Discrete Applied Mathematics 327 (2023) 185–196

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Primal and dual combinatorial dimensions

Pieter Kleer [a,*], Hans Simon [b,c]

[a] Tilburg University, Tilburg, The Netherlands
[b] Max Planck Institute for Informatics, Saarbrücken, Germany
[c] Department of Mathematics, Ruhr-University Bochum, Germany

ARTICLE INFO

Article history:
Received 23 August 2021
Received in revised form 14 November 2022
Accepted 21 November 2022
Available online xxxx

Keywords:
VC dimension
Pseudo-dimension
Fat-shattering dimension
Dual dimension

ABSTRACT

We give tight bounds on the relation between the primal and dual of various combinatorial dimensions, such as the pseudo-dimension and fat-shattering dimension, for multi-valued function classes. These dimensional notions play an important role in the area of learning theory. We first review some classical results that bound the dual dimension of a function class in terms of its primal, and after that give (almost) matching lower bounds. In particular, we give an appropriate generalization to multi-valued function classes of a well-known bound due to Assouad (1983), that relates the primal and dual VC-dimension of a binary function class.

1. Introduction

The Vapnik–Chervonenkis (VC) dimension [12] is a fundamental combinatorial dimension in learning theory used to characterize the complexity of *learning* a class X consisting of functions $f : Y \to \{0, 1\}$ where X and Y are given (possibly infinite) sets. Informally, the VC-dimension captures how rich or complex a class of functions is. Many extensions of the VC-dimension to multi-valued functions $f : Y \to Z$, for some given $Z \subseteq \mathbb{R}$, have been proposed in the literature, such as the Vapnik-dimension (also known as the uniform pseudo-dimension) [11], the Pollard-dimension (also known as pseudo-dimension) [6,10], and the fat-shattering dimension [7]. All these combinatorial dimensions are formally defined in Section 2.

Every (primal) class of functions can be identified with a *dual class* whose functions are of the form $g_y : X \to Z$ for $y \in Y$ defined by $g_y(f) = f(y)$ for $f \in X$. When interpreting a function class as a matrix A whose rows and columns are indexed by X and Y, respectively, the dual class is simply given by the transpose matrix $A^\top$. The (VC, pseudo-, etc.) dimension of the dual class is defined as the dimension of the matrix A^T.

Assouad [3] showed the following relation between the primal VC-dimension $\mathrm{VC}(A)$ and the dual VC-dimension $\mathrm{VC}^*(A)$:

$$\mathrm{VC}^*(A) \leq 2^{\mathrm{VC}(A)+1} - 1. \tag{1}$$

This has turned out to be a very useful inequality, e.g., in the context of so-called sample compression schemes [9]. In the case that $\mathrm{VC}^*(A)$ is a power of two, this immediately yields $\mathrm{VC}^*(A) \leq 2^{\mathrm{VC}(A)}$. It is known that this bound is tight for all values of $\mathrm{VC}^*(A)$, see, e.g., [8].

* Corresponding author.
E-mail addresses: p.s.kleer@tilburguniversity.edu (P. Kleer), hsimon@mpi-inf.mpg.de (H. Simon).

https://doi.org/10.1016/j.dam.2022.11.010

The purpose of this work is to understand the relation between the primal and dual of combinatorial dimensions for multi-valued function classes, in particular, for multi-valued functions where $Z = \{0, 1, \ldots, k\}$ for $k \in \mathbb{N}$. For the pseudo-dimension, as explained in Section 3, it can be shown that

$$\mathrm{Pdim}^*(A) \le k \cdot \left(2^{\mathrm{Pdim}(A)+1} - 1\right) \quad ,$$

which naturally generalizes Assouad's bound in (1).[1] Moreover, if $\mathrm{Vdim}^*(A)$ is a power of two, we even have

$$\mathrm{Pdim}^*(A) \le k \cdot 2^{\mathrm{Pdim}(A)} \quad . \tag{2}$$

Our first contribution is that the bound in (2) is in fact tight for every value of k provided that $\mathrm{Pdim}(A) \ge 2$ (Theorem 4.2). In case of $\mathrm{Pdim}(A) = 1$, we give an improved bound of $k+2$ (Theorem 4.1), and also show that this is tight (Theorem 4.2). We obtain similar bounds for the fat-shattering dimension (Theorem 4.5).

Remark 1.1. It is sometimes believed that Assouad's bound also holds for combinatorial dimensions other than the VC-dimension, see, e.g., [5]. Our results show that this is, unfortunately, not correct.

Outline. We continue in Section 2 with all the necessary definitions and notations, in particular the formal definitions of all combinatorial dimensions considered in this work. Then, in Section 3, we outline known results regarding the relations between various combinatorial dimensions and their duals. After that, in Section 4, we summarize our results, followed by their proofs in Section 5.

2. Preliminaries

For $k \ge 1$, we set $[k] := \{1, \ldots, k\}$ and $[k]_0 := [k] \cup \{0\}$. Let X and Y be disjoint sets and let $Z \subseteq \mathbb{R}$ be a subset of the reals. Consider a function $A : X \times Y \to Z$. For $x \in X$, we define $A_x : Y \to Z$ by $A_x(y) = A(x, y)$ and refer to A_x as a *row* of A. For $y \in Y$, we define $A_y : X \to Z$ by $A_y(x) = A(x, y)$ and refer to A_y as a *column* of A. The *transpose* of A is defined as the function $A^\top : Y \times X \to Z$ given by $A^\top(y, x) = A(x, y)$. As suggested by this terminology, we view A as a (possibly infinite) matrix with rows indexed by X, columns indexed by Y and with $A^\top$ as its transpose.

A matrix $A : X \times Y \to Z$ with $Z = \{0, 1\}$ is said to be *Boolean*. Let $d \ge 1$ be a positive integer. We denote by $B_d : X \times Y \to \{0, 1\}$ the Boolean matrix which is defined as follows:

1. $X = [2^d]$ and $Y = [d]$.
2. For every function $b : [d] \to \{0, 1\}$, there exists an $x \in [2^d]$ such that, for every $y \in [d]$, we have $B_d(x, y) = b(y)$.

Note that B_d is unique modulo renaming rows and columns.

Definition 2.1 (*Shattered Sets*)**.** Let $A : X \times Y \to Z$, with $Z \subseteq \mathbb{R}$, be a matrix and let $J \subseteq Y$ be a subset of its columns.

1. Let $Z = \{0, 1\}$. We say that J is *VC-shattered* by A if, for every function $b : J \to \{0, 1\}$, there exists an $x \in X$ such that, for every $y \in J$, we have $A(x, y) = b(y)$.
2. We say that J is *P-shattered* by A if there exists a function $\mathbf{t} : J \to \mathbb{R}$ such that the following holds: for every function $b : J \to \{0, 1\}$, there exists an $x \in X$ such that, for every $y \in J$, we have $A(x, y) \ge \mathbf{t}(y)$ iff $b(y) = 1$.
3. Let $\gamma > 0$. We say that J is *P_γ-shattered* by A if there exists a function $\mathbf{t} : J \to \mathbb{R}$ such that the following holds: for every function $b : J \to \{0, 1\}$, there exists an $x \in X$ such that, for every $y \in J$, we have
$$A(x, y) \begin{cases} \ge \mathbf{t}(y) + \gamma & \text{if } b(y) = 1 \\ < \mathbf{t}(y) - \gamma & \text{if } b(y) = 0 \end{cases} .$$
4. We say that J is *V-shattered* by A if there exists a number $t \in \mathbb{R}$ such that the following holds: for every function $b : J \to \{0, 1\}$, there exists an $x \in X$ such that, for every $y \in J$, we have $A(x, y) \ge t$ iff $b(y) = 1$.
5. Let $\gamma > 0$. We say that J is *V_γ-shattered* by A if there exists a number $t \in \mathbb{R}$ such that the following holds: for every function $b : J \to \{0, 1\}$, there exists an $x \in X$ such that, for every $y \in J$, we have
$$A(x, y) \begin{cases} \ge t + \gamma & \text{if } b(y) = 1 \\ < t - \gamma & \text{if } b(y) = 0 \end{cases} .$$

We will refer to $\mathbf{t} : J \to \mathbb{R}$ occurring in the definition of P- and the P_γ-shattered sets as the *thresholds used for shattering* J. Similarly, we will refer to $t \in \mathbb{R}$ occurring in the definition of V- and the V_γ-shattered sets as the *uniform threshold used for shattering* J.

