q of the same dimension, and b) every cube of Q is a finite union of simplices of Q^t of the same dimension. In particular, $\dim(Q) = \dim(Q^t)$ and Q and Q^t are homeomorphic as topological spaces, which implies $\mathbf{r}(Q) = \mathbf{r}(Q^t)$. It is well-known and easy to check that some triangulation Q^t of Q exists, see for example Lee and Santos (2017). Then

$$\begin{aligned} \mathbf{r}(\mathcal{C}) &\leq \mathbf{r}(Q(\mathcal{C})) + 2 = \mathbf{r}(Q^t(\mathcal{C})) + 2 \\ &\leq 2\dim(Q^t(\mathcal{C})) - 1 + 2 = 2\dim(Q(\mathcal{C})) - 1 + 2 = 2\mathbf{vc}(\mathcal{C}) + 1. \end{aligned}$$

To complete the proof, we thus need to prove Proposition 16 and Lemma 17. For the proof of Proposition 16, let us first the following property of $Q(\mathcal{C})$. Recall that a topological space Y is k-connected if for every $0 \leq \ell \leq k$, any continuous map from an ℓ-dimensional sphere S^ℓ into Y can be extended to a continuous map from the $(\ell+1)$-dimensional ball $B^{\ell+1}$ into Y. For $k = 1$, for example, this means that every copy of a circle in Y can be extended to a copy of a disc in Y. Figuratively, we can freely deflate balloons inside Y.

Proposition 18 *Let $\mathcal{C}$ be an extremal class, and let $Y \subseteq \mathcal{C}$ be an arbitrary non-empty convex, with respect to the convexity space $\mathcal{F}_\mathcal{C}$, subset of $\mathcal{C}$. Then $Q(Y) \subseteq Q(\mathcal{C})$ is k-connected for every $k \geq 0$.*

Proof The proof relies on the following two results from Chalopin, Chepoi, Moran, and Warmuth (2022) that emphasize the highly connected nature of extremal classes (the authors use the term *ample* for extremal):

- If $\mathcal{C}$ is extremal then every $\mathcal{C}_{X,t}$ is extremal.

- If $\mathcal{C}$ is extremal then the cubical complex $Q(\mathcal{C})$ is *collapsible*.

In Chalopin et al. (2022), the first bullet is Theorem 3.1 (6), but it is also a well-known property of extremal classes. The second is their Proposition 4.12, and, to our knowledge, it was not known before that. We refer the reader to the referenced paper for the definition of collapsibility; here we only use the standard fact[6] that collapsible spaces are k-connected for all k.

Now let us take a convex set $Y \subseteq \mathcal{C}$. By Definition 11 and the discussion afterwards, $Y = \mathcal{C}_{X,t}$ for some $X \subseteq [n]$ and $t: X \to \{0,1\}$. Then by the first item above, Y itself is extremal. And by the second item, $Q(Y)$ is collapsible, and hence k-connected for all k. ∎

Proof (Of Proposition 16). Let $c_1, \ldots, c_k \in \mathcal{C}$ be Radon independent, and let $\Delta = \Delta^{k-1}$. Recall that the abstract simplices of $\partial\Delta$ are precisely non-empty subsets of $[k]$. We are going to define $f: \partial\Delta \to Q = Q(\mathcal{C})$ such that for every simplex s with vertex set $V(s)$ in $\partial\Delta^{k-1}$,

$$f(s) \subseteq \mathtt{conv}_Q(C(s)),$$

where $C(s) = \{c_i : i \in V(s)\}$ and $\mathtt{conv}_Q(W) = Q(\mathtt{conv}(W)) \subseteq Q(\mathcal{C})$, for $W \subseteq \mathcal{C}$.

The map f is defined inductively. The ℓ-skeleton of $\partial\Delta$ comprises simplices of dimension at most ℓ. We construct f first on the 0-skeleton, then on the 1-skeleton, and so forth. For $\ell = 0$, we need to define f on $V(\partial\Delta) = [k]$. Naturally, we do it by putting $f(i) = c_i$, for $i \in [k]$. Note that here we use the fact that $\mathcal{C}$ is embedded into $Q(\mathcal{C})$, that is, each $c_i \in \mathcal{C} \subseteq \{0,1\}^n$ is treated as a point in $Q(\mathcal{C}) \subseteq [0,1]^n$. Then trivially, for every $i \in [k]$, $f(i) \in \mathtt{conv}_Q(c_i)$, as needed.

Now, assuming that f is defined on the ℓ-skeleton and we need to define it on the $(\ell + 1)$-skeleton. By the induction hypothesis, for every maximal $(\ell + 1)$-dimensional simplex s, the map f is already defined on its ℓ-skeleton ∂s. Moreover, for each ℓ-dimensional simplex s' in s,

$$f(s') \subseteq \mathtt{conv}_Q(C(s')) \subseteq \mathtt{conv}_Q(C(s)).$$

As $s \cong B^{\ell+1}$ and $\partial s \cong S^\ell$, applying Proposition 18 to s with $Y = \mathtt{conv}(C(s))$ gives an extension of f to s with $f(s) \subseteq \mathtt{conv}_Q(C(s))$.

This concludes the construction of the map f. Let us take any disjoint simplices s_1 and s_2 of $\partial\Delta^{k-1}$. In particular, $V(s_1)$ and $V(s_2)$ are disjoint. As $c_1, \ldots, c_k$ are Radon independent, the convex sets $\mathtt{conv}(C(s_1))$ and $\mathtt{conv}(C(s_2))$ are disjoint.

