Variants of VC dimension and their applications to dynamics

Guorong Gao, Jie Ma, Mingyuan Rong, and Tuan Tran

Abstract: Since its introduction by Vapnik and Chervonenkis in the 1960s, the VC dimension and its variants have played a central role in numerous fields. In this paper, we investigate several variants of the VC dimension and their applications to dynamical systems. First, we prove a new bound for a recently introduced generalization of VC dimension, which unifies and extends various extremal results on the VC, Natarajan, and Steele dimensions. This new bound allows us to strengthen one of the main theorems of Huang and Ye [Adv. Math., 2009] in dynamical systems. Second, we refine a key lemma of Huang and Ye related to a variant of VC dimension by providing a more concise and conceptual proof. We also highlight a surprising connection among this result, combinatorics, dynamical systems, and recent advances in communication complexity.

Keywords: VC dimension, variant, k-Natarajan dimension, dynamical system.

1. Introduction

The Vapnik–Chervonenkis dimension [30, 31], or VC dimension, is a combinatorial parameter of significant importance in various fields, including statistical learning theory [31, 5], probability [32], functional analysis [29], discrete and computational geometry [8, 16], model theory [27] and combinatorics [10]. The VC dimension of a family of binary vectors $\mathcal{H} \subseteq \{0,1\}^n$ is the maximum size of a set **shattered** by the family, i.e., a set $S \subseteq [n]$ such that the projection of $\mathcal{H}$ onto the coordinates of S equals $\{0,1\}^S$. A cornerstone result in this area is the Sauer–Shelah Lemma [26, 27, 31], which states that any family $\mathcal{H} \subseteq \{0,1\}^n$ with VC dimension d satisfies $|\mathcal{H}| \leq \sum_{i=0}^{d} \binom{n}{i}$. This

bound is tight, as exemplified by the family of all binary vectors of length n with at most d ones. Various proofs of the Sauer-Shelah lemma can be found in the literature, and numerous variants of the lemma are known (see, e.g., [28, 11, 9, 25, 2]).

In this paper, we study a generalization of the VC dimension recently introduced in computer science and game theory, a variant of the VC dimension applied to partial concept classes, and their implications in dynamical systems.

1.1. A generalization of VC dimension: k-Natarajan dimension

The concept of k-**Natarajan dimension** was introduced recently and has found applications in computer science, game theory [9], and machine learning [7]. For $r \geq k \geq 2$ and $n \geq 1$, the k-Natarajan dimension of a family $\mathcal{H} \subseteq \{1, \ldots, r\}^{[n]}$, denoted by $\dim_k(\mathcal{H})$, is defined as the maximum size of a subset $S \subseteq [n]$ such that the projection of $\mathcal{H}$ onto S contains a subfamily of the form $\prod_{i \in S} Y_i$, where each Y_i is a k-element subset of $\{1, \ldots, r\}$. When $r = k = 2$, this definition coincides with the VC dimension. For $k = 2$, it corresponds to the Natarajan dimension, and when $r = k$, it equals the Steele dimension. Our first result extends the Sauer–Shelah Lemma from the VC dimension to k-Natarajan dimension.

Theorem 1.1. *Let $r \geq k \geq 2$ and $n \geq d \geq 1$. For any family $\mathcal{H} \subseteq \{1, \ldots, r\}^{[n]}$ with $\dim_k(\mathcal{H}) \leq d$,*

$$|\mathcal{H}| \leq (k-1)^{n-d} \sum_{i=0}^{d} \binom{n-i-1}{d-i} \binom{r}{k}^{d-i} r^i.$$

Remarks.

(i) Theorem 1.1 gives $|\mathcal{H}| = O_{r,k,d}(n^d(k-1)^n)$. This bound is asymptotically tight, as shown by the family consisting of all vectors in which at most d coordinates are greater than $k-1$. Moreover, the bound is sharp whenever $r = k \geq 2$ (see (iii) and (v) below).

(ii) Weaker bounds were obtained by Daniely, Schapira and Shahaf [9, Theorem 1.5] and Charikar and Pabbaraju [7, Theorem 7] using a different method.

(iii) Theorem 1.1 recovers the classic Sauer–Shelah Lemma [31, 26, 27] by taking $r = k = 2$ and noting that $\sum_{i=0}^{d} \binom{n-i-1}{d-i} 2^i = \sum_{i=0}^{d} \binom{n}{i}$.

(iv) Specializing Theorem 1.1 to $k = 2$ yields an improved version of the Natarajan theorem [25].

(v) For $r = k$, Theorem 1.1 gives the same bound as in the Steele theorem [28].

Our novel approach in the proof of Theorem 1.1 is to construct a bijection between $\mathcal{H}$ and a classification table that captures the shattering relationship. This idea, inspired by [11], might be of independent interest. Clearly, Theorem 1.1 has broad applicability, since it unifies and enhances various extremal results on the VC dimension. As expected, in Theorem 1.6, we apply Theorem 1.1 to dynamical systems, improving one of the main theorems of Huang and Ye [17]. To the best of our knowledge, this is the first application of the k-Natarajan dimension to topological dynamics.

1.2. A variant of VC dimension to partial concept classes

Given integers $r \geq 2$ and $n \geq 1$, we consider partial concept classes $\mathcal{H} \subseteq \{1, \ldots, r, \star\}^{[n]}$, where each $h \in \mathcal{H}$ is a partial vector; specifically if $i \in [n]$ is such that $h(i) = \star$ then h is undefined at i. The (non-traditional) **VC dimension** of $\mathcal{H}$, denoted $\dim_{VC}(\mathcal{H})$, is the maximum size of a shattered set $S \subseteq Z$, where S is said to be **shattered** if the projection of $\mathcal{H}$ onto S contains $\{1, \ldots, r\}^S$. A family $\mathcal{F} \subseteq \{1, \ldots, r\}^{[n]}$ is called a **net** of $\mathcal{H}$ if for every $h \in \mathcal{H}$ there exists $f \in \mathcal{F}$ such that $h(i) \in \{1, \ldots, r, \star\} \setminus \{f(i)\}$ for all $i \in [n]$. The smallest possible size of a net of $\mathcal{H}$ is the **covering number** of $\mathcal{H}$, denoted $C(\mathcal{H})$.

To investigate maximal pattern entropy of a topological dynamical system (the notations for which will be introduced in Section 1.3), Huang and Ye [17] provide a crucial lemma showing that for sufficiently large n, if $\mathcal{H} \subseteq \{1, \ldots, r, \star\}^{[n]}$ is a partial concept class with $VC(\mathcal{H}) \leq d$, then

$$C(\mathcal{H}) \leq r^2 2^m \left(\frac{n}{m}\right)^{2m},$$

where $m := \log_{\frac{r+1}{r}} \binom{n}{\leq d} + 1$ and $\binom{n}{\leq d} := \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{d}$. In the following theorem, we strengthen this result and provide a more concise and conceptual proof.

Theorem 1.2. *Let $r \geq 2$ and $n \geq d \geq 1$. If $\mathcal{H} \subseteq \{1, \ldots, r, \star\}^{[n]}$ is a partial concept class with $VC(\mathcal{H}) \leq d$, then*

$$C(\mathcal{H}) \leq \binom{n}{\leq \log_{\frac{r}{r-1}} \binom{n}{\leq d}}.$$

(1)

Earlier versions of this lemma (see [31, 26, 27, 20, 22]) play a significant role in all aspects of local entropy theory for topological dynamical systems. For a thorough discussion we refer the reader to the survey by Glasner and Ye [12], and Chapter 12 of the book by Kerr and Li [23]. The case $r = 2$ of Theorem 1.2 is a result of Alon, Hanneke, Holzman and Moran [2, Theorem 12] in their study of PAC learning theory. Our proof extends their ideas.

When $r = 2$ and $d = 1$, the inequality (1) gives $C(\mathcal{H}) \leq \binom{n}{\leq \log(n+1)} \leq n^{\log(n+1)}$. More generally, one can show that the right-hand side of the inequality (1) is at most $n^{(r-1)d\log(n+1)}$. Note also that for $r = O(1)$ and $d = o(n)$, we have $\log_{\frac{r}{r-1}} \binom{n}{\leq d} = o(n)$ since $\binom{n}{\leq o(n)} = 2^{o(n)}$, which implies $C(\mathcal{H}) \leq \binom{n}{\leq o(n)} = 2^{o(n)}$.

Another result of Alon, Hanneke, Holzman and Moran [2, Theorem 11] shows that the inequality (1) is nearly tight when $r = 2$ and $d = O(1)$. Interestingly, its proof hinges on a recent breakthrough in communication complexity and its implications in graph theory by Balodis, Ben-David, Göös, Jain and Kothari [3].

Theorem 1.3 ([2]). *There is a partial concept class* $\mathcal{H} \subseteq \{1, 2, \star\}^{[n]}$ *with* $\mathrm{VC}(\mathcal{H}) \leq 1$ *and* $C(\mathcal{H}) \geq n^{(\log n)^{1-o(1)}}$.

Since Theorem 1.3 was stated in a different form in [2], we include its proof in Appendix A for completeness.