Definition 2.2 (*Combinatorial Dimensions*)**.** Let $A : X \times Y \to Z$ be a matrix. Let $\tau \in \{\mathrm{VC}, P, P_\gamma, V, V_\gamma\}$ be one of the shattering types mentioned in Definition 2.1. The *(primal) τ-dimension* of A is the size of a largest set $J \subseteq Y$ that is τ-shattered by A (resp. ∞ if there exist τ-shatterable sets of unbounded size). The *dual τ-dimension* of A is defined as the τ-dimension of $A^\top$.

[1] We refer to this as a classical result, rather than a contribution of this work.

We use the notations VC(A), Pdim(A), $\mathrm{P}_\gamma(A)$, Vdim(A) and $\mathrm{V}_\gamma(A)$ for the (primal) dimensions of type $\tau =$ VC, P, P_γ, V, V_γ, respectively. Here, VC(A) is the VC-dimension [12], Pdim(A) the pseudo-dimension [6,10], $\mathrm{P}_\gamma(A)$ the fat-shattering dimension [7], Vdim(A) the Vapnik-dimension [11] and $\mathrm{V}_\gamma(A)$ the fat-shattered version of the Vapnik-dimension, see, e.g., [1]. The corresponding dual dimensions are denoted by $\mathrm{VC}^*(A)$, $\mathrm{Pdim}^*(A)$, $P^*_\gamma(A)$, $\mathrm{Vdim}^*(A)$ and $V^*_\gamma(A)$, respectively. We remark that there also exist other dimensional notions, which are not discussed here, such as the *pseudo-rank* [2].

The matrix *obtained by thresholding the columns of* $A : X \times Y \to Z$ *at* $\mathbf{t} : Y \to \mathbb{R}$ is defined as the Boolean matrix $B : X \times Y \to \{0, 1\}$ such that, for all $x \in X$ and $y \in Y$, we have $B(x, y) = 1$ iff $A(x, y) \geq \mathbf{t}(y)$. For $I \subseteq X$ and $J \subseteq Y$, we denote the restriction of A to $I \times J$ by $A_{I,J}$. In other words: $A_{I,J}$ is the submatrix of A whose rows are indexed by I and whose columns are indexed by J. A *witness for the inequality* $\mathrm{Pdim}(A) \geq d$ is defined as a triple $(I, J, \mathbf{t})$ such that the following holds:

1. I is a subset of X of size 2^d, J is a subset of Y of size d and $\mathbf{t} : J \to \mathbb{R}$.
2. Every pattern $b : J \to \{0, 1\}$ occurs in exactly one row of the Boolean matrix obtained by thresholding the columns of $A_{I,J}$ at $\mathbf{t}$, i.e., $A_{I,J}$ equals B_d up to a permutation of its rows.

Remark 2.3. Let $k \geq 1$ be a positive integer. Consider a matrix $A : X \times Y \to [k]_0$. If a set $J \subseteq Y$ can be P-shattered by A with thresholds $\mathbf{t} : J \to \mathbb{R}$, then it can also be P-shattered with (suitably chosen) thresholds $\mathbf{t} : J \to [k]$. An analogous remark applies to V-shattering with a uniform threshold t.

When analyzing the P- or the V-dimension of a matrix with entries in $[k]_0$, we will assume that thresholds are taken from $[k]$ whenever we find that convenient.

3. Known relations

In this section we review some known relations between the combinatorial dimensions defined in Section 2.

3.1. Bounding P- in terms of V-dimension

It follows directly from the definitions that

$$\mathrm{Vdim}(A) \leq \mathrm{Pdim}(A) \ \text{ and } \mathrm{V}_\gamma(A) \leq \mathrm{P}_\gamma(A) \ .$$

This raises the question whether we can bound the P- in terms of the V-dimension (resp. the P_γ in terms of the V_γ-dimension). The gap between Pdim(A) and Vdim(A) cannot be bounded in general, as the following well-known example shows.

Example 3.1. Let X be the set of all monotone[2] functions from [0, 1] to [0, 1], $Y = [0, 1]$ and $A(x, y) = x(y)$ for $x \in X$. Then, as the following arguments show, we have $\mathrm{Vdim}(A) = 1$ and $\mathrm{Pdim}(A) = \infty$:

- Let $y_1 < y_2$ be two arbitrary elements of [0, 1], let t be an arbitrary uniform threshold and observe that no monotone function x can satisfy $x(y_1) \geq t$ and $x(y_2) < t$. Since this holds for any choice of y_1, y_2 and t, there can be no set of size 2 which is V-shattered by A. Hence $\mathrm{Vdim}(A) = 1$.
- Let $J = \{1/k : k \geq 1\}$, let $t_k = 1/k$ and $b_k \in \{0, 1\}$ for all $k \geq 1$. Consider the monotone function x such that, for every k and every $1/k \geq y > 1/(k + 1)$, we have
$$x(y) = \frac{1}{k + 1 - b_k} \ .$$
Then $x(1/k) \geq 1/k = t_k$ iff $b_k = 1$. It follows that $\mathrm{Pdim}(A) = \infty$.

In the sequel, we focus on matrices of the form $A : X \times Y \to [k]_0$. According to the following results of Ben-David et al. [4] (here expressed in our notation), the P- can exceed the V-dimension by factor k, but not by a larger factor[3]:

Theorem 3.2 ([4])**.** *For every matrix $A : X \times Y \to [k]_0$, we have*

$$\mathrm{Pdim}(A) \leq k \cdot \mathrm{Vdim}(A) \ .$$

Theorem 3.3 ([4])**.** *For every $d \geq 1$ and every $k \geq 1$, there exists a matrix $A : X \times Y \to [k]_0$ such that*

$$\mathrm{Vdim}(A) = d \ \text{ and } \ \mathrm{Pdim}(A) = k \cdot d.$$

Alon et al. [1] have bounded P_γ- in terms of the $V_{\gamma/2}$-dimension.