We claim that then $Q_1 = \mathtt{conv}_Q(C(s_1))$ and $Q_2 = \mathtt{conv}_Q(C(s_2))$ are also disjoint. Indeed, both Q_1 and Q_2 are cube subcomplexes of $Q(\mathcal{C})$, and hence their intersection $Q_1 \cap Q_2$ is also a cube subcomplex of $Q(\mathcal{C})$. Hence, if $Q_1 \cap Q_2$ is nonempty, it contains a 0-dimensional cube, that is, a vertex, p. But vertices of a cube complex of a class are in bijective correspondence with the elements of this class, that is, $p \in Q_1 = Q(\mathtt{conv}(C(s_1)))$ if and only if $p \in \mathtt{conv}(C(s_1))$, and similarly, $p \in \mathtt{conv}(C(s_1))$, a contradiction.

Thus, the images $f(s_1)$ and $f(s_2)$ are disjoint for any disjoint simplices s_1 and s_2 of $\partial\Delta^{k-1}$, and hence f is Radon, as needed. ∎

Finally, let us prove Lemma 17. Here we rely on two known results. The first is proved by Guilbault (2010) in his simple proof of the Topological Radon Theorem using the Borsuk-Ulam Theorem.

---

6. Chalopin et al. (2022), e.g., say that collapsibility is a stronger version of contractibility. It is well-known that a contraction of Y enables to extend any continuous map $S^\ell \cong \partial B^{\ell+1} \to Y$ to a continuous map $B^{\ell+1} \to Y$. Roughly speaking, the extension is constructed by emulating inside Y the contraction of S^ℓ to a point in $B^{\ell+1}$ using the contraction of Y.

Proposition 19 (Guilbault (2010), Proposition 3.1) *For every $k \geq 0$, there is a continuous function $\lambda_k \colon S^k \to \partial\Delta^{k+1}$ such that images of the antipodal points lie in disjoint simplices.*

The second result is a variant of the Borsuk-Ulan Theorem for general simplicial complexes that was proven by Gonçalves, Jaworowski, and Pergher (2002).

Theorem 20 (Gonçalves et al. (2002)) *Let K be a finite simplicial complex of dimension d. If $f \colon S^k \to K^d$ is continuous and $2d \leq k$, then there exists $x \in S^k$ so that $f(x) = f(-x)$.*

Proof (Of Lemma 17). The lower bound $d - 1 \leq \mathtt{r}(K)$ holds because K contains a d-dimensional simplex. For the upper bound $\mathtt{r}(K) \leq 2d - 1$, we use the two results above. Given a Radon map $f \colon \partial\Delta^{k+1} \to K$, we can use the map λ_k to get a *Borsuk-Ulam map* $g = f \circ \lambda_k$. That is, the map $g \colon S^k \to \mathbb{R}^d$ does not collapse antipodal points. By Theorem 20, we can deduce that $k \leq 2d - 1$. ∎

Acknowledgments

References

Noga Alon, Imre Bárány, Zoltán Füredi, and Daniel J Kleitman. Point selections and weak ε-nets for convex hulls. *Combinatorics, Probability and Computing*, 1(03):189–200, 1992.

Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus Vapnik-Chervonenkis dimension. *Sbornik: Mathematics*, 208(12):1724, 2017.

Patrick Assouad. Densité et dimension. *Annales de l'Institut Fourier (Grenoble)*, 33(3):233–282, 1983.

Ervin G. Bajmóczy and Imre Bárány. On a common generalization of Borsuk's and Radon's theorem. *Acta Mathematica Academiae Scientiarum Hungarica*, 34(3):347–350, 1979. doi: 10.1007/BF01896131. URL https://doi.org/10.1007/BF01896131.

Imre Bárány, Zoltán Füredi, and László Lovász. On the number of halving planes. *Combinatorica*, 10(2):175–183, 1990.

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.

Béla Bollobás and Andrew J. Radcliffe. Defect Sauer results. *Journal of Combinatorial Theory, Series A*, 72(2):189–208, 1995.

Béla Bollobás, Imre Leader, and Andrew 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.

Olivier Bousquet, Steve Hanneke, Shay Moran, and Nikita Zhivotovskiy. Proper learning, Helly number, and an optimal SVM bound. In *Proceedings of the 33rd Conference on Learning Theory*, 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.

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. What convex geometries tell about shattering-extremal systems. *The Electronic Journal of Combinatorics*, pages P3–40, 2022.

Ludvig Danzer, Branko Grünbaum, and Victor Klee. *Helly's Theorem and Its Relatives*. Proceedings of symposia in pure mathematics: Convexity. American Mathematical Society, 1963. URL https://books.google.com/books?id=I1l5HAAACAAJ.

Andreas W. M. Dress. Towards a theory of holistic clustering. In Boris G. Mirkin, Fred R. McMorris, Fred S. Roberts, and Andrey Rzhetsky, editors, *Mathematical Hierarchies and Biology, Proceedings of a DIMACS Workshop, November 13-15, 1996*, volume 37 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 271–290. DIMACS/AMS, 1996. doi: 10.1090/DIMACS/037/19. URL https://doi.org/10.1090/dimacs/037/19.

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

Bernd Gärtner and Emo Welzl. Vapnik-Chervonenkis dimension and (pseudo-) hyperplane arrangements. *Discrete Comput. Geom.*, 12(4):399–432, 1994. doi: 10.1007/BF02574389.