1.3. Applications to dynamical systems

The central object of study in topological dynamics is a topological dynamical system (TDS) (X, T), where X is a nonempty compact metrizable space and $T\colon X \to X$ is a continuous map. Ever since Adler, Konheim and McAndrew [1] introduced entropy into dynamical systems theory 60 years ago, it has played a very central role in the study of topological dynamical systems (see, for example, the surveys [15, 21]). They associated to any topological dynamical system (X, T) a topological invariant $h_{\mathrm{top}}(T) \in \mathbb{R}_+ \cup \{\infty\}$, called the **topological entropy** of (X, T), which measures the uncertainty or disorder of the system. Systems with positive topological entropy are random in certain sense, and systems with zero topological entropy are said to be **deterministic** even though they may exhibit complicated behaviours. Comparing to positive entropy systems, deterministic systems are much less understood. In order to distinguish between deterministic systems, Huang and Ye [17] introduced the concept of maximal pattern entropy $h^*_{\mathrm{top}}(T)$ of a topological

dynamical system (X, T). It is known that both $h_{\text{top}}(T)$ and $h_{\text{top}}^*(T)$ take value in $\{0, \log 2, \log 3, \ldots\} \cup \{\infty\}$, and that $h_{\text{top}}(T) > 0$ implies $h_{\text{top}}^*(T) = \infty$ (see [1, 17]). Hence maximal pattern entropy is especially useful for deterministic systems.

We now proceed to give more details. Let (X, T) be a TDS. Given two finite open covers $\mathcal{U}, \mathcal{V}$ (of X), their **joint** is defined as $\mathcal{U} \vee \mathcal{V} := \{A \cap B : A \in \mathcal{U}, B \in \mathcal{V}\}$. Clearly, $\mathcal{U} \vee \mathcal{V}$ is also an open cover of size at most $|\mathcal{U}||\mathcal{V}|$, and at least 1. For a finite open cover $\mathcal{U}$, let $p_{X,\mathcal{U}}^*\colon \mathbb{N} \to \mathbb{N}$ be the function given by

$$p_{X,\mathcal{U}}^*(n) = \max_{S \subset \mathbb{N} \cup \{0\}, |S| = n} N\big(\bigvee_{i \in S} T^{-i}\mathcal{U}\big),$$

where $N(\mathcal{V})$ denotes the minimum size of a subcover chosen from an open cover $\mathcal{V}$. It is easy to see that the sequence $\{\log p_{X,\mathcal{U}}^*(n)\}_{n \in \mathbb{N}}$ is sub-additive, and thus by Fekete's lemma, the limit $\lim_{n \to +\infty} \frac{1}{n} \log p_{X,\mathcal{U}}^*(n)$ exists. Denote this limit by $h_{\text{top}}^*(T, \mathcal{U})$. The **maximal pattern entropy** of (X, T) is then defined as

$$h_{\text{top}}^*(T) := \sup_{\mathcal{U}} h_{\text{top}}^*(T, \mathcal{U}),$$

where the supremum is over all finite open covers $\mathcal{U}$.[1]

We say a topological dynamical system (X, T) is **null** if $h_{\text{top}}^*(T) = 0$, that is, $h_{\text{top}}^*(T)$ attains the minimum possible value. From the definition of $h_{\text{top}}^*(T)$, we find that (X, T) is null if and only if $p_{X,\mathcal{U}}^*(n)$ grows sub-exponentially in n for each finite open cover $\mathcal{U}$. For such systems, an intriguing conjecture of Huang and Ye [17] further rules out the intermediate growth between polynomial and exponential.

Conjecture 1.4 ([17])**.** *If (X, T) is a null TDS, then $p_{X,\mathcal{U}}^*$ is of polynomial order for each finite open cover $\mathcal{U}$.*

The conjecture was repeated in the survey on local entropy theory by Glasner and Ye [12]. It was shown to be true for interval maps by Li [24], circle maps by Yang [33], and most interestingly for zero-dimensional systems by Huang and Ye [17]. A refinement of an open cover $\mathcal{U}$ of X is a new open cover $\mathcal{V}$ of X such that every set in $\mathcal{V}$ is contained in some set in $\mathcal{U}$. We say X is **zero-dimensional** if every finite open cover of X has a clopen (closed and open) refinement. Using the Karpovsky–Milman Theorem [20], which is a generalization of the Sauer–Shelah Lemma, Huang and Ye [17, Theorem 5.4] verified Conjecture 1.4 for such spaces X.

---

[1]This definition is independent of a choice of metric.

Theorem 1.5 ([17])**.** *Let* (X,T) *be a TDS and let* $\mathcal{U} = \{U_1, \ldots, U_r\}$ *be a clopen partition of* X. *Then* $h^*_{\text{top}}(T,\mathcal{U}) = \log \ell$ *for some* $\ell \in \{1, \ldots, r\}$. *Moreover, one of the following alternatives holds.*

(a) *If* $\ell = r$, *then* $p^*_{X,\mathcal{U}}(n) = r^n$ *for all* $n \in \mathbb{N}$.
(b) *If* $\ell = r - 1$, *then there exists* $c > 0$ *such that* $\ell^n \leq p^*_{X,\mathcal{U}}(n) \leq n^c \ell^n$ *for all* $n \geq 2$.
(c) *If* $2 \leq \ell \leq r - 2$, *then there exists* $c > 0$ *such that* $\ell^n \leq p^*_{X,\mathcal{U}}(n) \leq n^{c \log n} \ell^n$ *for all* $n \geq 2$.
(d) *If* $\ell = 1$, *then there exists* $c > 0$ *such that* $p^*_{X,\mathcal{U}}(n) \leq n^c$ *for all* $n \geq 2$.

Alternative (d) tells us that Conjecture 1.4 does hold when X is zero-dimensional. Indeed, let $\mathcal{U}$ be any finite open cover of X. Then $\mathcal{U}$ has a clopen refinement $\mathcal{V} = \{V_1, \ldots, V_r\}$, which forms a partition of X. Since $\mathcal{V}$ is a refinement of $\mathcal{U}$, $p^*_{X,\mathcal{U}}(n) \leq p^*_{X,\mathcal{V}}(n)$ for all n. Moreover, as (X,T) is a null TDS, $h^*_{\text{top}}(T,\mathcal{U}) = 0$, and so it follows from Theorem 1.5 (d) that $p^*_{X,\mathcal{V}}(n) \leq n^c$ for $n \geq 2$. Therefore, $p^*_{X,\mathcal{U}}(n) \leq n^c$ for all $n \geq 2$, as desired.

Huang and Ye [17] handled alternatives (b)–(d) separately. Using our combinatorial lemma (i.e., Theorem 1.1), we improve the multiplicative factor in (c) from quasipolynomial to polynomial in n, and provide a unified treatment of (b)–(d) as follows.

Theorem 1.6. *Let* (X,T) *be a TDS, and let* $\mathcal{U} = \{U_1, \ldots, U_r\}$ *be a clopen partition of* X. *Then* $h^*_{\text{top}}(T,\mathcal{U}) = \log \ell$ *for some* $\ell \in \{1, \ldots, r\}$. *Moreover, one of the following alternatives holds:*

(a) *If* $\ell = r$, *then* $p^*_{X,\mathcal{U}}(n) = r^n$ *for all* $n \in \mathbb{N}$.
(b) *If* $\ell \leq r - 1$, *then there exists* $c > 0$ *such that* $\ell^n \leq p^*_{X,\mathcal{U}}(n) \leq n^c \ell^n$ *for all* $n \geq 2$.

For general topological spaces, the following result, due to Huang and Ye [17, a special case of Theorem 4.5], represents the current state of the art of Conjecture 1.4.

Theorem 1.7 ([17])**.** *If* (X,T) *is a null TDS, then for each finite open cover* $\mathcal{U}$ *there is a constant* $c = c(\mathcal{U})$ *such that* $p^*_{X,\mathcal{U}}(n) \leq n^{c \log n}$ *for every* $n \geq 2$.

At the heart of Huang and Ye's argument in Theorem 1.7 is an earlier version of Theorem 1.2. To illustrate the application of Theorem 1.2 in dynamical systems, we provide a (simplified) proof of Theorem 1.7 in Appendix B.

1.4. Notation and organization

We adopt standard notation throughout. In particular, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. Given a set S and $k \in \mathbb{N}$, we write $\binom{S}{k}$ for the collection of all k-subsets of S. The notation Y^Z refers to the family of all functions (or vectors) from a set Z to a set Y. For $h \in Y^Z$ and $S \subseteq Z$, the projection of h onto S, denoted $h|_S$, is the map in Y^S defined by $i \mapsto h(i)$. Similarly, for $\mathcal{H} \subseteq Y^Z$ and $S \subseteq Z$, the projection of $\mathcal{H}$ onto S is given by

$$\mathcal{H}|_S = \{h|_S : h \in \mathcal{H}\}.$$

For two nonnegative functions f and g of some underlying parameter n, we write $f = O(g)$ if there exist positive constants C and n_0 such that $f(n) \leq Cg(n)$ for all $n \geq n_0$, and $f = o(g)$ if $\lim_{n \to \infty} f(n)/g(n) = 0$. To simplify the presentation, we systematically omit floor and ceiling signs. Unless specified otherwise, all logarithms are taken to base 2.