2 A function $f : [0, 1] \to [0, 1]$ is monotone if $f(x) \leq f(y)$ for all $x \leq y$.

3 See [4, Theorem 7-8] and the proof of [4, Theorem 7].

Theorem 3.4 (*[1]*). *For every matrix $A : X \times Y \to [0, 1]$ and every $0 < \gamma \le 1/2$, we have*[4]

$$\mathrm{P}_\gamma(A) \le \left(\left\lceil \frac{1}{\gamma} \right\rceil - 1 \right) \cdot \mathrm{V}_{\gamma/2}(A) \le \left(\left\lceil \frac{1}{\gamma} \right\rceil - 1 \right) \cdot \mathrm{Pdim}(A) \ . \tag{3}$$

Proof. The thresholds $t_1, \dots, t_d$ used for P_γ-shattering $d := \mathrm{P}_\gamma(A)$ many columns of A must belong to the interval $[\gamma, 1-\gamma]$. Any threshold t_i can be rounded to the closest multiple of γ. Denote the latter by $\hat{t}_i$. The inequality (3) becomes now evident from the following observations. First, by using the thresholds $\hat{t}_i$ instead of t_i, the width of shattering may drop from γ to $\gamma/2$ (but not beyond). Second, $\hat{t}_1, \dots, \hat{t}_d$ can take on at most

$$r := \left\lceil \frac{1-2\gamma}{\gamma} \right\rceil + 1 = \left\lceil \frac{1}{\gamma} \right\rceil - 1$$

different values. By the pigeonhole principle, there is some $t \in \{\hat{t}_1, \dots, \hat{t}_d\}$ that can be used for $V_{\gamma/2}$-shattering d/r many points. □

3.2. Bounding dual dimension in terms of its primal

A well-known result due to Assouad [3] already mentioned in Section 1, which we will refer to as *Assouad's bound*, states that one can upper bound the dual VC-dimension in terms of the (primal) VC-dimension as follows:

Theorem 3.5 (*[3]*). *For every matrix $A : X \times Y \to \{0, 1\}$, we have*

$$\mathrm{VC}^*(A) \le 2^{\mathrm{VC}(A)+1} - 1 \ . \tag{4}$$

Note that, under the assumption that $\mathrm{VC}^*(A)$ is a power of two, this means

$$\mathrm{VC}^*(A) \le 2^{\mathrm{VC}(A)} \ . \tag{5}$$

The bound in (5) is known to be tight for every value of $\mathrm{VC}(A)$, see, e.g., [8].

Assouad's bound has the following immediate implications:

$$\log \mathrm{VC}^*(A) < \mathrm{VC}(A) + 1 \ \text{ and } \ \lfloor \log \mathrm{VC}^*(A) \rfloor \le \mathrm{VC}(A) \ . \tag{6}$$

In Appendix we show that the Assouad's bound also holds for $\mathrm{Vdim}(A)$ and $V_\gamma(A)$, based on the notion of *uniform Ψ-dimension* as defined in [1]. These observations are summarized in the following statements.

Corollary 3.6. *For every matrix $A : X \times Y \to [0, 1]$, we have*

$$\mathrm{Vdim}^*(A) \le 2^{\mathrm{Vdim}(A)+1} - 1 \ \text{ and } \ V_\gamma^*(A) \le 2^{\mathrm{V}_\gamma(A)+1} - 1 \ . \tag{7}$$

If $\mathrm{Vdim}^*(A)$, respectively $V_\gamma^*(A)$, is a power of two, this means

$$\mathrm{Vdim}^*(A) \le 2^{\mathrm{Vdim}(A)} \ \text{ and } \ V_\gamma^*(A) \le 2^{\mathrm{V}_\gamma(A)} \ . \tag{8}$$

Combining Theorem 3.2 (applied to $A^\top$) with Corollary 3.6, one can directly obtain the following result:

Theorem 3.7. *For every matrix $A : X \times Y \to [k]_0$, the following holds:*

1. $\mathrm{Pdim}^*(A) \le k \cdot \left(2^{\mathrm{Vdim}(A)+1} - 1\right) \le k \cdot \left(2^{\mathrm{Pdim}(A)+1} - 1\right)$.
2. *If $\mathrm{Vdim}^*(A)$ is a power of two, then $\mathrm{Pdim}^*(A) \le k \cdot 2^{\mathrm{Vdim}(A)} \le k \cdot 2^{\mathrm{Pdim}(A)}$.*

Similarly, combining Theorem 3.4 with Corollary 3.6, one can directly obtain the following result.

Corollary 3.8. *For every matrix $A : X \times Y \to [0, 1]$, the following holds:*

$$P_\gamma^*(A) \le \left(\left\lceil \frac{1}{\gamma} \right\rceil - 1 \right) \cdot \left(2^{\mathrm{V}_{\gamma/2}(A)+1} - 1\right) \le \left(\left\lceil \frac{1}{\gamma} \right\rceil - 1 \right) \cdot \left(2^{\mathrm{Pdim}(A)+1} - 1\right).$$

4. Our results

In this section we describe our new contributions, that complement those mentioned in Section 3. We first discuss results related to the pseudo-dimension. We start with a result showing that the upper bound on $\mathrm{Pdim}^*(A)$ in Theorem 3.7 can be improved by a factor 2 (roughly) for matrices A with $\mathrm{Vdim}(A) = 1$.

[4] In [1], one finds a factor $2\lceil 1/(2\gamma) \rceil$ at the place of factor $\lceil 1/\gamma \rceil$. We find the latter (and slightly smaller) factor preferable because of its simpler form.

Theorem 4.1. *Let $A : X \times Y \to [k]_0$ with $k \geq 1$ be a matrix with* $\mathrm{Vdim}(A) = 1$. *Then* $\mathrm{Pdim}^*(A) \leq k + 2$.

The next result implies that the upper bound on $\mathrm{Pdim}^*(A)$ in the second statement of Theorem 3.7 is tight for matrices with $\mathrm{Vdim}(A) \geq 2$, as well as the upper bound on $\mathrm{Pdim}^*(A)$ in Theorem 4.1 whenever $\mathrm{Vdim}(A) = 1$.

Theorem 4.2. *The following two lower bounds hold:*

1. *For every $d \geq 2$ and every $k \geq 1$, there exists a matrix $A : X \times Y \to [k]_0$ such that*

$$\mathrm{Pdim}(A) = d \ , \ \ \mathrm{Vdim}^*(A) = 2^d \ \ \text{and} \ \ \mathrm{Pdim}^*(A) = k \cdot 2^d.$$

2. *For every $k \geq 1$, there exists a matrix $A : X \times Y \to [k]_0$ such that* $\mathrm{Vdim}(A) = \mathrm{Pdim}(A) = 1$ *and* $\mathrm{Pdim}^*(A) = k + 2$.

In combination with a technical tool defined in Section 5.2, we also obtain the following corollary. It stands in stark contrast to Assouad's bound for the VC-dimension.

Corollary 4.3. *There exists a matrix $A : X \times Y \to [0, 1]$, such that* $\mathrm{Pdim}(A) = 1$ *and* $\mathrm{Pdim}^*(A) = \infty$.

We next move to our results for the fat-shattering dimensions. The first result here implies that the upper bound on $\mathrm{P}_\gamma(A)$ from Theorem 3.4 is tight up to a small constant factor:

Theorem 4.4. *For every $d \geq 1$, there exists a matrix $A : X \times Y \to [0, 1]$ such that* $\mathrm{Vdim}(A) = d$ *and, for all $k \geq 1$,*

$$\mathrm{P}_{1/(2k)}(A) \geq k \cdot d \ .$$

Finally, our last result implies that the bound on $P^*_\gamma(A)$ from Corollary 3.8 is tight up to a small constant factor.