Daciberg Gonçalves, Jan Jaworowski, and Pedro Pergher. *G*-coincidences for maps of homotopy spheres into CW-complexes. *Proceedings of the American Mathematical Society*, 130(10):3111–3115, 2002.

Craig R. Guilbault. An elementary deduction of the Topological Radon Theorem from Borsuk–Ulam. *Discrete & Computational Geometry*, 43:951–954, 2010.

Eduard Helly. Über mengen konvexer körper mit gemeinschaftlichen punkte. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 32:175–176, 1923. URL http://eudml.org/doc/145659.

Andreas F Holmsen and Donggyu Lee. Radon numbers and the fractional Helly theorem. *Israel Journal of Mathematics*, 241(1):433–447, 2021.

Jan Jaworowski and Marek Izydorek. Antipodal coincidence for maps of spheres into complexes. In *preprint (Lecture at the Conference on Topological Fixed Point Theory and its Applications in Torun*, 1993.

David Kay and Eugene W Womble. Axiomatic convexity theory and relationships between the Carathéodory, Helly, and Radon numbers. *Pacific Journal of Mathematics*, 38(2):471–485, 1971.

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.

James F. Lawrence. Lopsided sets and orthant-intersection of convex sets. *Pacific J. Math.*, 104:155–173, 1983.

Carl W Lee and Francisco Santos. Subdivisions and triangulations of polytopes. In *Handbook of discrete and computational geometry*, pages 415–447. Chapman and Hall/CRC, 2017.

Friedrich W. Levi. On Helly's theorem and the axioms of convexity. *J. Indian Math. Soc*, 15:65–76, 1951.

Nick Littlestone and Manfred K. Warmuth. Relating data compression and learnability. *Unpublished manuscript*, 1986.

Jiří Matoušek, Anders Björner, Günter M Ziegler, et al. *Using the Borsuk-Ulam theorem: lectures on topological methods in combinatorics and geometry*, volume 2003. Springer, 2003.

Shay Moran. Shattering-extremal systems. *arXiv preprint*, 1211.2980, 2012.

Shay Moran and Manfred Warmuth. Labeled compression schemes for extremal classes. In *Proceedings of the* 27th *International Conference on Algorithmic Learning Theory*, pages 34–49. Springer, 2016.

Shay Moran and Amir Yehudayoff. Sample compression schemes for VC classes. *Journal of the ACM*, 63(3):1–10, 2016.

Shay Moran and Amir Yehudayoff. On weak ϵ-nets and the Radon number. *Discret. Comput. Geom.*, 64(4):1125–1140, 2020. doi: 10.1007/S00454-020-00222-Y. URL https://doi.org/10.1007/s00454-020-00222-y.

Alain Pajor. *Sous-espaces ℓ_1^n des Espaces de Banach*. Travaux en Cours. Hermann, Paris, 1985.

Johann Radon. Mengen konvexer körper, die einen gemeinsamen punkt enthalten. *Mathematische Annalen*, 83(1):113–115, Mar 1921. ISSN 1432-1807. doi: 10.1007/BF01464231. URL https://doi.org/10.1007/BF01464231.

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 I. P. 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.

Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory (A)*, 13(1):145–147, 1972.

Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning - From Theory to Algorithms*. Cambridge University Press, 2014. ISBN 978-1-10-705713-5.

Saharon Shelah. A combinatorial problem, stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41(1):247–261, 1972. doi: 10.2140/pjm.1972.41.247.

M.L.J. van de Vel. *Theory of Convex Structures*, volume 50 of *North-Holland mathematical library*. North-Holland, 1993. ISBN 9780444815057. URL https://books.google.com/books?id=xt9-lAEACAAJ.

Vladimir N. Vapnik and Alexey Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Proc. USSR Acad. Sci.*, 181(4):781–783, 1968.

Daniel Vendruscolo, Patricia E Desideri, Pedro LQ Pergher, et al. Some generalizations of the Borsuk-Ulam theorem. *Plubl. Math. Debrecen*, 78:583–593, 2011.

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.

Avi Wigderson. *A Theory Revolutionizing Technology and Science*. Princeton University Press, Princeton, 2019. ISBN 9780691192543. doi: doi:10.1515/9780691192543. URL https://doi.org/10.1515/9780691192543.

Appendix A. Examples

We now give examples that demonstrate that our bounds relating the Radon number, dual VC dimension, and VC dimension are nearly tight. These examples reveal that the lower bounds in Theorems B, C, and D are tight, while the upper bounds are tight within constant multiplicative factors. The determination of more precise bounds is left open for future research. This is summarized in Table 1 below. The detailed analysis of the examples appears in Appendix B.

	Theorem B (VC vs. dual VC)	Theorem C (VC vs. Radon)	Theorem D (dual VC vs. Radon)
Upper Bound	Examples 4 and 5	Examples 3, 4, and 5	Example 3
Lower Bound	Example 2	Examples 2, 3, and 5	Examples 2 and 5

Table 1: Summary of examples demonstrating the tightness of our bounds. Our lower bounds are tight, while the upper bounds are tight within constant multiplicative factors.

Our initial example demonstrates that for non-extremal concept classes $\mathcal{C}$, the Radon number cannot be upper-bounded solely as a function of the VC or dual VC dimensions. (However, the lower bound by the dual VC dimension remains valid even for non-extremal classes, as stated in Theorem D.)