   We give the proof of Theorem 1.1 in Section 2 and the proof of Theorem 1.2 in Section 3. We then derive Theorem 1.6 from Theorem 1.1 in Section 4.1. In Section 4.2 we use Theorem 1.6 to study the complexity of symbolic dynamics (see Theorem 4.2). Finally, we close the paper with some concluding remarks in Section 5.

2. k-Natarajan dimension

In this section we provide a proof of Theorem 1.1. Throughout the section, we view elements of product spaces as vectors. For a vector x, we denote by x_i the value of the i-th coordinate. For integers a and b, we employ the interval notation

$$[a, b] := \{x \in \mathbb{Z} : a \leq x \leq b\}.$$

The following lemma makes up the bulk of the proof of Theorem 1.1.

Lemma 2.1. *Given integers r and k with $r \geq k \geq 2$, let Ω denote the alphabet*

$$\{b_1, b_2, \ldots, b_{k-1}\} \cup \left\{c_A \colon A \in \binom{[r]}{k}\right\}$$

of size $k - 1 + \binom{r}{k}$. Then for every family $\mathcal{H} \subseteq [r]^{[n]}$, there exist n maps $\varphi_1, \ldots, \varphi_n$ and $n + 1$ families $\mathcal{H}_0 := \mathcal{H}, \mathcal{H}_1, \ldots, \mathcal{H}_n$ (see the diagram below) with the following properties.

(P1) For $0 \leq i \leq n$, $\mathcal{H}_i$ is a subfamily of $\Omega^{[i]} \times [r]^{[i+1,n]}$.

(P2) *For $1 \leq i \leq n$, $\varphi_i \colon \mathcal{H}_{i-1} \to \mathcal{H}_i$ is a bijection that preserves all but the i-th coordinate.*

(P3) *Every element in $\mathcal{H}_n$ has at most $\dim_k(\mathcal{H})$ coordinates in $\{c_A \colon A \in \binom{[r]}{k}\}$.*

$$[r]^{[n]} \qquad \Omega^{[1]} \times [r]^{[2,n]} \qquad \Omega^{[2]} \times [r]^{[3,n]} \qquad \cdots \qquad \Omega^{[n]}$$
$$\cup| \qquad\qquad \cup| \qquad\qquad \cup| \qquad\qquad\qquad \cup|$$
$$\mathcal{H}_0 = \mathcal{H} \xrightarrow{\ \varphi_1\ } \mathcal{H}_1 \xrightarrow{\ \varphi_2\ } \mathcal{H}_2 \xrightarrow{\qquad\quad} \cdots \xrightarrow{\ \varphi_n\ } \mathcal{H}_n$$

Assuming Lemma 2.1, we now give a proof of Theorem 1.1.

Proof of Theorem 1.1 (assuming Lemma 2.1). Let us consider a family $\mathcal{H} \subseteq [r]^{[n]}$ with $\dim_k(\mathcal{H}) \leq d$. Applying Lemma 2.1 to $\mathcal{H}$, we get

$$\Omega, \varphi_1, \ldots, \varphi_n, \mathcal{H}_1, \ldots, \mathcal{H}_n.$$

Set $\Omega' = \{b_1, \ldots, b_{k-1}\}, \Omega'' = \{c_A \colon A \in \binom{[r]}{k}\}$ and $\varphi = \varphi_n \circ \cdots \circ \varphi_1$. By (P3), for every $x \in \mathcal{H}$, $\varphi(x)$ has at least $n - d$ coordinates in Ω'. Hence $\mathcal{H}$ is partitioned into $d + 1$ subfamilies $\mathcal{G}_0, \ldots, \mathcal{G}_d$, where

$$\mathcal{G}_i = \{x \in \mathcal{H} \colon \varphi(x)_{n-i} \in \Omega' \text{ and } \varphi(x)|_{[n-i]} \text{ has exactly } n-d \text{ coordinates in } \Omega'\}.$$

Denote

$$\mathcal{F}_i = \{(\varphi_{n-i} \circ \cdots \circ \varphi_1)(x) \colon x \in \mathcal{G}_i\}.$$

From (P1) and (P2), we get $\mathcal{F}_i \subseteq \Omega^{[n-i]} \times [r]^{[n-i+1,n]}$ and $|\mathcal{F}_i| = |\mathcal{G}_i|$. To bound the size of $\mathcal{F}_i$, let y be any vector of $\mathcal{F}_i$. Then $y = (\varphi_{n-i} \circ \cdots \circ \varphi_1)(x)$ for some $x \in \mathcal{G}_i$. By (P2), $\varphi(x) = (\varphi_n \circ \cdots \circ \varphi_{n-i+1})(y)$ and y agree in the first $n-i$ coordinates. On the other hand, as $x \in \mathcal{G}_i$, we know that $\varphi(x)_{n-i} \in \Omega'$, and that $\varphi(x)|_{[n-i]}$ has exactly $n - d$ (respectively $d - i$) coordinates in Ω' (respectively Ω''). Therefore, $y_{n-i} \in \Omega'$, and $y|_{[n-i]}$ has exactly $n - d$ (respectively $d - i$) coordinates in Ω' (respectively Ω''). From this, we obtain

$$|\mathcal{F}_i|_{[n-i]}| \leq \binom{n-i-1}{n-d-1}(k-1)^{n-d}\binom{r}{k}^{d-i} = (k-1)^{n-d}\binom{n-i-1}{d-i}\binom{r}{k}^{d-i}.$$

Combining this with the trivial bound $|\mathcal{F}_i|_{[n-i+1,n]}| \leq r^i$ yields

$$|\mathcal{F}_i| \leq |\mathcal{F}_i|_{[n-i]}| \cdot |\mathcal{F}_i|_{[n-i+1,n]}| \leq (k-1)^{n-d}\binom{n-i-1}{d-i}\binom{r}{k}^{d-i} \cdot r^i.$$

Therefore, we have

$$|\mathcal{H}| = \sum_{i=0}^{d} |\mathcal{G}_i| = \sum_{i=0}^{d} |\mathcal{F}_i| \le (k-1)^{n-d} \sum_{i=0}^{d} \binom{n-i-1}{d-i} \binom{r}{k}^{d-i} r^i,$$

finishing the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Construction. We recursively define $\varphi_1, \mathcal{H}_1, \ldots, \varphi_n, \mathcal{H}_n$. Suppose we have defined

$$\mathcal{H}_0, \varphi_1, \mathcal{H}_1, \ldots, \varphi_{i-1}, \mathcal{H}_{i-1}$$

for some $1 \le i \le n$. We shall construct a map $\varphi_i \colon \mathcal{H}_{i-1} \to \Omega^{[i]} \times [r]^{[i+1,n]}$ and let $\mathcal{H}_i = \varphi_i(\mathcal{H}_{i-1})$. For $x \in \mathcal{H}_{i-1}$, the i-th *block* of $\mathcal{H}_{i-1}$ containing x is the family

$$\partial^{(i)}(x) = \{y \in \mathcal{H}_{i-1} \colon x \text{ and } y \text{ agree in all but possibly the } i\text{-th coordinate}\}.$$

Note that $x \in \partial^{(i)}(x)$. Since $y \in \partial^{(i)}(x)$ if and only if $\partial^{(i)}(x) = \partial^{(i)}(y)$, $\mathcal{H}_{i-1}$ is decomposed into i-th blocks. We thus only need to define φ_i on each i-th block. Given $x = (x_1, \ldots, x_n) \in \mathcal{H}_{i-1}$, order the elements of $\partial^{(i)}(x)$ as

$$(x_1, \ldots, x_{i-1}, s_1, x_{i+1}, \ldots, x_n), \ldots, (x_1, \ldots, x_{i-1}, s_t, x_{i+1}, \ldots, x_n),$$

where $1 \le s_1 < \ldots < s_t \le r$.

For $1 \le j \le t$, define $\varphi_i(x_1, \ldots, x_{i-1}, s_j, x_{i+1}, \ldots, x_n)$ to be the vector

- Except the i-th coordinate, the other coordinates of

$$\varphi_i(x_1, \ldots, x_{i-1}, s_j, x_{i+1}, \ldots, x_n)$$

  and x are the same;
- The i-th coordinate of $\varphi_i(x_1, \ldots, x_{i-1}, s_j, x_{i+1}, \ldots, x_n)$ is b_j if $1 \le j \le k-1$, and $c_{\{s_1, \ldots, s_{k-1}, s_j\}}$ if $j \ge k$.

As these vectors $\varphi_i(x_1, \ldots, x_{i-1}, s_j, x_{i+1}, \ldots, x_n)$ lie in $\Omega^{[i]} \times [r]^{[i+1,n]}$, $\mathcal{H}_i = \varphi_i(\mathcal{H}_{i-1})$ is a subfamily of $\Omega^{[i]} \times [r]^{[i+1,n]}$, thereby verifying (P1).

To verify the other properties, we first make some simple observations that follow readily from the definition of φ_i.

Observation 2.2. *The following hold for every $i \in [n]$ and $x \in \mathcal{H}_{i-1}$.*

(a) $\varphi_i(x)$ preserves all but the i-th coordinate of x.
(b) The restriction of φ_i to $\partial^{(i)}(x)$ is injective.