Theorem 4.5. *The following two lower bounds hold:*

1. *For every $d \geq 2$, there exists a matrix $A : X \times Y \to [0, 1]$ such that* $\mathrm{Pdim}(A) = d$ *and, for all $k \geq 1$,*

$$P^*_{1/(2k)}(A) \ \geq \ k \cdot 2^d \ .$$

2. *There exists a matrix $A : X \times Y \to [0, 1]$ such that* $\mathrm{Pdim}(A) = 1$ *and, for all $k \geq 2$,*

$$P^*_{1/(2k)}(A) \ \geq \ k + 2 \ .$$

5. Proofs

Section 5.1 is devoted to the proof of Theorem 4.1. In Section 5.2, we make some considerations which will allow for an easier presentation of our lower bound constructions, that are given in Sections 5.3 and 5.4.

5.1. Proof of Theorem 4.1

For the case of binary functions ($k = 1$), the assertion of the theorem collapses to the claim that $\mathrm{VC}^*(A) \leq 3$ for every Boolean matrix A with $\mathrm{Vdim}(A) = 1$. This is an immediate consequence of (4). Suppose now that $k \geq 2$. It suffices to show that $\mathrm{Pdim}^*(A) \geq k + 3$ implies that $\mathrm{Vdim}(A) \geq 2$ (i.e., we give a proof by contradiction). Pick a witness $(I, J, \mathbf{t})$ for $\mathrm{Pdim}^*(A) \geq k + 3$. More concretely (using Remark 2.3):

- $I = \{x_1, \ldots, x_{k+3}\}$, $J \subseteq Y$ with $|J| = 2^{k+3}$ and $\mathbf{t} : I \to [k]$, say $\mathbf{t}(x_i) = t_i$.
- The matrix obtained by thresholding the rows of $A_{I,J}$ at $\mathbf{t}$ equals $B^\top_{k+3}$.

We may assume that, after renumbering the rows appropriately, one has $t_1 \leq \cdots \leq t_{k+3}$. We decompose I (and hence the rows of $A_{I,J}$) into maximal blocks such that $\mathbf{t}$ assigns the same threshold to every row index of the same block. While I is P-shattered by A, every block in I has a uniform threshold and is therefore even V-shattered by A. Since any threshold t_i is taken from $[k]$, the total number k' of blocks is bounded by k. A block that is different from the first and from the last block is said to be an *inner block*. We proceed by case analysis (we argue afterwards why at least one of the cases occurs).

Case 1: One of the blocks of I is of size at least 4.

Since, as mentioned above already, every block has a uniform threshold, it follows that $\mathrm{Vdim}^*(A) \geq 4$. An application of (6) yields $\mathrm{Vdim}(A) \geq \log \mathrm{Vdim}^*(A) \geq 2$.

Case 2: The first or the last block of I is of size at least 3.

For reasons of symmetry, we may assume that the first block contains 3 rows. Consider the following (4×2)-submatrix of $B_{k+3}^{\top}$:

$$\begin{array}{cc} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ \hline 1 & 1 \end{array}$$

The first three rows are taken from the first block and the last row is taken from the last block. The separation line between the third and the last row is only intended to illustrate the transition from one block to another. Remember that the rows of the first block are thresholded at t_1 while the rows of the last block are thresholded at $t_{k'} > t_1$. Hence, if we threshold *all rows* (or *all columns*) of $A_{I,J}$ at t_1, then the above submatrix of $B_{k+3}^{\top}$ will remain unchanged. Since this submatrix equals B_2, we may conclude that $\mathrm{Vdim}(A) \geq 2$.

Case 3: One of the inner blocks of I is of size at least 2, say block b.

The argument is similar to that given in Case 2. The relevant submatrix of $B_{k+3}^{\top}$ (with one row of the first block, two rows of block b, one row of the last block and two separation lines inbetween) now looks as follows:

$$\begin{array}{cc} 0 & 0 \\ \hline 0 & 1 \\ 1 & 0 \\ \hline 1 & 1 \end{array}$$

Since $t_1 < t_b < t_{k'}$, thresholding all rows (or all columns) of $A_{I,J}$ at t_b will leave the above submatrix of $B_{k+3}^{\top}$ unchanged. We may conclude that $\mathrm{Vdim}(A) \geq 2$.

Since $A_{I,J}$ has $k+3$ rows (with $k \geq 2$), it can be argued that one of the three above cases must occur. Suppose first that $k = 2$. Then there are at most 2 blocks and 5 rows. It follows that the first or the last block contains at least 3 rows. Suppose now that $k \geq 3$. If the first or last block has three or more rows then Case 2 occurs. Otherwise, if the first and the last block contain at most two rows, respectively, then at least $k-1$ rows are left for the $k'-2 \leq k-2$ inner blocks. By the pigeonhole principle, there must be an inner block with at least two rows. This completes the proof of Theorem 4.1. □

5.2. Preparatory considerations for the remaining proofs

Consider again the Boolean matrix B_d with d columns and 2^d rows that had been defined in Section 2. It is evident that B_d, or any matrix which equals B_d up to a permutation of its rows, satisfies the following conditions:

(i) **Distinctness Condition:** The rows of B_d are pairwise distinct.
(ii) **General Balance Condition:** For any $k \in [d]$, any choice of k distinct columns of B_d and any pattern $\mathbf{b} \in \{0, 1\}^k$, there are exactly 2^{d-k} rows of B_d which realize the pattern $\mathbf{b}$ within the chosen columns.

The general balance condition implies the following:

(iii) **1st Balance Condition:** Each column of B_d has as many zeros as ones.
(iv) **2nd Balance Condition:** For any two distinct columns of B_d, any pattern from $\{0, 1\}^2$ is realized within these columns by the same number of rows.

Remark 5.1 (*Proof Templates*). Consider a matrix $A : X \times Y \to [k]_0$. The following template for proving assertions like $\mathrm{Pdim}(A) \leq d$ will prove itself quite useful.

- Assume for contradiction that $\mathrm{Pdim}(A) \geq d + 1$.
- Pick a witness $(I, J, \mathbf{t})$ for this inequality.
- Exploit the fact that the matrix B obtained by thresholding the columns of $A_{I,J}$ at $\mathbf{t}$ must be equal to B_{d+1} (up to a permutation of its rows).
- Prove that B violates one of the conditions that B_{d+1} must satisfy.

Sometimes the following (slightly simpler) template can be used instead:

- Take a fixed but arbitrary function $\mathbf{t} : Y \to [k]$.
- Let B be the matrix obtained by thresholding the columns of A at $\mathbf{t}$.
- Show that no more than d columns of B have at least 2^d zeros and at least 2^d ones.

This also shows that $\mathrm{Pdim}(A) \leq d$ because no submatrix of B with $d + 1$ columns and 2^{d+1} rows has a chance to satisfy the first balance condition.