Example 1 (Singletons) *The class of singletons $\mathcal{C} = \{1_{\{i\}} : i \in [n]\}$ is not extremal, and*

1. *$\mathtt{vc}(\mathcal{C}) = 1$.*

2. *$\mathtt{vc}^\star(\mathcal{C}) = 1$,*

3. *$\mathtt{r}(\mathcal{C}) = n$.*

This illustrates that for non-extremal classes, the Radon number can be unbounded while both the VC and dual VC dimensions remain as low as 1.

The upcoming examples pertain to Theorems B, C, and D. Refer to Table 1 for a roadmap linking each example to its theorem.

Example 2 (Cube) *The cube $\mathcal{C} = \{0, 1\}^d$ is a maximum class satisfying:*

1. *$\mathtt{vc}(\mathcal{C}) = d$,*

2. $\text{vc}^\star(\mathcal{C}) = \lfloor \log d \rfloor$,

3. $\text{r}(\mathcal{C}) = \lfloor \log(2d+2) \rfloor$.

This demonstrates the tightness of the first lower bound in Theorem C and the lower bound in Theorem B.

Example 3 (Dented cube) *The dented cube $\mathcal{C}$ is obtained by removing the all-1 function from the cube $\{0,1\}^{d+1}$. It is a maximum class satisfying:*

1. $\text{vc}(\mathcal{C}) = d$,

2. $\text{vc}^\star(\mathcal{C}) = \lfloor \log(d+1) \rfloor$,

3. $\text{r}(\mathcal{C}) = d + 1$.

This example demonstrates that the second lower bound in Theorem C is tight. Additionally, it shows that the upper bounds in Theorem C and Theorem D are tight within a constant multiplicative factor.

Example 4 *Let*

$$\mathcal{C} = \{01010101, 11010101, 10010101, 01110101,$$
$$01100101, 01011101, 01011001, 01010111, 01010110\}.$$

This class is maximum and satisfies:

1. $\text{vc}(\mathcal{C}) = 1$,

2. $\text{vc}^\star(\mathcal{C}) = 3$ *(any three out of the four blue concepts are dually shattered)*,

3. $\text{r}(\mathcal{C}) = 3$.

This confirms that the upper bounds in Theorem B and Theorem C are tight for $d = 1$.

A.1. Hyperplane Arrangements

We conclude with a geometric example of linear classifiers in $\mathbb{R}^d$. Consider a collection of hyperplanes $H_1, \ldots, H_n \subseteq \mathbb{R}^d$ where each H_i is defined as $H_i = \{x : \langle a_i, x \rangle = b_i\}$ with $a_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$. For each point $x \in \mathbb{R}^d \setminus (\cup_{i=1}^n H_i)$, associate a sign vector $c_x \in \{\pm\}^n$ defined by:

$$c_x(i) = \begin{cases} + & \text{if } \langle a_i, x \rangle > b_i, \\ - & \text{if } \langle a_i, x \rangle < b_i. \end{cases}$$

Define the class $\mathcal{C} = \{c_x : x \in \mathbb{R}^d\}$. Each $c \in \mathcal{C}$ corresponds to a cell induced by the hyperplane arrangement (see Figure 3 for an illustration). We say that the hyperplanes are generic if for every subset of hyperplanes $\{H_{i_1}, \ldots, H_{i_k}\}$, the intersection $\cap_{j=1}^k H_{i_j}$ has co-dimension k (and every $d + 1$ hyperplanes do not intersect). For every d, n, there are generic hyperplane arrangements, and as the name suggests a "typical" arrangement is generic. If the hyperplanes are generic, the corresponding $\mathcal{C}$ is a maximum class of VC dimension d (Moran, 2012).
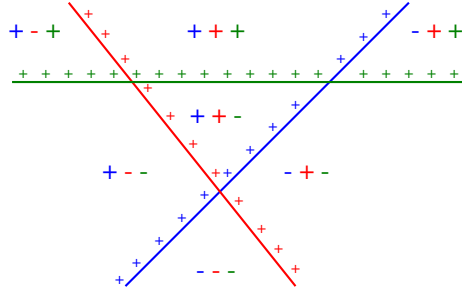
Figure 3: A 2-dimensional hyperplane arrangement consisting of 3 lines and 7 cells. The class $\mathcal{C}$ is maximum. Its VC dimension is two because $|\mathcal{C}| < 8$. Its dual VC dimension is one because $|\mathcal{C}^\star| < 4$. Its Radon number is three, because the three concept $+ - +, - + +, - - -$ are Radon indepedent.

The convexity space $\mathcal{F} = \mathcal{F}_C$ corresponds to the standard notion of convexity. The half-space $\mathcal{C}_{i,y}$ corresponds to the positive or negative, depending on y, half-space defined by H_i in $\mathbb{R}^d$. The classical Radon Theorem implies that $\mathbf{r}(\mathcal{C}) \leq d + 1$ for all hyperplane arrangements in $\mathbb{R}^d$ (not necessarily generic). Theorem D says that $\mathbf{vc}^\star(\mathcal{C}) \leq \mathbf{r}(\mathcal{C}) \leq d + 1$.

Take $d + 1$ generic hyperplanes that support a simplex. The total number of cells we get is $2^{d+1} - 1$. The VC dimension is d. The Radon number is $d + 1$ because the cells obtained by flipping (with respect to the simplex) all hyperplanes but one are Radon independent. The dual VC dimension is however at most $\log(d + 1)$. In fact, we obtained a different description of Example 3.