(c) *Suppose $x = (x_1, \ldots, x_n) \in \mathcal{H}$ and the i-th coordinate of $\varphi_i(x)$ is c_A for some $A \in \binom{[r]}{k}$. Then $\mathcal{H}_{i-1} \supseteq \partial^{(i)}(x)$ contains $\{x_1\} \times \cdots \times \{x_{i-1}\} \times A \times \{x_{i+1}\} \times \cdots \times \{x_n\}$.*

Proof of Lemma 2.1 (continuation). (P2) As $\mathcal{H}_i = \varphi_i(\mathcal{H}_{i-1})$, evidently φ_i is surjective. So what's left is to show that φ_i is injective. Suppose $\varphi_i(x) = \varphi_i(y)$. By Observation 2.2 (a), φ_i preserves all but the i-th coordinate, so x and y agree in all but possibly the i-th coordinate. Thus, x and y are two vectors of $\partial^{(i)}(x)$ with $\varphi_i(x) = \varphi_i(y)$. But by Observation 2.2 (b), the map φ_i restricted to $\partial^{(i)}(x)$ is injective, so one must have $x = y$. This proves (P2).

(P3) Let $d = \dim_k(\mathcal{H})$. Suppose for the contrary that there is a vector $y \in \mathcal{H}_n$ together with a $(d+1)$-element subset $I \subseteq [n]$ such that for every $i \in I$, $y_i = c_{A_i}$ for some $A_i \in \binom{[r]}{k}$. By repeatedly applying Observation 2.2 (a) and (c) in a reversed ordering with respect to the coordinates, we find that $\mathcal{H}|_I$ contains $\prod_{i \in I} A_i$, implying $\dim_k(\mathcal{H}) \geq d + 1$, a contradiction. This finishes the proof. $\square$

3. A variant of VC dimension to partial concept classes

In this section, we provide a short proof of Theorem 1.2. We need to show that any partial concept class with bounded VC-dimension admits a small net. We shall construct the net via an algorithm. We first fix some notations and make some innocuous observations.

Let $\mathcal{H} \subseteq \{1, \ldots, r, \star\}^Z$ be a partial concept class. Define the **shattering strength** of $\mathcal{H}$, which is denoted by $s(\mathcal{H})$, to be the number of subsets $S \subseteq Z$ that are shattered by $\mathcal{H}$. By convention, the shattering strength of the empty class is 0, and the empty set is shattered by all nonempty classes (and so the shattering strength of any nonempty class is at least 1). It is easy to see that $s(\mathcal{H}) \leq \binom{|Z|}{\leq \mathrm{VC}(\mathcal{H})}$. For $(i, j) \in Z \times [r]$, we denote

$$\mathcal{H}_{i \to j} = \{h \in \mathcal{H} \colon h(i) = j\}.$$

Define the VC-minority function $M_{\mathcal{H}} \colon Z \to [r]$ of $\mathcal{H}$ by letting $M_{\mathcal{H}}(i)$ be the value $j \in [r]$ which minimises $s(\mathcal{H}_{i \to j})$, with an arbitrary tie-breaking rule. Observe that for any $i \in Z$,

$$(2) \qquad (r-1) \cdot s(\mathcal{H}) \geq s(\mathcal{H}_{i \to 1}) + \ldots + s(\mathcal{H}_{i \to r}).$$

In particular,

$$(3) \qquad s(\mathcal{H}_{i \to j}) \leq \tfrac{r-1}{r} \cdot s(\mathcal{H}), \text{ where } j = M_{\mathcal{H}}(i).$$

To see (2), for any subset $S \subseteq Z$ with $i \notin S$, we consider the contribution of the pair $S, S \cup \{i\}$ to both sides of the inequality. We note that every set S that is shattered by one of the classes $\mathcal{H}_{i \to 1}, \ldots, \mathcal{H}_{i \to r}$ is also shattered by $\mathcal{H}$, and if S is shattered by all of the $\mathcal{H}_{i \to j}$ then both S and $S \cup \{i\}$ are shattered by $\mathcal{H}$.

We shall use the following algorithm to construct a small net of a given partial concept class.

The algorithm. Fix a partial concept class $\mathcal{H} \subseteq \{1, \ldots, r, \star\}^{[n]}$. For any partial function $h \in \mathcal{H}$, the algorithm will output an index set $A^n \subseteq [n]$ and a total function $f \in \{1, \ldots, r\}^{[n]}$. Set $A^0 = \emptyset$ and $\mathcal{H}^0 = \mathcal{H}$. For $i = 1, \ldots, n$, do the following:

(1) Compute the value of the VC-minority function of $\mathcal{H}^{i-1}$ at i. Denote this value by j.
(2) If $h(i) \neq j$, then set $f(i) = j$, $A^i = A^{i-1}$ and $\mathcal{H}^i = \mathcal{H}^{i-1}$.
(3) If $h(i) = j$, then set $f(i) = j + 1 \pmod{r}$, $A^i = A^{i-1} \cup \{i\}$ and $\mathcal{H}^i = (\mathcal{H}^{i-1})_{i \to j}$.

The outputs of the algorithm satisfy the following properties.

Lemma 3.1.

(a) $h \in \mathcal{H}^i$ for every $0 \leq i \leq n$. In particular, $s(\mathcal{H}^i) \geq 1$ for every $0 \leq i \leq n$.
(b) $h(i) \neq f(i)$ for every $i \in [n]$.
(c) f is determined by A^n.
(d) A^n is a subset of $[n]$ of size at most $\log_{\frac{r}{r-1}} s(\mathcal{H})$.

Proof. Properties (a) and (b) are easy to verify. For the others, write $A^n = \{a_1 < a_2 < \cdots < a_k\}$ and let $a_0 = 0$, $a_{k+1} = n + 1$. Given $i \in [n]$, there must exist $\ell = \ell(i) \in [k + 1]$ with $a_{\ell-1} < i \leq a_\ell$. Denote by $j = j(i)$ the value of the VC-minority function of $\mathcal{H}^{a_{\ell-1}}$ at i. A simple induction on ℓ shows that $A^i = \{a_1, \ldots, a_{\ell-1}\}$, $\mathcal{H}^i = \mathcal{H}^{a_{\ell-1}}$ and $f(i) = j$ when $a_{\ell-1} < i < a_\ell$, while $A^i = \{a_1, \ldots, a_\ell\}$, $\mathcal{H}^i = (\mathcal{H}^{a_{\ell-1}})_{i \to j}$ and $f(i) = j + 1 \pmod{r}$ when $i = a_\ell$. From this we see that for every $i \in [n]$, $A^i, \mathcal{H}^i$ and $f(i)$ are uniquely determined by A^n and i. This implies (c).

From (3) and the above discussion, we get $s(\mathcal{H}^{a_\ell}) \leq \frac{r-1}{r} \cdot s(\mathcal{H}^{a_{\ell-1}})$ for every $1 \leq \ell \leq k$. Together with (a), this implies $1 \leq s(\mathcal{H}^{a_k}) \leq (\frac{r-1}{r})^k \cdot s(\mathcal{H}^{a_0}) = (\frac{r-1}{r})^k \cdot s(\mathcal{H})$, which results in $k \leq \log_{\frac{r}{r-1}} s(\mathcal{H})$, as desired. $\square$

From Lemma 3.1, we quickly deduce Theorem 1.2.

Proof of Theorem 1.2. Let $\mathcal{H} \subseteq \{1, \ldots, r, \star\}^{[n]}$ be a partial concept class with $\text{VC}(\mathcal{H}) \leq d$. Run the algorithm for each $h \in \mathcal{H}$, and let $\mathcal{F}$ be the family consisting of all outputs f. By property (b) in Lemma 3.1, $\mathcal{F}$ is a net of $\mathcal{H}$. Finally, from properties (c) and (d) in Lemma 3.1, we find

$$|\mathcal{F}| \leq \binom{n}{\leq \log_{\frac{r}{r-1}} s(\mathcal{H})} \leq \binom{n}{\leq \log_{\frac{r}{r-1}} \binom{n}{\leq d}},$$

where the second inequality holds since $s(\mathcal{H}) \leq \binom{n}{\leq \text{VC}(\mathcal{H})} \leq \binom{n}{\leq d}$. This completes our proof. □

4. Applications to dynamical systems

As mentioned earlier, in this section we study the complexity function $p^*_{X,\mathcal{U}}(n)$ when $\mathcal{U}$ is a clopen partition of X. We derive the main result of this section, namely Theorem 1.6, from Theorem 1.1. In Section 4.2, we use Theorem 1.6 to study the complexity of symbolic dynamics, improving another result of Huang and Ye [17].

We begin with an auxiliary lemma that is used to prove the statements given later in the subsections.