We next introduce matrices that, though not being Boolean, are close relatives of the matrix B_d.

$$
B_4 = \left[\begin{array}{c|cc|c}
0&0&0&0\\
0&0&0&1\\
0&0&1&0\\
0&0&1&1\\
0&1&0&0\\
0&1&0&1\\
0&1&1&0\\
0&1&1&1\\
1&0&0&0\\
1&0&0&1\\
1&0&1&0\\
1&0&1&1\\
1&1&0&0\\
1&1&0&1\\
1&1&1&0\\
1&1&1&1
\end{array}\right]
\qquad
A = \left[\begin{array}{c|cc|c}
0&1&1&2\\
0&1&1&3\\
0&1&2&2\\
0&1&2&3\\
0&2&1&2\\
0&2&1&3\\
0&2&2&2\\
0&2&2&3\\
1&1&1&2\\
1&1&1&3\\
1&1&2&2\\
1&1&2&3\\
1&2&1&2\\
1&2&1&3\\
1&2&2&2\\
1&2&2&3
\end{array}\right]
\qquad
B = \left[\begin{array}{c|cc|c}
0&0&0&1\\
0&0&0&1\\
0&0&1&1\\
0&0&1&1\\
0&1&0&1\\
0&1&0&1\\
0&1&1&1\\
0&1&1&1\\
1&0&0&1\\
0&0&0&1\\
0&0&1&1\\
0&0&1&1\\
0&1&0&1\\
0&1&0&1\\
0&1&1&1\\
0&1&1&1
\end{array}\right]
$$

Fig. 1. The B_4-based matrix with 3 column blocks of sizes 1, 2 and 1, respectively (left); the matrix A (middle); and the matrix A thresholded at 2, resulting in matrix B (right).

Definition 5.2. Let k and $d_1, \dots, d_k$ be positive integers and let $D = d_1 + \cdots + d_k$ denote their sum. The *B_D-based matrix with k column blocks of sizes $d_1, \dots, d_k$* is the matrix $A : X \times Y \to [k]_0$, where $X = [2^D]$ and $Y = [D]$, that results from the following procedure:

1. Decompose the D columns of B_D into k blocks of sizes $d_1, \dots, d_k$. The blocks are consecutively numbered from 1 to k.
2. Obtain A from B_D by replacing any 1-entry (resp. 0-entry) in a column belonging to block $b \in [k]$ by b (resp. by $b - 1$).

The *$B_D^\top$-based matrix with k row blocks of sizes $d_1, \dots, d_k$* is defined analogously. That is, we start with the transpose of B_D. We now have k row blocks of sizes $d_1, \dots, d_k$ and consider the matrix $A : X \times Y \to [k]_0$, where $X = [D]$ and $Y = [2^D]$, and in conditions 1. and 2. We replace 'column' by 'row'. Clearly, the $B_D^\top$-based matrix with k row blocks of sizes $d_1, \dots, d_k$ coincides with the transpose of the B_D-based matrix with k column blocks of sizes $d_1, \dots, d_k$.

Example 5.3. The matrix A shown below in Fig. 1 is the B_4-based matrix with three blocks of sizes 1, 2 and 1, respectively. We can think of B_4 as having the same block structure. If block $j \in \{1, 2, 3\}$ of A is thresholded at j, then we obtain again the original matrix B_4. If the columns of A are thresholded at the uniform threshold 2, then we obtain the matrix B (also shown in Fig. 1) whose second block coincides with the second block of B_4.

Note that the matrix A resulting from the above procedure has the property that, for any two columns y_1 in block b_1 and y_2 in block $b_2 > b_1$ and any row x, we have $A(x, y_1) \le A(x, y_2)$. We will refer to this property as *block monotonicity*.

At this point we also bring into play the matrix $\dot{A}$, which is defined as the matrix A augmented with a row of zeros. Formally, we assume that $0 \notin X$ and define $\dot{A} : (X \cup \{0\}) \times Y \to Z$ as the extension of A which satisfies $\dot{A}(0, y) = 0$ for all $y \in Y$. The (technical) use of $\dot{A}$ will become clear in Section 5.4 (in particular, this is explained after Definition 5.9), but it is already included in the statements that follow.

Lemma 5.4.

1. *Let $D = d_1 + \cdots + d_k$ and let A be the B_D-based matrix with k column blocks of sizes $d_1, \dots, d_k$. Then* $\mathrm{Vdim}(A) = \mathrm{Vdim}(\dot{A}) = \max_{j \in [k]} d_j$ *and* $\mathrm{Pdim}(A) = D$.
2. *Let $D = d_1 + \cdots + d_k$ and let A be the $B_D^\top$-based matrix with k row blocks of sizes $d_1, \dots, d_k$. Then* $\mathrm{Vdim}^*(A) = \max_{j \in [k]} d_j$ *and* $\mathrm{Pdim}^*(A) = D$.

Proof. It suffices to prove the first assertion because the second-one is then obtained by dualization. We first show that the pseudo-dimension of A equals D. Let $\mathbf{t} : [D] \rightarrow [k]$ be the mapping that assigns to every column in block $j \in [k]$ the threshold j. Then the matrix obtained by thresholding the columns of A at $\mathbf{t}$ equals B_D. It follows that $\mathrm{Pdim}(A) \geq \mathrm{VC}(B_D) = D$. Of course $\mathrm{Pdim}(A)$ cannot exceed D so that $\mathrm{Pdim}(A) = D$. Next, set $d_{max} = \max_{j\in[k]} d_j$. Pick some index $j_{max} \in [k]$ such that $d_{j_{max}} = d_{max}$. We still have to show that $\mathrm{Vdim}(A) = \mathrm{Vdim}(\dot{A}) = d_{max}$. Thresholding the columns of A at the uniform threshold j_{max}, we obtain a matrix B whose entries equal that of B_D within block j_{max}. This shows that $\mathrm{Vdim}(A) \geq d_{max}$. The inequality $\mathrm{Vdim}(\dot{A}) \leq d_{max}$ can be seen as follows. Pick a fixed but arbitrary $J \subseteq [D]$ of size $1+d_{max}$ and a fixed but arbitrary uniform threshold $t \in [k]$. Let B be the matrix obtained by thresholding the columns of $\dot{A}$ at t. The set J must contain two columns belonging to two different blocks, say column y_1 in block b_1 and column y_2 in block $b_2 > b_1$. By the block-monotonicity of A (which implies block-monotonicity for $\dot{A}$ as well), no row of B can assign label 1 to y_1 and label 0 to y_2. Since J and t were arbitrary choices, it follows that no set of size $1 + d_{max}$ can be V-shattered by $\dot{A}$. □

Setting $d_1 = \cdots = d_k = d$ in Lemma 5.4, we obtain the following result (which is almost the same as Theorem 3.2):

Corollary 5.5. *For every $d \geq 1$ and every $k \geq 1$, there exists a matrix $A : X \times Y \rightarrow [k]_0$ such that $\mathrm{Vdim}(A) = \mathrm{Vdim}(\dot{A}) = d$ and $\mathrm{Pdim}(A) = k \cdot d$.*

5.3. Proof of Theorem 4.2

Theorem 4.2 is a direct consequence of the following two results:

Lemma 5.6. *Let $d \geq 2$ and $k \geq 1$ be given. For $D = k \cdot 2^d$, let A be the $B_D^\top$-based matrix with k row blocks of size 2^d, respectively. Then*

$$\mathrm{Pdim}(A) = \mathrm{Pdim}(\dot{A}) = d\ ,\ \ \mathrm{Vdim}^*(A) = 2^d \ \text{ and } \ \mathrm{Pdim}^*(A) = k \cdot 2^d\ .$$

Proof. The identities $\mathrm{Vdim}^*(A) = 2^d$ and $\mathrm{Pdim}^*(A) = k \cdot 2^d$ are immediate from Lemma 5.4. Hence it suffices to verify the identity $\mathrm{Pdim}(A) = \mathrm{Pdim}(\dot{A}) = d$. Using (8), we can infer from $\mathrm{Vdim}^*(A) = 2^d$ that $\mathrm{Pdim}(\dot{A}) \geq \mathrm{Pdim}(A) \geq \mathrm{Vdim}(A) \geq d$. Hence the proof can be accomplished by showing that $\mathrm{Pdim}(\dot{A}) \leq d$. For sake of brevity, set

$$s = 2^d \ \text{ and } \ \bar{d} = 1 + d\ .$$

Assume for the sake of contradiction that $\mathrm{Pdim}(\dot{A}) \geq \bar{d}$ and fix some witness $(I, J, \mathbf{t})$ for this inequality, i.e.,

1. $I \subseteq [D]_0$, $|I| = 2^{\bar{d}}$, $J \subseteq [2^D]$, $|J| = \bar{d}$ and $\mathbf{t} : J \rightarrow [k]$ assigns a threshold to each column of $\dot{A}_{I,J}$.[5]
2. The matrix $B : I \times J \rightarrow \{0, 1\}$ obtained by thresholding the columns of $\dot{A}_{I,J}$ at $\mathbf{t}$ equals $B_{\bar{d}}$ (modulo row permutations).