To make the dual VC dimension larger, let $p_1, \ldots, p_{d+1} \in \mathbb{R}^d$ be $d + 1$ points that are shattered by hyperplanes $h_1, \ldots, h_{2^{d+1}}$. These points can be chosen, for example, as the vertices of a full-dimensional simplex. The hyperplanes can be chosen to be generic so that we get a maximum class of VC dimension d. By construction, the dual VC dimension is at least $d + 1$, and, as discussed above, it is also at most $d + 1$.

Example 5 *There are maximum classes $\mathcal{C}$ defined by arrangements of generic hyperplanes in $\mathbb{R}^d$ fulfilling:*

1. $\mathbf{vc}(\mathcal{C}) = d$,

2. $\mathbf{vc}^\star(\mathcal{C}) = d + 1$,

3. $\mathbf{r}(\mathcal{C}) = d + 1$.

This example demonstrates the tightness of the lower bounds in Theorem C and Theorem D, and that the upper bounds in Theorem B and Theorem C are tight within a factor of 2.

A more general type of classes, called *pseudogeometric range spaces*, was defined and studied in Gärtner and Welzl (1994). Intuitively, this class is an abstraction of the class obtained from generic hyperplanes in $\mathbb{R}^d$, with pseudo-hyperplanes instead of hyperplanes, where pseudo-hyperplanes are topological abstractions of linear hyperplanes. Their definition, however, is combinatorial and is inductively defined based on the following two ideas. First, a one-dimensional pseudogeometric range space is a, combinatorially defined, line. Second, for $d \geq 2$, a class is a d-dimensional pseudogeometric range space if its intersection with any pseudo-hyperplane from

19

it is a $(d-1)$-dimensional pseudogeometric range space; here again, the intersection is defined combinatorially.

Our work suggests a different way to define and think of pseudo-hyperplane arrangements. Let us first imagine a generic hyperplane arrangement realized in $\mathbb{R}^d$, and denote by $\mathcal{C}$ its concept class. The cube complex $Q = Q(\mathcal{C})$ can be thought of as a dual object to the geometric realization. Each point in Q is a cell in $\mathcal{C}$. Two adjacent cells are connected by an edge in Q, four cells that meet at a point define a square, and so forth. Thus, Q, as a topological space, can be embedded into $\mathbb{R}^d$.

Based on this observation, it is natural to suggest that a maximum class $\mathcal{C}$ of VC dimension d is a d-dimensional pseudogeometric range space in the sense of Gärtner and Welzl (1994) if and only if its cube complex $Q(\mathcal{C})$ is embeddable into $\mathbb{R}^d$. This equivalence will be the topic of a future study.

Appendix B. Proofs and Analysis of Examples

Here we analyze all examples in Appendix A. Some of the bounds in our main results are deduced as corollaries.

Proposition 21 (First Lower Bound in Theorem C) *For any class $\mathcal{C}$,*

$$\lfloor \log(2\mathtt{vc}(\mathcal{C}) + 2) \rfloor \leq \mathtt{r}(\mathcal{C}).$$

Proof Suppose $\mathtt{vc}(\mathcal{C}) = d$, and, without losing generality, assume that $\mathcal{C}$ shatters $[d]$. Let $k = \lfloor \log(2d+2) \rfloor$. In particular, $D = (2^k - 2)/2 \leq d$. We say that (I, J) is a (non-trivial, ordered) partition of $[k]$ if $I, J \neq \emptyset$, $I \sqcup J = [k]$, and $1 \in I$. Let $\mathcal{I}$ be a set of all such partitions; the last condition ensures that precisely one of (I, J) and (J, I) is in $\mathcal{I}$. It is easy to see that then $|\mathcal{I}| = (2^k - 2)/2$, and let us take an arbitrary bijection $(I, J) \mapsto x_{I,J}$ from $\mathcal{I}$ to $[D]$.

Let us now take $c_i \in \mathcal{C}$, for $i \in [k]$, such that for every $(I, J) \in \mathcal{I}$, $c_i(x_{I,J}) = 0$ if $i \in I$ and 1 if $i \in J$. Such c_i exist because $\mathcal{C}$ shatters $[D] \subseteq [d]$. We claim that the concepts $c_1, \ldots, c_k$ are Radon independent. Indeed, let $(I, J) \in \mathcal{I}$ be a partition of $[k]$, and let $x = x_{I,J}$. Then $c_i(x) = 0$ for all $i \in I$, and $c_j(x) = 1$ for all $j \in J$. But then $I \subseteq \mathcal{C}_{x,0}$ and $J \subseteq \mathcal{C}_{x,1}$, which implies that $\mathtt{conv}_{\mathcal{C}}(I) \cap \mathtt{conv}_{\mathcal{C}}(J) \subseteq \mathcal{C}_{x,0} \cap \mathcal{C}_{x,1} = \emptyset$, as needed. ∎

Proposition 22 (Example 2 - Cube) *For the cube $\mathcal{C} = \{0, 1\}^d$:*

1. $\mathtt{vc}(\mathcal{C}) = d$,

2. $\mathtt{vc}^\star(\mathcal{C}) = \lfloor \log d \rfloor$,

3. $\mathtt{r}(\mathcal{C}) = \lfloor \log(2d + 2) \rfloor$.

This example also confirms the last lower bound in Theorem C.