Lemma 4.1. *Let (X,T) be a TDS and let $\mathcal{U} = \{U_1, \ldots, U_r\}$ be a clopen partition of X. Then, for any subset $S \subseteq \mathbb{N} \cup \{0\}$,*

$$N\left(\bigvee_{i \in S} T^{-i}\mathcal{U}\right) = \#\left\{f \in [r]^S : \bigcap_{i \in S} T^{-i}U_{f(i)} \neq \emptyset\right\}.$$

Proof. Since $\mathcal{U}$ is a partition of X, the sets in the open cover $\bigvee_{i \in S} T^{-i}\mathcal{U}$ are pairwise disjoint. Hence the minimum size of a subcover of $\bigvee_{i \in S} T^{-i}\mathcal{U}$ is exactly the number of nonempty sets in $\bigvee_{i \in S} T^{-i}\mathcal{U}$. Moreover, every set in $\bigvee_{i \in S} T^{-i}\mathcal{U}$ is of the form $\bigcap_{i \in S} T^{-i}U_{f(i)}$ for some function $f \in \{1, \ldots, r\}^S$. Therefore, $N\left(\bigvee_{i \in S} T^{-i}\mathcal{U}\right) = \#\left\{f \in [r]^S : \bigcap_{i \in S} T^{-i}U_{f(i)} \neq \emptyset\right\}$. □

4.1. An application in topological dynamical systems

Given a total class $\mathcal{H} \subseteq \{1, \ldots, r\}^Z$ and a positive integer k, we say a subset $S \subseteq Z$ is *k-Natarajan shattered* by $\mathcal{H}$ if $\mathcal{H}|_S$ contains a subclass of the form $\prod_{i \in S} Y_i$, where Y_i is a k-element subset of $\{1, \ldots, r\}$ for each $i \in S$. Then, $\dim_k(\mathcal{H})$ equals the maximum size of a k-Natarajan shattered set.

Proof of Theorem 1.6 (assuming Theorem 1.1). Let $\mathcal{U} = \{U_1, \ldots, U_r\}$ be a clopen partition of X. For each $n \in \mathbb{N}$, let V_n be a set of n nonnegative integers such that $N\left(\bigvee_{i \in V_n} T^{-i}\mathcal{U}\right) = p^*_{X,\mathcal{U}}(n)$. Define

$$\mathcal{H}_n = \Big\{ h \in [r]^{V_n} : \bigcap_{i \in V_n} T^{-i} U_{h(i)} \neq \emptyset \Big\}.$$

Then, by Lemma 4.1, $N\left(\bigvee_{i \in V_n} T^{-i}\mathcal{U}\right) = |\mathcal{H}_n|$, resulting in

$$p^*_{X,\mathcal{U}}(n) = |\mathcal{H}_n|.$$

Let ℓ be the maximum integer such that $\limsup_{n \to \infty} \dim_\ell(\mathcal{H}_n) = \infty$. Then we have $\ell \in \{1, \ldots, r\}$. We consider the upper bound on $p^*_{X,\mathcal{U}}(n)$ first. If $\ell = r$, then clearly $p^*_{X,\mathcal{U}}(n) \leq |\mathcal{U}|^n = r^n = \ell^n$. Now suppose $\ell \leq r - 1$, then we have $\dim_{\ell+1}(\mathcal{H}_n) = O(1)$. Since $\mathcal{H}_n \subseteq \{1, \ldots, r\}^{V_n}$, $|V_n| = n$ and $\dim_{\ell+1}(\mathcal{H}_n) = O(1)$, we derive from Theorem 1.1 that $|\mathcal{H}_n| \leq n^{O(1)}\ell^n$, and so

$$p^*_{X,\mathcal{U}}(n) = |\mathcal{H}_n| \leq n^{O(1)}\ell^n \quad \text{when} \ \ell \leq r - 1.$$

We proceed to lower bound $p^*_{X,\mathcal{U}}(n)$. Because $\limsup_{n \to \infty} \dim_\ell(\mathcal{H}_n) = \infty$, there exist a sequence $\{n(k)\}_{k \in \mathbb{N}}$ of positive integers and a sequence $\{W_k\}_{k \in \mathbb{N}}$ of sets satisfying

(i) W_k is ℓ-Natarajan shattered by $\mathcal{H}_{n(k)}$.
(ii) W_k is a subset of $V_{n(k)}$ of size k;

For every $k \in \mathbb{N}$, we have

$$p^*_{X,\mathcal{U}}(k) \overset{(ii)}{\geq} N\Big(\bigvee_{i \in W_k} T^{-i}\mathcal{U} \Big) = \#\Big\{ f \in [r]^{W_k} : \bigcap_{i \in W_k} T^{-i} U_{f(i)} \neq \emptyset \Big\}$$

$$\geq \big|\mathcal{H}_{n(k)}\big|_{W_k}\big| \overset{(i)}{\geq} \ell^{|W_k|} \overset{(ii)}{=} \ell^k,$$

where the first equality follows from Lemma 4.1, and in the second inequality we used the definition of $\mathcal{H}_{n(k)}$. We therefore get $p^*_{X,\mathcal{U}}(n) = \ell^n$ when $\ell = r$ and

$$\ell^n \leq p^*_{X,\mathcal{U}}(n) \leq n^{O(1)}\ell^n \quad \text{when } \ell \leq r - 1.$$

From this we find $h^*_{\text{top}}(T, \mathcal{U}) = \limsup_{n \to +\infty} \frac{1}{n} \log p^*_{X,\mathcal{U}}(n) = \log \ell$, where $\ell \in \{1, \ldots, r\}$. This completes our proof. $\qquad \square$

4.2. An application in symbolic dynamics

Given an integer $r \geq 2$, we consider the product set $\Omega_r = \{1, \ldots, r\}^{\mathbb{N}}$. Topology on $\{1, \ldots, r\}$ is discrete, and Ω_r is endowed with the product topology. Since $\{1, \ldots, r\}$ is compact and metrizable, so is the product space Ω_r. Concretely, one can equip Ω_r with the metric $d((x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}) = \sum_{n \geq 1} 2^{-n} \mathbf{1}_{x_n \neq y_n}$. Let $T \colon \Omega_r \to \Omega_r$ be the shift $T(x_n)_{n \in \mathbb{N}} := (x_{n+1})_{n \in \mathbb{N}}$. A **subshift** is a closed T-invariant subset of Ω_r.

Consider a subshift $X \subseteq \Omega_r$. We see that (X, T) is a TDS. For $1 \leq i \leq r$, U_i denotes the clopen set $X \cap \{x \in \Omega_r \colon x_1 = i\}$. Then $\mathcal{U}_0 := \{U_1, \ldots, U_r\}$ is a clopen partition of X. Hence we can define

$$p_X^*(n) := p_{X, \mathcal{U}_0}^*(n) \text{ for } n \in \mathbb{N}, \text{ and } h^*(X) := h_{\text{top}}^*(T, \mathcal{U}_0).$$

As a direct application of Theorem 1.6, we get the following result.

Theorem 4.2. *For any subshift (X, T) on r letters, one has $h^*(X) = \log \ell$ for some $\ell \in \{1, \ldots, r\}$. Moreover, one of the following alternatives holds.*

(a) *If $\ell = r$, then $p_X^*(n) = r^n$ for all $n \in \mathbb{N}$.*
(b) *If $\ell \leq r - 1$, then there exists a constant $c > 0$ such that $\ell^n \leq p_X^*(n) \leq n^c \ell^n$ for all $n \geq 2$.*

We remark that for $\ell \leq r - 1$, a much weaker bound of the form $p_X^*(n) \leq n^{O(\log n)} \ell^n$ was obtained by Huang and Ye [17, Theorem 5.5].

For the rest of this section, we give a more explicit expression for $p_X^*(n)$. Huang and Ye [17] claimed, without a proof, that

$$(4) \qquad p_X^*(n) = p_{X, \mathcal{U}_0}^*(n) = \max_{0 \leq s_1 < \ldots < s_n} \# \{x_{1+s_1} \ldots x_{1+s_n} \colon x \in X\}.$$

For the reader's convenience, we provide a proof of this simple fact. Let $s_1 < \ldots < s_n$ be any sequence of n nonnegative integers. Since $\mathcal{U}_0$ is a clopen partition of X, it follows from Lemma 4.1 that

$$N\left(\bigvee_{i \in [n]} T^{-s_i} \mathcal{U} \right)$$

$$= \#\left\{ f \in [r]^{[n]} \colon \bigcap_{i \in [n]} T^{-s_i} U_{f(i)} \neq \emptyset \right\}$$

$$= \#\left\{ f \in [r]^{[n]} \colon \text{ there is } x \in X \text{ with } x \in \bigcap_{i \in [n]} T^{-s_i} U_{f(i)} \right\}$$

$$= \#\left\{ f \in [r]^{[n]} \colon \text{ there is } x \in X \text{ with } T^{s_i} x \in U_{f(i)} \text{ for every } 1 \leq i \leq n \right\}$$

$$= \#\Big\{f \in [r]^{[n]} : \text{ there is } x \in X \text{ with } x_{1+s_1} = f(1), \dots, x_{1+s_n} = f(n)\Big\}$$
$$= \# \{x_{1+s_1} \dots x_{1+s_n} : x \in X\} \,.$$

Taking the maximum over all sequences $s_1 < \dots < s_n$ yields (4).