Note that $s = 2^d$ is the block-size in matrix A. Let $A_{I,J}$ and B inherit the block structure of A, i.e., a block I_b in A induces a block $I_b \cap I$ in $A_{I,J}$ resp. in B. Then s is an upper bound on the block size in B. Moreover B has $|I| = 2^{\bar{d}} = 2s$ many rows. We proceed with a series of easy-to-prove claims.

Claim 1: If $i_1 < i_2$, $B(i_1, j) = 1$ and $B(i_2, j) = 0$, then the indices i_1 and i_2 must be in the same row block of B.

Proof: Let $t_j = \mathbf{t}(j)$. From $B(i_1, j) = 1$ and $B(i_2, j) = 0$, we infer that $A(i_2, j) < t_j \leq A(i_1, j)$. If i_1 and i_2 were in different row blocks, this would contradict the block-monotonicity of A.

Claim 2: Each column in B starts with a 0-entry and ends with a 1-entry.

Proof: Assume for contradiction that there is a column j of B that starts with a 1-entry. By the first balance condition, there must be s zeros among the remaining entries. But, to make this possible, there must be a block of size at least $s+1$. However, as observed above already, all blocks in B are of size at most s. We arrived at a contradiction. Hence each column of B starts with a 0-entry. For reasons of symmetry, it ends with a 1-entry.

Claim 3: B must have a column j_1 whose second entry is a one and a column j_2 whose second-last entry is a zero.

Proof: If not, then B would have two all-zeros rows or two all-ones rows. That would contradict the distinctness condition.

Claim 4: The second entry in column j_1 of B (which is 1-entry) must be followed by $s-1$ 0-entries. The second-last entry of column j_2 of B (which is a 0-entry) must be preceded by $s - 1$ 1-entries.

Proof: If not, we would obtain a contradiction to the first balance condition or to the fact that the blocks in B are of size at most s. Compare with the proof of Claim 2.

[5] Recall from the definition of $\dot{A}$ that this matrix is obtained from A by adding an all-zeros row which is indexed by 0.

We are now in the position to derive the desired contradiction. Column j_1 of B starts with 010^{s-1}. It follows that entries number $2, \ldots, s+1$ belong to the same row block. A similar inspection of column j_2 reveals that the entries with numbers $s, \ldots, 2s$ belong to the same row block. Thus, all entries, with the possible exception of the first- and the last-one, belong to the same row block. It follows that there is a block of size at least $2s-2$. Since s is the maximal block size, we have $2s-2 \le s$, or equivalently, $s \le 2$. This is in contradiction to $d \ge 2$ and $s = 2^d \ge 4$. $\square$

Lemma 5.7. *Let $k \ge 2$ and let A be the $B_{k+2}^{\top}$-based matrix with k row blocks of sizes*

$$d_j = \begin{cases} 2 & \text{for } j = 1, k \\ 1 & \text{for } j = 2, \ldots, k-1 \end{cases}.$$

Then $\mathrm{Pdim}(A) = 1$ *and* $\mathrm{Pdim}^*(A) = k+2$.

Proof. The identity $\mathrm{Pdim}^*(A) = k+2$ is immediate from Lemma 5.4. Clearly $\mathrm{Pdim}(A) \ge 1$. Hence it suffices to show that $\mathrm{Pdim}(A) \le 1$. The rows of A have indices $1, \ldots, k+2$. Assume for contradiction that $\mathrm{Pdim}(A) \ge 2$ and pick a witness $(I, J, \mathbf{t})$ for this inequality so that the following holds:

- $J = \{j_1, j_2\} \subset [2^{k+2}]$, $I \subseteq [k+2]$ with $|I| = 4$ and $\mathbf{t} : J \to [k]$, say $\mathbf{t}(j_1) = t_1$ and $\mathbf{t}(j_2) = t_2$.
- The matrix $B : I \times J \to \{0, 1\}$ obtained by thresholding the columns of $A_{I,J}$ at $\mathbf{t}$ equals B_2 (up to row permutations).

Consequently B satisfies the distinctness condition and the balance conditions. Consider the smallest index i_1 and the second-smallest index i_2 in I. Note that, since $|I| = 4$ and the last row block of A is of size 2, neither i_1 nor i_2 belongs to the last block, i.e., either $i_1 = 1$ and $i_2 = 2$ (so that i_1 and i_2 represent the first row block) or i_2 belongs to one of the inner blocks consisting of a single row only. In both cases, the remaining indices in I do not belong to the same row block as i_2 or i_1. We proceed by case analysis:

Case 1: $B(i_1, j) = B(i_2, j) = 0$ for all $j \in J$.

Then the first two rows of B realize the all-zeros pattern, which contradicts the distinctness condition.

Case 2: There exist $j \in J$ and $i' \in \{i_1, i_2\}$ such that $B(i', j) = 1$.

Then the block-monotonicity in A implies that all remaining entries in column j of B are 1-entries, which contradicts the first balance condition.

In both cases, we arrived at a contradiction. $\square$

Lemma 5.7 does not cover the case $k = 1$ in the second assertion of Theorem 4.2. But this case is easy to handle: setting $A = B_3^{\top}$, we obtain $\mathrm{Pdim}(A) = \mathrm{VC}(A) = 1$ and $\mathrm{Pdim}^*(A) = \mathrm{VC}^*(A) = 3$.

For technical reasons, we will later also need the following result:

Lemma 5.8. *Let $k \ge 2$ and let A be the $B_{k+1}^{\top}$-based matrix with k row blocks of sizes*

$$d_j = \begin{cases} 2 & \text{for } j = k \\ 1 & \text{for } j = 1, \ldots, k-1 \end{cases}.$$

Then $\mathrm{Pdim}(\dot{A}) = \mathrm{Pdim}(A) = 1$ *and* $\mathrm{Pdim}^*(A) = k+1$.

Proof. The proof is quite similar to the proof of Lemma 5.7. The identity $\mathrm{Pdim}^*(A) = k+1$ is again immediate from Lemma 5.4. Clearly $\mathrm{Pdim}(\dot{A}) \ge \mathrm{Pdim}(A) \ge 1$. Hence it suffices to show that $\mathrm{Pdim}(\dot{A}) \le 1$. The rows of $\dot{A}$ have indices $0, 1, \ldots, k+1$. Assume for contradiction that $\mathrm{Pdim}(\dot{A}) \ge 2$ and pick a witness $(I, J, \mathbf{t})$ for this inequality so that the following holds:

- $J = \{j_1, j_2\} \subset [2^{k+2}]$, $I \subseteq [k+1]_0$ with $|I| = 4$ and $\mathbf{t} : J \to [k]$, say $\mathbf{t}(j_1) = t_1$ and $\mathbf{t}(j_2) = t_2$.
- The matrix $B : I \times J \to \{0, 1\}$ obtained by thresholding the columns of $A_{I,J}$ at $\mathbf{t}$ equals B_2 (up to row permutations).