Proof The equality $\mathtt{vc}(\mathcal{C}) = d$ is trivial. The bound $\mathtt{vc}^\star(\mathcal{C}) = \lfloor \log d \rfloor$ is also trivial and known from the sharpness of the bound in Theorem 9; see Assouad (1983). From Proposition 21, $\mathtt{r}(\mathcal{C}) \geq \lfloor \log(2d + 2) \rfloor$. It remains to prove that $\mathtt{r}(\mathcal{C}) \leq \log(2d + 2)$. The argument goes along the lines of the proof of Proposition 21.

We first prove the followng technical claim: *for $I, J \subseteq \mathcal{C}$, $\mathrm{conv}_\mathcal{C}(I) \cap \mathrm{conv}_\mathcal{C}(J) = \emptyset$ if and only if there is $x \in [d]$ that separates I and J, that is, such that $c(x) = 0$ for all $c \in I$ and $c(x) = 1$ for all $c \in J$, or the other way round.*

Indeed, if there is such x, then $\mathrm{conv}_\mathcal{C}(I)$ and $\mathrm{conv}_\mathcal{C}(J)$ are trivially disjoint. In the other direction, suppose $\mathrm{conv}_\mathcal{C}(I) \cap \mathrm{conv}_\mathcal{C}(J) = \emptyset$. As explained below Proposition 18, every non-empty convex set Y in $\mathcal{F}_\mathcal{C}$ is of the form $Y = \mathcal{C}_{X,t}$ for $X \subseteq [d]$ and $t \colon X \to \{0,1\}$. Write $\mathrm{conv}_\mathcal{C}(I) = \mathcal{C}_{X_I, t_I}$ and $\mathrm{conv}_\mathcal{C}(J) = \mathcal{C}_{X_J, t_J}$. As $\mathrm{conv}_\mathcal{C}(I)$ and $\mathrm{conv}_\mathcal{C}(J)$ are disjoint, there is $x \in X_I \cap X_J$ such that $t_I(x) = 0$ and $t_J(x) = 1$, or vice versa. But, by the definition of $\mathcal{C}_{X,t}$, this means that $c(x) = 0$ for all $c \in I$ and $c(x) = 1$ for all $c \in J$, or the other way round, as needed.

Now, suppose that $c_1, \ldots, c_k \in \mathcal{C}$ are Radon shattered. We see that for each non-trivial ordered partition (I, J) of $[k]$, there is $x_{I,J} \in [d]$ that separates $\{c_i : i \in I\}$ and $\{c_j : j \in J\}$. Moreover, it is easy to see that the map $(I, J) \mapsto x_{I,J}$ on $\mathcal{I}$ is one-to-one. As $|\mathcal{I}| = (2^k - 2)/2$, it follows that $d \geq (2^k - 2)/2$. ∎

Proposition 23 (Lower Bound for Maximum Classes in Theorem C) *For any maximum class $\mathcal{C} \neq \{0,1\}^n$, we have $\mathrm{vc}(\mathcal{C}) + 1 \leq \mathrm{r}(\mathcal{C})$.*

We use the following known property of extremal classes (see Chalopin, Chepoi, Moran, and Warmuth (2022) or Chornomaz (2022)). For a class $\mathcal{C} \subseteq \{0,1\}^n$, a set $X \subseteq [n]$ is called *minimal non-shattered* if X is not shattered by $\mathcal{C}$ and Y is shattered by $\mathcal{C}$ for every $Y \subsetneq X$.

Lemma 24 *For every extremal class $\mathcal{C}$, if X is minimal non-shattered by $\mathcal{C}$, then X contains a unique forbidden trace t; that is, there is $t \colon X \to \{0,1\}$ such that*

- *For every $t' \colon X \to \{0,1\}$ such that $t' \neq t$, there is $c \in \mathcal{C}$ such that $t'(x) = c(x)$ for all $x \in X$.*

- *There is no $c \in \mathcal{C}$ such that $t(x) = c(x)$ for all $x \in X$.*

Proof [Of Proposition 23] Let $\mathcal{C} \neq \{0,1\}^n$ be a maximum class of VC dimension $0 < d < n$. Fix $X \subseteq [n]$ of size $|X| = d + 1$. Because $\mathcal{C}$ is maximum, the set X is minimal non-shattered. Let t be its unique forbidden trace. Define a family of traces $t_x \colon X \to \{0,1\}$ for $x \in X$, as $t_x(y) = t(y)$ for $y \neq x$ and $t_x(x) = 1 - t(x)$. For each $x \in X$, let $c_x \in \mathcal{C}$ be a concept witnessing trace t_x.

We claim that the family $\{c_x : x \in X\}$ is Radon independent, so that $\mathrm{r}(\mathcal{C}) \geq d + 1$. Indeed, let (I, J) be a non-trivial partition of X. By choice of $(c_i : i \in I)$,

$$\mathrm{conv}_\mathcal{C}(\{c_i : i \in I\}) \subseteq \mathcal{C}_{J, t|_J}$$

where $t|_J$ is a restriction of t to J. Similarly, $\mathrm{conv}_\mathcal{C}(\{c_j : j \in J\}) \subseteq \mathcal{C}_{I, t|_I}$. Hence,

$$\mathrm{conv}_\mathcal{C}(\{c_i : i \in I\}) \cap \mathrm{conv}_\mathcal{C}(\{c_j : j \in J\}) \subseteq \mathcal{C}_{J, t_J} \cap \mathcal{C}_{I, t_I} = \mathcal{C}_{X, t}.$$

But $\mathcal{C}_{X,t} = \emptyset$ because t is a forbidden trace of $\mathcal{C}$ on X. ∎

Proposition 25 (Example 3 - Dented cube) *Let $\mathcal{C} = \{0,1\}^{d+1} \setminus \{(1, 1, \ldots, 1)\}$ be the dented cube. Then,*

1. $\mathtt{vc}(\mathcal{C}) = d$,

2. $\mathtt{vc}^\star(\mathcal{C}) = \lfloor \log(d+1) \rfloor$,

3. $\mathtt{r}(\mathcal{C}) = d + 1$.