In a special case when X is the closure of the orbit of a word $a \in \Omega_r$ under the shift map T, one can further simplify (4). Indeed, for every $x \in X = \overline{\{a, Ta, T^2a, \dots\}}$ and for every $n_0 \in \mathbb{N}$, there exists $y \in \{a, Ta, T^2a, \dots\}$ such that $d(x, y) < 2^{-n_0}$. Suppose $y = T^{m-1}a$ for some $m \in \mathbb{N}$. Then the condition $d(x, y) < 2^{-n_0}$ forces $x_i = y_i = a_{m+i}$ for all $i \le n_0$. Together with (4), this implies

$$p_X^*(n) = \max_{0 \le s_1 < \dots < s_n} \# \{a_{m+s_1} \dots a_{m+s_n} : m \in \mathbb{N}\} \,.$$

The right-hand side of the equation is also called the **maximal pattern complexity** of a. Actually, this concept was introduced by Kamae and Zamboni [18, 19], and was the inspiration behind the work of Huang and Ye [17] on the maximal pattern entropy.

5. Concluding remarks

In this paper we study two variants of the VC dimension and their connections among dynamical systems, combinatorics and theoretical computer science. One intriguing question that deserves further investigation is the tightness of the bounds in Theorem 1.2 for $r \ge 3$. Theorem 1.3 tells us that these bounds are essentially tight for $r = 2$. Some of the arguments in the proof of Theorem 1.3 do generalise to larger r. Let us recall some notions from hypergraph theory. Let $G = (V, E)$ be an r-graph. The r-partition number of G, denoted by $f_r(G)$, is the minimum number of complete r-partite r-graphs needed to partition the edge set of G. The chromatic number $\chi(G)$ of G is the minimum k for which there exists a coloring $c \colon V \to [k]$ such that every edge $e \in E$ contains two vertices u, v with $c(u) \ne c(v)$. One can easily extend Lemma A.2 to r-graphs as follows.

Proposition 5.1. *For every r-graph G with $f_r(G) = n$, there exists a partial concept class $\mathcal{H} \subseteq \{1, \dots, r, \star\}^{[n]}$ with $\mathrm{VC}(\mathcal{H}) \le 1$ and $C(\mathcal{H}) \ge \chi(G)$.*

What is missing is an analogue of Theorem A.1 for r-graphs.

Question 5.2 (A hypergraph Alon–Saks–Seymour problem)**.** *Let $r \ge 3$. For every n, is there an r-graph G such that $f_r(G) = n$ and $\chi(G) \ge n^{c_r \log n}$, where $c_r > 0$ is a constant depending only on r?*

Appendix A. Proof of Theorem 1.3

In this section we present a proof of Theorem 1.3, due to Alon, Hanneke, Holzman and Moran [2]. The proof exploits a recent breakthrough in communication complexity and graph theory, namely Theorem A.1 below, which provides a near-optimal solution to the Alon–Saks–Seymour problem in graph theory (for background on this problem, see the survey by Bousquet, Lagoutte and Thomassé [6]). Let $G = (V, E)$ be a simple graph. Recall that the chromatic number of G, denoted by $\chi(G)$, is the minimum k for which there exists a labelling $c \colon V \to [k]$ such that every edge $\{u, v\} \in E$ satisfies $c(u) \neq c(v)$. The **biclique partition number** of G, denoted $\mathrm{bp}(G)$, is the minimum number of bicliques (i.e. complete bipartite graphs) needed to partition the edges of G. The following result follows from a recent line of breakthroughs by Göös [13]; Göös, Lovett, Meka, Watson and Zuckerman [14]; Balodis, Ben-David, Göös, Jain and Kothari [3]:

Theorem A.1 ([3])**.** *For every positive integer n there exists a graph G with* $\mathrm{bp}(G) = n$ *and*

$$\chi(G) \geq n^{(\log n)^{1-o(1)}},$$

where the term $o(1)$ tends to zero as n goes to infinity.

The following result allows us to use the graph G promised by Theorem A.1 to construct a partial concept class $\mathcal{H}$ with small VC dimension and large covering number.

Lemma A.2. *For every graph G with $\mathrm{bp}(G) = n$, there exists a partial concept class $\mathcal{H} \subseteq \{1, 2, \star\}^{[n]}$ with $\mathrm{VC}(\mathcal{H}) \leq 1$ and $C(\mathcal{H}) \geq \chi(G)$.*

We remark that our proof also gives that $|\mathcal{H}| = |V(G)|$. Before proving Lemma A.2, let us deduce Theorem 1.3 from it.

Proof of Theorem 1.3 assuming Lemma A.2. Let G be the graph given by Theorem A.1. Then we have $\mathrm{bp}(G) = n$ and $\chi(G) \geq n^{(\log n)^{1-o(1)}}$. By Lemma A.2, there exists a partial concept class $\mathcal{H} \subseteq \{1, 2, \star\}^{[n]}$ with $\mathrm{VC}(\mathcal{H}) \leq 1$ and $C(\mathcal{H}) \geq \chi(G) \geq n^{(\log n)^{1-o(1)}}$, proving Theorem 1.3. □

To complete the proof of Theorem 1.3, it remains to prove Lemma A.2, the task we now pursue.

Proof of Lemma A.2. Suppose we have a partition of $E(G)$ as disjoint union of $\mathcal{B}(L_i, R_i)$ for $i \in [n]$, where $\mathcal{B}(L_i, R_i)$ denotes the edge set of the complete

bipartite graph with parts L_i and R_i. For each $v \in V(G)$, let h_v be a partial function in $\{1, 2, \star\}^{[n]}$ given by

$$h_v(i) = \begin{cases} 1 & \text{if } v \in L_i \\ 2 & \text{if } v \in R_i \\ \star & \text{otherwise.} \end{cases}$$

Set $\mathcal{H} = \{h_v \colon v \in V(G)\}$. We have to show that the partial concept class $\mathcal{H} \subseteq \{1, 2, \star\}^{[n]}$ satisfies $\mathrm{VC}(\mathcal{H}) \leq 1$ and $C(\mathcal{H}) \geq \chi(G)$.

Suppose for the contrary that $\mathrm{VC}(\mathcal{H}) \geq 2$. Then there must exist two distinct coordinates $i, j \in [n]$ such that $\{i, j\}$ is shattered by $\mathcal{H}$. In particular, we can find two vertices $u, v \in V(G)$ with $(h_u(i), h_u(j)) = (1, 1)$ and $(h_v(i), h_v(j)) = (2, 2)$. From the definitions of h_u and h_v, we get $u \in L_i \cap L_j$ and $v \in R_i \cap R_j$. Hence $\{u, v\}$ is covered by both $\mathcal{B}(L_i, R_i)$ and $\mathcal{B}(L_j, R_j)$, which contradicts the assumption that $\cup_{i=1}^n \mathcal{B}(L_i, R_i)$ is an edge partition of G.

It remains to show that $C(\mathcal{H}) \geq \chi(G)$. Indeed, from the definition of $C(\mathcal{H})$, there exists a family $\mathcal{C} \subseteq \{1, 2\}^{[n]}$ that satisfies

(i) $|\mathcal{C}| = C(\mathcal{H})$;
(ii) for each vertex $v \in V$ there is a total function $c_v \in \mathcal{C}$ such that $h_v(i) \neq c_v(i)$ for all $i \in [n]$.

Assign to each vertex $v \in V$ the color $c_v \in \mathcal{C}$. We claim that this is a proper coloring, and so $C(\mathcal{H}) = |\mathcal{C}| \geq \chi(G)$, as desired. Indeed, let $\{u, v\}$ be any edge in G. Since $\cup_{i=1}^n \mathcal{B}(L_i, R_i)$ is an edge partition of G, $\{u, v\} \in \mathcal{B}(L_i, R_i)$ for some $i \in [n]$. Let $u \in L_i$ and $v \in R_i$. By the definitions of h_u and h_v, we thus obtain $h_u(i) = 1$ and $h_v(i) = 2$. It then follows from (ii) that $c_u(i) = 2$ and $c_v(i) = 1$. We conclude that u and v are assigned different colors, completing our proof. $\qquad\square$

Appendix B. Proof of Theorem 1.7

In this section we present Huang and Ye's proof [17] of Theorem 1.7, with several simplifications. Let (X, T) be a TDS. For a finite open cover $\mathcal{U}$ of X, define $L(\mathcal{U}) = \limsup\limits_{n \to +\infty} \frac{\log p_{X, \mathcal{U}}^*(n)}{\log^2(n+1)}$. Then one can restate Theorem 1.7 as follows.

Theorem B.1. *If (X, T) is a null TDS, then for every finite open cover $\mathcal{U}$*

$$L(\mathcal{U}) = O(1).$$

Throughout the section, A^c denotes the complement $X \setminus A$ of A.

Proof of Theorem B.1. Suppose for the contrary that $L(\mathcal{U}) = \infty$. By Lemma B.3 below, there is an open cover $\mathcal{V} = \{V_1, V_2\}$ of size 2 such that $L(\mathcal{V}) = \infty$. Then, for each $d \in \mathbb{N}$, there exists a finite set $S \subset \mathbb{N} \cup \{0\}$ with $N\big(\bigvee_{i \in S} T^{-i}\mathcal{V}\big) > (|S| + 1)^{4d \log(|S|+1)}$. For each $x \in X$, let h_x be a partial function in $\{1, 2, \star\}^S$ defined as

$$h_x(i) = \begin{cases} 1 & \text{if } x \in T^{-i}(V_1^c) \\ 2 & \text{if } x \in T^{-i}(V_2^c) \\ \star & \text{otherwise.} \end{cases}$$

Since $\{V_1, V_2\}$ is a cover of X, V_1^c and V_2^c are disjoint, and hence h_x is well-defined. Consider the partial concept class $\mathcal{H} := \{h_x : x \in X\}$.