Consequently B satisfies the distinctness condition and the balance conditions. Consider the smallest index i_1 and the second-smallest index i_2 in I. Note that, since $|I| = 4$ and the last row block of A is of size 2, neither i_1 nor i_2 belongs to the last block. All blocks of A, except for the last-one, are of size 1. Thus the indices in $I \setminus \{i_1, i_2\}$ do not belong to the same row block as i_2 or i_1. The proof can now be accomplished by the same kind of case analysis that we have used at the end of the proof of Lemma 5.7. As before, a contradiction is obtained in any case. $\square$

5.4. Proofs of Corollary 4.3 and of Theorems 4.4 and 4.5

Matrices A with the properties as prescribed by Theorems 4.4 and 4.5 are easy to construct by means of a suitable operation that merges matrices of a given matrix family into a single matrix.

Definition 5.9 (*Merge-Operation*)**.** Let $(A_k)_{k\geq 1}$ with $A_k : X_k \times Y_k \rightarrow [k]_0$ be a given family of matrices. Let X (resp. Y) denote the disjoint union of the sets X_k (resp. Y_k) with $k \geq 1$. Assume that $X \cap Y = \emptyset$. For every $x \in X$, let $k(x)$ denote the unique k such that $x \in X_k$. The notation $k(y)$ is understood analogously. The matrix $A : X \times Y \rightarrow [0, 1]$ given by

$$A(x, y) = \begin{cases} \frac{A_{k(x)}(x,y)}{k(x)} & \text{if } k(y) = k(x) \\ 0 & \text{otherwise} \end{cases},$$

is called the *merge of the family* $(A_k)_{k\geq 1}$.

The merge-operation reveals why we introduce the matrix $\dot{A}$: The pseudo-dimension (or any other combinatorial dimension for that matter) of the matrix A restricted to the columns Y_k is nothing more than the pseudo-dimension of the functions in A_k augmented with an infinite number of functions that are zero everywhere. The pseudo-dimension of this function class clearly equals the pseudo-dimension of the matrix $\dot{A}_k$. The merge-operation has the following properties:

Lemma 5.10. *Let $A : X \times Y \rightarrow [0, 1]$ be the merge of the family $(A_k)_{k\geq 1}$. Then the following holds:*

1. $\mathrm{P}_{1/(2k)}(A) \geq \mathrm{Pdim}(A_k)$ *for all* $k \geq 1$.
2. *Let* $d_0 \in \mathbb{N}$. *If* $\sup_k \mathrm{Pdim}(\dot{A}_k) \leq d_0$, *then* $\mathrm{Pdim}(A) \leq d_0$.
3. $\mathrm{V}_{1/(2k)}(A) \geq \mathrm{Vdim}(A_k)$ *for all* $k \geq 1$.
4. *Let* $d_0 \in \mathbb{N}$. *If* $\sup_k \mathrm{Vdim}(\dot{A}_k) \leq d_0$, *then* $\mathrm{Vdim}(A) \leq d_0$.

Proof. We only prove the first two assertions of the lemma; the proofs of other two assertions are quite similar.

Note that, for $k = k(x) = k(y)$, A coincides with A_k except for scaling down the values $0, 1, \dots, k$ by factor k. Since A_k takes integer values, each set that can be P-shattered by A_k can actually be $P_{1/2}$-shattered. After down-scaling, the width of shattering becomes $1/(2k)$. From these observations, the first assertion of the lemma easily follows.

We proceed with the proof of the second assertion. Let d be an arbitrary but fixed positive integer such that Y contains a set of size d that is P-shattered by A. It suffices to show that $d \leq d_0$. Fix some witness $(I, J, \mathbf{t})$ so that the following holds:

1. $I \subset X$, $|I| = 2^d$, $J \subset Y$, $|J| = d$ and $\mathbf{t} : J \rightarrow \mathbb{N}$ assigns a threshold $t_y := \mathbf{t}(y)$ to every $y \in J$.
2. The matrix B obtained by thresholding the columns of $A_{I,J}$ at $\mathbf{t}$ equals B_d (with rows indexed by I and columns indexed by J).

It follows that B satisfies the distinctness condition and the balance conditions.

Claim 1. *For every $y \in J$, we have $t_y > 0$.*

Proof. $t_y \leq 0$ would imply that column y of B has no 0-entry, which contradicts the first balance condition. □

Claim 2. *The mapping $y \mapsto k(y)$ assigns the same value to all $y \in J$.*

Proof. Assume to the contrary that there exist $y_1, y_2 \in J$ such that $k(y_1) \neq k(y_2)$. Then, for every row x of B, at least one of the entries $B[x, y_1]$ and $B[x, y_2]$ equals 0 (because $k(x)$ cannot be equal to both, $k(y_1)$ and $k(y_2)$). By the first balance condition, any column in B has as many 0- as 1-entries. Since this is particularly true for the columns y_1 and y_2, it follows that, for every row x of B, exactly one of the entries $B[x, y_1]$ and $B[x, y_2]$ equals 0. Thus column y_2 of B is the entry-wise logical negation of the column y_1. This, however, contradicts the second balance condition. □

Claim 3. *Let k_1 denote the common k-value of all $y \in J$. Then any row x in B with $k(x) \neq k_1$ has 0-entries only.*

Proof. This is straightforward. □

We conclude from Claims 2 and 3 that $d \leq \mathrm{Pdim}(\dot{A}_{k_1})$ and, by assumption, the latter quantity is at most d_0, which concludes the proof. □

We now conclude with the proof of Corollary 4.3 and Theorems 4.4 and 4.5.

Proof of Corollary 4.3. Lemma 5.8 tells us that for every $k \geq 2$ there exists a matrix A_k such that $\mathrm{Pdim}(\dot{A}_k) = \mathrm{Pdim}(A_k) = 1$ and $\mathrm{Pdim}^*(A_k) = k+1$. Let A be the merge of the family $(A_k)_{k\geq 2}$. The first two claims of Lemma 5.10 imply that $\mathrm{Pdim}(A) = 1$ while the dual of the first claim yields $\mathrm{Pdim}^*_{1/(2k)}(A) \geq \mathrm{Pdim}^*(A_k) = k + 1$. Since $\mathrm{Pdim}^*(A) \geq \mathrm{Pdim}^*_{1/(2k)}(A)$ for every k, it follows that $\mathrm{Pdim}^*(A) = \infty$. □

Proof of Theorem 4.4. Corollary 5.5 tells us that for every $d \geq 1$ and $k \geq 1$, there exists a matrix $A_k : X \times Y \rightarrow [k]_0$ such that $\mathrm{Vdim}(A_k) = \mathrm{Vdim}(\dot{A}_k) = d$ and $\mathrm{Pdim}(A_k) = k \cdot d$. Let A be the merge of the family $(A_k)_{k\geq 1}$. The first claim of Lemma 5.10 tells us then that $P_{1/(2k)}(A) \geq \mathrm{Pdim}(A_k) = k \cdot d$. The last two claims of Lemma 5.10 imply that $\mathrm{Vdim}(A) = d$. □

Proof of Theorem 4.5. We start with the proof of the first claim in Theorem 4.5. Lemma 5.6 tells us that there exists a matrix A_k with $\mathrm{Pdim}(\dot{A}_k) = \mathrm{Pdim}(A_k) = d$ and $\mathrm{Pdim}^*(A_k) = k \cdot 2^d$. Let A be the merge of the family $(A_k)_{k\geq 1}$. The first two claims in Lemma 5.10 imply that $\mathrm{Pdim}(A) = d$ while the dual of the first claim yields $\mathrm{Pdim}^*_{1/(2k)} \geq k \cdot 2^d$. This proves the first claim of Theorem 4.5. In order to prove the second claim, we can instead use Lemma 5.8 and start with the existence of a matrix A_k for which $\mathrm{Pdim}(\dot{A}) = \mathrm{Pdim}(A) = 1$ and $\mathrm{Pdim}^*(A_k) = k + 1$. Then apply the same argument as for the first claim. □

Data availability

No data was used for the research described in the article.