Proof The first item is trivial and the second follows from Proposition 22. We only prove $\mathtt{r}(\mathcal{C}) = d+1$. Proposition 23 says that $\mathtt{r}(\mathcal{C}) \geq d+1$. So we only need to show that $\mathtt{r}(\mathcal{C}) \leq d+1$. Suppose, towards contradiction, that there are $d+2$ Radon shattered concepts $c_1, \ldots, c_{d+2} \in \mathcal{C}$. Let us define partitions $I_i = \{c_i\}$ and $J_i = \{c_j : j \in [d+2] - \{i\}\}$ for $i \in [d+2]$. Write $\mathtt{conv}_\mathcal{C}(J_i) = \mathcal{C}_{X_i, t_i}$ for some $X_i \subseteq [n]$ and $t_i : X_i \to \{0, 1\}$. Because

$$\emptyset = \mathtt{conv}_\mathcal{C}(I_i) \cap \mathtt{conv}_\mathcal{C}(J_i) = I_i \cap \mathcal{C}_{X_i, t_i}$$

there is $x_i \in X_i$ such that $c_i(x_i) \neq t_i(x_i)$ and $c_j(x_i) = t_i(x_i)$ for all $j \neq i$. But this implies that the map $[d+2] \ni i \mapsto x_i \in [d+1]$ is one-to-one, which is a contradiction. ∎

Proposition 26 (Example 4) *For the class*

$$\mathcal{C} = \{01010101, 11010101, {\color{blue}10010101}, 01110101,$$
$${\color{blue}01100101}, 01011101, {\color{blue}01011001}, 01010111, {\color{blue}01010110}\},$$

it holds that

1. $\mathtt{vc}(\mathcal{C}) = 1$,

2. $\mathtt{vc}^\star(\mathcal{C}) = 3$ *(any three out of the four blue concepts are dually shattered),*

3. $\mathtt{r}(\mathcal{C}) = 3$.

Proof It can be verified that $\mathtt{vc}(\mathcal{C}) = 1$ and that it indeed dually shatters any three out of the four blue concepts, witnessing $\mathtt{vc}^*(\mathcal{C}) \geq 3$. This analysis can be simplified by noticing that $\mathcal{C}$ is highly symmetric: it remains the same under permutations that exchange the pairs of even and odd vertices, for example, $01234567 \mapsto 23671045$. The class $\mathcal{C}$ is maximum because $|\mathcal{C}| = 9 = \binom{8}{\leq 1}$. Theorem C implies that $3 \leq \mathtt{vc}^\star(\mathcal{C}) \leq \mathtt{r}(\mathcal{C}) \leq 2\mathtt{vc}(\mathcal{C}) + 1 = 3$, and so $\mathtt{vc}^\star(\mathcal{C}) = \mathtt{r}(\mathcal{C}) = 3$, as claimed. ∎

Proposition 27 (Theorem D)

1. *Every concept class $\mathcal{C}$ satisfies $\mathtt{vc}^\star(\mathcal{C}) \leq \mathtt{r}(\mathcal{C})$.*

2. *For general classes, $\mathtt{r}(\mathcal{C})$ is not upper-bounded by $\mathtt{vc}(\mathcal{C})$ or $\mathtt{vc}^*(\mathcal{C})$.*

3. *If $\mathcal{C}$ is extremal, then $\mathtt{r}(\mathcal{C}) \leq 2^{\mathtt{vc}^\star(\mathcal{C})+2} - 1$, and this bound is tight up to a multiplicative factor.*

Proof The fact that $\text{vc}^\star(\mathcal{C}) \leq \text{r}(\mathcal{C})$ was already proved in the discussion following Theorem C. The fact that $\text{r}(\mathcal{C})$ is, in general, unbounded with respect to $\text{vc}(\mathcal{C})$ or $\text{vc}^*(\mathcal{C})$ is demonstrated by the class of singletons (Example 1). For the third item, let $\mathcal{C}$ be extremal. Theorem C says that $\text{r}(\mathcal{C}) \leq 2\text{vc}(\mathcal{C}) + 1$. Theorem 9 says that $\text{vc}(\mathcal{C}) \leq 2^{\text{vc}^*(\mathcal{C})+1} - 1$. Thus,

$$\text{r}(\mathcal{C}) \leq 2\text{vc}(\mathcal{C}) + 1 \leq 2(2^{\text{vc}^*(\mathcal{C})+1} - 1) + 1 = 2^{\text{vc}^*(\mathcal{C})+2} - 1.$$

The tightness up to a factor of 2 is demonstrated by the dented cube (Example 3) where $d + 1 = 2^k - 1$. In this case, $\text{r}(\mathcal{C}) = 2^k - 1$ and $\text{vc}^*(\mathcal{C}) = \lfloor \log(2^k - 1) \rfloor = k - 1$. ∎

Appendix C. Discussion of the Main Proof

We conclude with a discussion of the Radon number, which is an interesting concept that merits a discussion in its own right. Most of the relevant concepts have been introduced in the main proof (Section 2).