Claim B.2. $C(\mathcal{H}) = N\big(\bigvee_{i \in S} T^{-i}\mathcal{V}\big)$.

Proof. Let $\mathcal{F} \subseteq \{1, 2\}^S$ be a total class. We can infer from the definitions of $\mathcal{H}$ and h_x that

$\mathcal{F}$ is a net of $\mathcal{H}$

$\iff$　for each $h_x \in \mathcal{H}$ there is $f \in \mathcal{F}$ such that $h_x(i) \neq f(i)$ for all $i \in S$

$\iff$　for each $x \in X$ there is $f \in \mathcal{F}$ such that $x \notin T^{-i}(V_{f(i)}^c)$ for all $i \in S$

$\iff$　for each $x \in X$ there is $f \in \mathcal{F}$ such that $x \in T^{-i}V_{f(i)}$ for all $i \in S$

$\iff$　for each $x \in X$ there is $f \in \mathcal{F}$ such that $x \in \bigcap_{i \in S} T^{-i}V_{f(i)}$

$\iff$　$\big\{\bigcap_{i \in S} T^{-i}V_{f(i)} : f \in \mathcal{F}\big\}$ is a cover of X.

Letting $\mathcal{F}$ be a net of minimum size of $\mathcal{H}$, this yields the lower bound

$$C(\mathcal{H}) = |\mathcal{F}| \geq |\{\bigcap_{i \in S} T^{-i}V_{f(i)} : f \in \mathcal{F}\}| \geq N\big(\bigvee_{i \in S} T^{-i}\mathcal{V}\big).$$

For the upper bound, let $\mathcal{W}$ be a minimum subcover of $\bigvee_{i \in S} T^{-i}\mathcal{V}$. Then $\mathcal{W}$ can be written as $\mathcal{W} = \{\bigcap_{i \in S} T^{-i}V_{f(i)} : f \in \mathcal{F}\}$, where $\mathcal{F} \subseteq \{1, 2\}^S$ is a total class with $|\mathcal{F}| = |\mathcal{W}|$. As $\mathcal{W}$ is a cover of X, we find that $\mathcal{F}$ is a net of $\mathcal{H}$. Thus $C(\mathcal{H}) \leq |\mathcal{F}| = |\mathcal{W}| = N\big(\bigvee_{i \in S} T^{-i}\mathcal{V}\big)$. We are done. $\square$

From Claim B.2 we obtain $C(\mathcal{H}) = N\big(\bigvee_{i \in S} T^{-i}\mathcal{V}\big) \geq (|S| + 1)^{4d \log(|S|+1)}$. Thus, by the remark after Theorem 1.2, there exists a size-d subset $W \subseteq S$

with $\mathcal{H}|_W \supseteq \{1,2\}^W$. Let $x \in X$ be an element such that $h_x|_W$ lies in $\{1,2\}^W$. Suppose x is contained in $\bigcap_{i \in W} T^{-i} V_{f(i)}$ for some $f \in \{1,2\}^W$, then we must have $h_x(i) \neq f(i)$ for all $i \in W$, and so f is uniquely determined by $h_x|_W$. Therefore, we have

$$N\Big(\bigvee_{i \in W} T^{-i} \mathcal{V} \Big) \geq |\mathcal{H}|_W \cap \{1,2\}^W| = 2^{|W|}.$$

Letting $|W| = d \to \infty$ yields $h_{\text{top}}^*(T, \mathcal{V}) \geq 1 > 0$, a contradiction. $\square$

The rest of this section is devoted to establishing the following lemma that was used in the proof Theorem B.1.

Lemma B.3. *Let (X, T) be a TDS. Suppose that $\mathcal{U}$ is a finite open cover of X with $L(\mathcal{U}) = +\infty$. Then there is an open cover $\mathcal{V} = \{V_1, V_2\}$ of size two with $L(\mathcal{V}) = +\infty$.*

As in [17], we follow the arguments of Blanchard [4]. We shall make use of basic properties of the function L.

Proposition B.4. *Let (X, T) be a TDS. Then the following properties hold.*

(i) *(**Monotone**) If $\mathcal{V}$ is a refinement of $\mathcal{U}$, then $L(\mathcal{U}) \leq L(\mathcal{V})$.*
(ii) *(**Subadditive**) If $\mathcal{U}, \mathcal{V}$ are open covers of X, then $L(\mathcal{U} \vee \mathcal{V}) \leq L(\mathcal{U}) + L(\mathcal{V})$.*

Proof. For (i), consider any set S of nonnegative integers. Since $\mathcal{V}$ is a refinement of $\mathcal{U}$, $\bigvee_{i \in S} T^{-i} \mathcal{V}$ is a refinement of $\bigvee_{i \in S} T^{-i} \mathcal{U}$, and so $N(\bigvee_{i \in S} T^{-i} \mathcal{U}) \leq N(\bigvee_{i \in S} T^{-i} \mathcal{V})$. It follows that $p_{X,\mathcal{U}}^*(n) \leq p_{X,\mathcal{V}}^*(n)$ for every $n \in \mathbb{N}$, which implies $L(\mathcal{U}) \leq L(\mathcal{V})$.

For every $n \in \mathbb{N}$ we have

$$\begin{aligned}
p_{X,\mathcal{U} \vee \mathcal{V}}^*(n) &= \max_S N\Big(\bigvee_{i \in S} T^{-i}(\mathcal{U} \vee \mathcal{V}) \Big) \\
&= \max_S N\Big(\bigvee_{i \in S} (T^{-i}\mathcal{U} \vee T^{-i}\mathcal{V}) \Big) \\
&\leq \max_S N\Big(\bigvee_{i \in S} T^{-i}\mathcal{U} \Big) \cdot N\Big(\bigvee_{i \in S} T^{-i}\mathcal{V} \Big) \\
&\leq \max_S N\Big(\bigvee_{i \in S} T^{-i}\mathcal{U} \Big) \cdot \max_T N\Big(\bigvee_{i \in T} T^{-i}\mathcal{V} \Big) \\
&= p_{X,\mathcal{U}}^*(n) \cdot p_{X,\mathcal{V}}^*(n),
\end{aligned}$$

where the maximums are taken over all size-n subsets $S, T \subset \mathbb{N} \cup \{0\}$. Taking the logarithm and then dividing by $\log^2(n+1)$, we obtain $L(\mathcal{U} \vee \mathcal{V}) \leq L(\mathcal{U}) + L(\mathcal{V})$.　□

Proof of Lemma B.3. Let $\mathcal{U} = \{U_1, \ldots, U_k\}$. We first observe that $U_1 \neq X$. Suppose otherwise that $U_1 = X$, then for every subset $S \subseteq \mathbb{N} \cup \{0\}$ we have $X \in \bigvee_{i \in S} T^{-i}\mathcal{U}$, and so the open cover $\bigvee_{i \in S} T^{-i}\mathcal{U}$ has a subcover of size 1, which implies $L(\mathcal{U}) = 0$, a contradiction.

Given a subset $A \subseteq X$, we shall use $\mathrm{diam}(A)$ to denote the diameter of A. We now inductively construct a sequence of closed sets $(A_n)_{n \geq 0}$ with the following three properties

(P1)　$U_1^c = A_0 \supseteq A_1 \cdots$;
(P2)　$\mathrm{diam}(A_n) \leq 2^{-n}$ for every $n \geq 1$;
(P3)　$L(\mathcal{U}_n) = +\infty$ for every $n \geq 0$, where $\mathcal{U}_n := \{A_n^c, U_2, \ldots, U_k\}$.