Acknowledgments

We thank the anonymous reviewers for their valuable comments. The first author thanks Tim Roughgarden for discussions that (indirectly) lead to the questions studied in this work.

Appendix. Assouad's bound for uniform dimensions

We say that $J \subseteq Y$ is *VC-shattered* by $A : X \times Y \to \{0, 1, *\}$ if, for every function $b : J \to \{0, 1\}$, there exists an $x \in X$ such that, for every $y \in J$, we have $B(x, y) = b(y)$. We first note that (4) is also valid for every matrix of the form $A : X \times Y \to \{0, 1, *\}$: the central observation in the proof is that B_d contains $B^\top_{\lfloor \log d \rfloor}$ as a submatrix. This implies that $\mathrm{VC}(A) \geq \lfloor \log \mathrm{VC}^*(A) \rfloor$, which is equivalent to (4).

Consider now a matrix of the general form $A : X \times Y \to Z$ with $Z \subseteq \mathbb{R}$. Making use of the concept of uniform Ψ-dimensions from [4], the result of Assouad can be extended to several other combinatorial dimensions. Let Ψ denote a family of substitutions of the form $\psi : \mathbb{R} \to \{0, 1, *\}$. Denote by $\psi(A)$ the matrix obtained from A by performing the substitution ψ entry-wise. The *uniform Ψ-dimension* of A is then defined as

$$\Phi_U(A) = \sup_{\psi \in \Psi} \mathrm{VC}(\psi(A)) \ .$$

Let Ψ_Y denote the set of all collections $\bar{\psi} = (\psi_y)_{y\in Y}$ with $\psi_y \in \Psi$. Denote by $\bar{\psi}(A)$ the matrix obtained from A by replacing each entry $A(x, y)$ with $\psi_y(A(x, y))$. The *(non-uniform) Ψ-dimension* of A is defined as

$$\Phi(A) = \sup_{\bar{\psi} \in \Psi_Y} \mathrm{VC}(\bar{\psi}(A)) \ .$$

As usual, we get the corresponding dual dimensions by setting $\Phi^*(A) = \Phi(A^\top)$ and $\Phi^*_U(A) = \Phi_U(A^\top)$. Note that $\psi(A^\top) = \psi(A)^\top$ while $\bar{\psi}(A^\top)$ is not generally equal to $\bar{\psi}(A)^\top$.

As noted in [4], several popular combinatorial dimensions can be viewed as (uniform or non-uniform) ψ-dimension. Here we are particularly interested in the P-, P_γ-, V-and V_γ-dimension:

Remark A.1. We next explain how to interpret known dimensions as special cases of the Ψ-dimension.

1. If Ψ is the set of mappings ψ_t of the form $\psi_t(a) = \mathrm{sgn}(a - t)$ for some $t \in \mathbb{R}$, then $\Phi(A) = \mathrm{Pdim}(A)$ and $\Phi_U(A) = \mathrm{Vdim}(A)$ (see [4]).
2. If Ψ is the set of mappings ψ_t of the form

$$\psi_t(a) = \begin{cases} 1 & \text{if } a \geq t + \gamma \\ 0 & \text{if } a < t - \gamma \\ * & \text{otherwise} \end{cases}$$

for some $t \in \mathbb{R}$ and $\gamma > 0$, then $\Phi(A) = \mathrm{P}_\gamma(A)$ and $\Phi_U(A) = V_\gamma(A)$.

The following calculation, with ψ ranging over all functions in Ψ, shows that Theorem 3.5 can be extended to any uniform Ψ-dimension at the place of the VC-dimension:

$$\begin{aligned} \Phi^*_U(A) &= \Phi_U(A^\top) = \sup_\psi \mathrm{VC}(\psi(A^\top)) = \sup_\psi \mathrm{VC}(\psi(A)^\top) \\ &= \sup_\psi \mathrm{VC}^*(\psi(A)) \leq \sup_\psi \left(2^{\mathrm{VC}(\psi(A))+1} - 1\right) \\ &= 2^{\sup_\psi \mathrm{VC}(\psi(A))+1} - 1 = 2^{\Phi_U(A)+1} - 1 \ . \end{aligned}$$

We remark that a similar argument for the non-uniform Ψ-dimension fails as it then no longer holds that $\bar{\psi}(A^\top) = \bar{\psi}(A)^\top$ (which is the argument we use in the third equality above).

References

[1] N. Alon, S. Ben-David, N. Cesa-Bianchi, D. Haussler, Scale-sensitive dimensions, uniform convergence, and learnability, J. ACM 44 (4) (1997) 615–631.
[2] M. Anthony, J. Ratsaby, Large-width bounds for learning half-spaces on distance spaces, Discrete Appl. Math. 243 (2018) 73–89.
[3] P. Assouad, Densité et dimension, Ann. Inst. Fourier 33 (3) (1983) 233–282.
[4] S. Ben-David, N. Cesabianchi, D. Haussler, P.M. Long, Characterizations of learnability for classes of [n]-valued functions, J. Comput. System Sci. 50 (1) (1995) 74–86.
[5] S. Hanneke, A. Kontorovich, M. Sadigurschi, Sample compression for real-valued learners, in: Algorithmic Learning Theory, PMLR, 2019, pp. 466–488.
[6] D. Haussler, Decision theoretic generalizations of the PAC model for neural net and other learning applications, Inform. and Comput. 100 (1) (1992) 78–150.
[7] M.J. Kearns, R.E. Schapire, Efficient distribution-free learning of probabilistic concepts, J. Comput. System Sci. 48 (3) (1994) 464–497.
[8] J. Matoušek, Lectures on Discrete Geometry, Vol. 108, Springer, 2002.
[9] S. Moran, A. Yehudayoff, Sample compression schemes for VC classes, J. ACM 63 (3) (2016) 21:1–21:10.
[10] D. Pollard, Empirical processes: theory and applications, in: NSF-CBMS Regional Conference Series in Probability and Statistics, JSTOR, 1990, pp. i–86.
[11] V.N. Vapnik, Inductive principles of the search for empirical dependences (methods based on weak convergence of probability measures), in: Proceedings of the Second Annual Workshop on Computational Learning Theory, Morgan Kaufmann, San Mateo, CA, 1989, pp. 3–21.
[12] V.N. Vapnik, A.Y. Chervonenkis, On uniform convergence of the frequencies of events to their probabilities, Theory Probab. Appl. 16 (2) (1971) 264–281.