The classical Radon Theorem asserts that the maximal size of a Radon-independent set in $\mathbb{R}^d$ is $d+1$ with respect to the usual convexity structure. Equivalently, it says that the maximal k for which there is a *linear* Radon map $f \colon \partial \Delta^{k+1} \to \mathbb{R}^d$ is $d - 1$. The Topological Radon Theorem extends it by allowing f to be any continuous function. This opens the way to defining Radon numbers for arbitrary topological spaces, as we no longer depend on neither convexity nor linearity.

Other concepts can be formulated similarly. For example, we can define the *Borsuk-Ulam number* $\text{bu}(X)$ of a topological space X as a maximal d such that there is a Borsuk-Ulam map $f \colon S^d \to X$; that is, a continuous map that does not collapse antipodal points. This way, Theorem 19 proves that for all topological spaces X,

$$\text{r}(X) \leq \text{bu}(X).$$

The Topological Radon theorem ($\text{r}(\mathbb{R}^d) = d - 1$) becomes a consequence of the Borsuk-Ulam theorem ($\text{bu}(\mathbb{R}^d) = d - 1$), and Lemma 17 ($\text{r}(K) \leq 2d - 1$) a consequence of Theorem 20 ($\text{bu}(K) \leq 2d - 1$). It is natural to ask if the opposite direction holds.

Open Question 4 *Is it true that for an arbitrary topological space X, we have $\text{r}(S) = \text{bu}(S)$? Is it true for the special case of simplicial complexes?*

The question can be thought of as a rigidity property of antipodality. A positive answer says that the existence of a Borsuk-Ulam map implies the existence of a Radon map. The Borsuk-Ulam condition just requires that there is no collapse of antipodal points. The Radon condition requires e.g. that the image of a point is disjoint from all of its "opposite faces".

In the one-dimensional case, the answer seems to be positive. For the special case of simplicial complexes, if there is a Borsuk-Ulam map from the circle S^1 to a graph, then the graph is not a path (or a collection of disjoint paths). Then in the graph, there is either a cycle or a vertex of degree at least three. In both cases, there is a Radon map from $\partial \Delta^2$ to this graph. For general topological spaces X, consider the path space of X. Namely, the space of continuous maps $[0, 1] \to X$. If there are two paths whose union is connected but is not a path then there is a Radon map (there is a point of "degree three"). Otherwise, the space X itself is a path, which is a contradiction to the existence of a Borsuk-Ulam map.

The argument for the one-dimensional case is different than the proof that a Radon map yields a Borsuk-Ulam map using Theorem 19. Even for the one-dimensional case, there is no proof of this kind. That is, there is a one-dimensional space X and a Borsuk-Ulam map $f : S^1 \to X$ so that for every continuous map $g : \partial\Delta^2 \to S^1$, the map $g \circ f$ is not Radon. The space X can be chosen to be the circle S^1, and the map f can be chosen to be the map that wraps around three times.

In the Topological Radon Theorem, we can deal with two types of maps: $\partial\Delta^{d+1} \to X$ and $\Delta^{d+1} \to X$. We used the first option, but it also makes sense to use the second. If we denote this modified version of the Radon number by $\mathbf{r}'(X)$, then $\mathbf{r}'(X) \leq \mathbf{r}(X)$ for all X. For Euclidean space, we have $\mathbf{r}'(\mathbb{R}^d) = \mathbf{r}(\mathbb{R}^d)$ for all d, which could explain why this distinction is not made in the context of the Topological Radon Theorem. However, while $\mathbf{r}(S^d) = \mathbf{bu}(S^d) = d$ as witnessed by the identity map, it seems that proving $\mathbf{r}'(S^d) = d - 1$ should not be too hard.

We were not able to find an *explicit* definitions of topological Radon or Borsuk-Ulam numbers in the literature. They were, however, implicitly studied. For example, the aforementioned Gonçalves et al. (2002) addressed $\mathbf{bu}(K)$, Jaworowski and Izydorek (1993) studied the Borsuk-Ulam number of contractible simplicial complexes, and Vendruscolo et al. (2011) of products of manifolds. The latter paper, in fact, explicitly introduces a setup akin to (and even more general than) our definition of $\mathbf{bu}(S)$, and also has a good list of relevant references. The most systematic approach that we are aware of is developed in the book Matoušek et al. (2003) under the name of $\mathbb{Z}_2$-coindex of a space. This setup, nevertheless, is different because the target space is also required to have an involution, and the maps to be equivariant.

Finally, let us discuss the *topological Radon number* of a class $\mathcal{C}$, defined as the Radon number of the cube complex $Q = Q(\mathcal{C})$:

$$\mathbf{r}_t(\mathcal{C}) = \mathbf{r}(Q).$$

The bound $\mathbf{r}(\mathcal{C}) \leq \mathbf{r}_t(\mathcal{C}) + 2$ from Proposition 16 is specific to extremal classes. For example, in the (non-extremal) singletons class $\mathcal{C}$ in Example 1, the complex Q is a collection of n isolated points, and so $\mathbf{r}_t(\mathcal{C}) = 0$ while $\mathbf{r}(\mathcal{C}) = n$. The connection between $\mathbf{r}$ and $\mathbf{r}_t$ for extremal classes utilizes their topological nature in an essential way.