Clearly, $A_0 = U_1^c$ is a closed set that satisfies (P1)–(P3). Suppose that we have already constructed A_n. Since A_n is a closed subset of the compact set X, A_n is also compact, and so we can cover A_n by a finite number of closed balls $B_1, \ldots, B_\ell$ of radius 2^{-n-2}. For $1 \leq i \leq \ell$, let $A_{n+1,i} = A_n \cap B_i$ and $\mathcal{U}_{n+1,i} = \{A_{n+1,i}^c, U_2, \ldots, U_k\}$. Since $\mathcal{U}_n = \{A_n^c, U_2, \ldots, U_k\}$ is an open cover of X and $A_{n+1,i}^c$ is an open set containing A_n^c, we find that $\mathcal{U}_{n+1,i}$ is also an open cover of X. The definition of $A_{n+1,i}$ gives $A_{n+1,i}^c = (A_n \cap B_i)^c = A_n^c \cup B_i^c$. Hence

$$\bigcap_{i=1}^{\ell} A_{n+1,i}^c = \bigcap_{i=1}^{\ell} (A_n^c \cup B_i^c) = A_n^c \cup \left(A_n \cap \bigcap_{i=1}^{\ell} B_i^c\right) = A_n^c \cup \emptyset = A_n^c,$$

where the third identity holds since A_n is covered by $B_1, \ldots, B_\ell$. It follows that every set in $\bigvee_{i=1}^{\ell} \mathcal{U}_{n+1,i}$ is contained in some set in $\mathcal{U}_n = \{A_n^c, U_2, \ldots, U_k\}$. In other words, the open cover $\bigvee_{i=1}^{\ell} \mathcal{U}_{n+1,i}$ is a refinement of $\mathcal{U}_n$. Using Proposition B.4, we thus obtain

$$+\infty = L(\mathcal{U}_n) \leq L\left(\bigvee_{i=1}^{\ell} \mathcal{U}_{n+1,i}\right) \leq \sum_{i=1}^{\ell} L(\mathcal{U}_{n+1,i}).$$

So $L(\mathcal{U}_{n+1,i}) = +\infty$ for some $i \in [\ell]$. Set $A_{n+1} = A_{n+1,i}$. As $A_{n+1} = A_n \cap B_i$, A_{n+1} is a closed subset of A_n of diameter $\mathrm{diam}(A_{n+1}) \leq \mathrm{diam}(B_i) \leq 2 \cdot 2^{-n-2} = 2^{-n-1}$. Therefore, A_{n+1} has the desired properties.

From (P3) and the observation at the beginning of the proof, we see that $A_n^c \neq X$, and so $A_n \neq \emptyset$. This, together with (P1), (P3) and the compactness

of X, yields $\bigcap_{n \geq 0} A_n = \{x\}$ for some $x \in X$. Because $x \in A_0 = U_1^c$ and $\mathcal{U} = \{U_1, \ldots, U_k\}$ is a cover of X, there is $\ell \in \{2, \ldots, k\}$ with $x \in U_\ell$. Since U_ℓ is open, there exists $\varepsilon > 0$ such that U_ℓ contains an open ball $B(x, \varepsilon)$ centered at x of radius $\varepsilon > 0$. On the other hand, since $\bigcap_{n \geq 0} A_n = \{x\}$ and $\lim_{n \to \infty} \operatorname{diam}(A_n) = 0$, one has $A_n \subseteq B(x, \varepsilon)$ for n sufficient large. For such an n, let $V_1 = A_n^c$ and $V_2 = U_\ell$. Since $V_1 \cup V_2 \supseteq B(x, \varepsilon)^c \cup B(x, \varepsilon) = X$, $\mathcal{V} = \{V_1, V_2\}$ is a subcover of $\mathcal{U}_n$. Finally, we have

$$L(\mathcal{V}) \geq L(\mathcal{U}_n) = +\infty.$$

This completes our proof. $\qquad\square$

Acknowledgements

References

[1] Roy L. Adler, Alan G. Konheim, and M. Harry McAndrew. Topological entropy. *Transactions of the American Mathematical Society* **114**.2 (1965), 309–319. MR0175106

[2] Noga Alon, Steve Hanneke, Ron Holzman, and Shay Moran. A Theory of PAC Learnability of Partial Concept classes. *Proc. FOCS* (2021), 658–671. MR4399723

[3] Kaspars Balodis, Shalev Ben-David, Mika Göös, Siddhartha Jain, and Robin Kothari. Unambiguous DNFs and Alon–Saks–Seymour. *Proc. FOCS* (2021), 116–124. MR4399674

[4] François Blanchard. A disjointness theorem involving topological entropy. *Bulletin de la Société Mathématique de France* **121**.4 (1993), 565–578. MR1254749

[5] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik–Chervonenkis dimension. Journal of the ACM **36** (1989), 929–965. MR1072253

[6] Nicolas Bousquet, Aurélie Lagoutte, and Stéphan Thomassé. Clique versus independent set. *Eurepan Journal of Combinatorics* **40** (2014), 73–92. MR3191490

[7] Moses Charikar and Chirag Pabbaraju. A characterization of list learnability. *Proc. STOC* (2023), 1713–1726. MR4617501

[8] Bernard Chazelle and Emo Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete and Computational Geometry* **4**(5) (1989), 467–489. MR1014739

[9] Amit Daniely, Michael Schapira and Gal Shahaf. Inapproximability of truthful mechanisms via generalizations of the Vapnik–Chervonenkis dimension. *SIAM Journal on Computing* **47**.1 (2018), 96–120. MR3745005

[10] Guo-Li Ding, Paul Seymour and Peter Winkler. Bounding the vertex cover number of a hypergraph. *Combinatorica* **14**.1 (1994), 23–34. MR1273198

[11] Zoltán Füredi and Attila Sali. Optimal multivalued shattering. *SIAM Journal on Discrete Mathematics* **26**.2 (2012), 737–744. MR2967495

[12] Eli Glasner and Xiangdong Ye. Local entropy theory. *Ergodic Theory and Dynamical Systems* **29**.2 (2009), 321–356. MR2486773

[13] Mika Göös. Lower bounds for clique vs. independent set. In *Proceedings of the 56th Symposium on Foundations of Computer Science (FOCS)*, pages 1066–1076. IEEE, 2015. MR3473357

[14] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing* **45**.5 (2016), 1835–1869. MR3561784

[15] Boris Hasselblatt and Anatole Katok. Principal structures. In *Handbook of Dynamical Systems*, North-Holland, Amsterdam, 1A (2002), 1–203. MR1928518

[16] David Haussler and Emo Welzl. ε-nets and simplex range queries. Discrete and Computational Geometry **2** (1987), 127–151. MR0884223

[17] WEN HUANG AND XIANGDONG YE. Combinatorial lemmas and applications to dynamics. *Advances in Mathematics* **220**.6 (2009), 1689–1716. MR2493178

[18] TETURO KAMAE AND LUCA ZAMBONI. Sequence entropy and the maximal pattern complexity of infinite words. *Ergodic Theory and Dynamical Systems* **22** (2002), 1191–1199. MR1926282

[19] TETURO KAMAE AND LUCA ZAMBONI. Maximal pattern complexity for discrete systems. *Ergodic Theory and Dynamical Systems* **22** (2002), 1201–1214. MR1926283

[20] MARK G. KARPOVSKY AND VITALI D. MILMAN. Coordinate density of sets of vectors. *Discrete Mathematics* **24** (1978), 177–184. MR0522926

[21] ANATOLE KATOK. Fifty years of entropy in dynamics: 1958–2007. In *Frontiers in Entropy Across the Disciplines: Panorama of Entropy: Theory, Computation, and Applications* (2023), 353–404. MR4545797

[22] DAVID KERR AND HANFENG LI. Independence in topological and C^*-dynamics. *Mathematische Annalen* **338**.4 (2007), 869–926. MR2317754

[23] DAVID KERR AND HANFENG LI. Ergodic Theory. Independence and Dichotomies. Springer, Cham, 2016. MR3616077

[24] JIAN LI. Chaos and entropy for interval maps. *Journal of Dynamics and Differential Equations* **23** (2011), 333–52. MR2802890

[25] BALAS K. NATARAJAN. On learning sets and functions. *Machine Learning* **4** (1989), 67–97. MR1137519

[26] NORBERT SAUER. On the density of families of sets. *Journal of Combinatory Theory, Series A* **13** (1972), 145–147. MR0307902

[27] SAHARON SHELAH. A combinatorial problem: Stability and order for models and theories in infinitary language. *Pacific Journal of Mathematics* **41** (1972), 247–261. MR0307903

[28] J. MICHAEL STEELE. Existence of submatrices with all possible columns. *Journal of Combinatorial Theory, Series A* **24** (1978), 84–88. MR0477507

[29] MICHEL TALAGRAND. Sharper Bounds for Gaussian and Empirical Processes. *Annals of Probability* **22**.1 (1994), 28–76. MR1258865

[30] VLADIMIR VAPNIK AND ALEXEY CHERVONENKIS. The uniform convergence of frequencies of the appearance of events to their

probabilities. *Doklady of Russian Academy of Sciences* **181** (1968), 781–783. MR0231431

[31] Vladimir Vapnik and Alexey Chervonenkis. The uniform convergence of relative frequencies of events to their probabilities. *Probability Theory and Its Applications* **16** (1971), 264–280. MR0288823

[32] Roman Vershynin. High-Dimensional Probability. *Cambridge University Press*, 2018 MR3837109

[33] Yini Yang. Some properties of circle maps with zero topological entropy. *Nonlinearity* **34** (2021), 2781–2799. MR4260778

Guorong Gao
School of Mathematics and Statistics
Fuzhou University
Fuzhou
China
E-mail: grgao@fzu.edu.cn

Jie Ma
School of Mathematical Sciences
University of Science and Technology of China
Hefei
China
Yau Mathematical Sciences Center
Tsinghua University
Beijing
China
E-mail: jiema@ustc.edu.cn

Mingyuan Rong
School of Mathematical Sciences
University of Sciences and Technology of China
Hefei
China
E-mail: rong_ming_yuan@mail.ustc.edu.cn

Tuan Tran
School of Mathematical Sciences
University of Sciences and Technology of China
Hefei
China
E-mail: trantuan@ustc.edu.cn