# WHERE PARITY DIMENSION PARTS FROM DS DIMENSION: EXPLICIT WITNESSES, AND THE THRESHOLD AT DIMENSION THREE

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. A finite nonempty set $Q \subseteq \mathcal{Y}^D$ of patterns is a *pseudo-cube* on the finite index set $D$ when every point of $Q$ has, at every coordinate, a second point of $Q$ agreeing with it off that coordinate and differing there; the largest $|D|$ for which a function class has a pseudo-cube in its trace on $D$ is its DS dimension, the invariant that characterises multiclass PAC learnability. Li (arXiv:2607.07423) refines the notion over $\mathbb{F}_2$: $Q$ is an *even* pseudo-cube when it carries a nonzero $\eta \in \mathbb{F}_2^Q$ all of whose coordinate-deletion marginals vanish, and the resulting invariant, the parity dimension, is what survives an autoregressive rollout. Li proves ParDim $\leq$ DSdim with equality on binary classes, and that the inequality can be strict — but the strictness is derived indirectly, no pseudo-cube without an even certificate is exhibited, and nothing there says at which dimension the two invariants can first part.

We prove that *every pseudo-cube on at most two coordinates carries an even certificate*, for an arbitrary label set; hence $\mathrm{ParDim}(\mathcal{H}) \geq \min(\mathrm{DSdim}(\mathcal{H}), 2)$, and the two dimensions agree whenever the DS dimension is at most 2. At dimension three they need not: we exhibit a 24-point minimal pseudo-cube inside a $3 \times 3 \times 4$ grid, no nonempty subclass of which carries an even certificate. Three is therefore the least dimension at which parity dimension and DS dimension can differ. We also verify Li's separation directly on the 28-row table its construction produces, independently of the rollout non-increase theorem from which that paper derives it. The computations rest on an elementary normal form — a finite class carries an even certificate exactly when some nonempty subclass has all of its *own* coordinate-deletion fibres of even cardinality — which makes parity dimension a decidable combinatorial condition on a finite table. Every theorem below is machine-checked in Lean 4 over Mathlib, in its finite-table form and with no appeal to native evaluation; the exhaustive searches, and the elementary observation that carries the finite tables back to Li's rollout class, lie outside the formal development and are marked as such.

## 1. INTRODUCTION

**Two dimensions on a finite table.** Let $\mathcal{Y}$ be a set of labels, $\mathcal{X}$ a domain, and $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ a class of functions. For a finite $D \subseteq \mathcal{X}$ write $\mathcal{H}|_D \subseteq \mathcal{Y}^D$ for the trace of $\mathcal{H}$ on $D$; and for a finite $V \subseteq \mathcal{Y}^D$, a coordinate $x \in D$ and a pattern $v \in \mathcal{Y}^D$ write

$$V_x(v) = \{\, w \in V : \ w|_{D\setminus\{x\}} = v|_{D\setminus\{x\}} \,\}$$

for the *coordinate-deletion fibre* of $v$ inside $V$: the set of points of $V$ that agree with $v$ everywhere except possibly at $x$.

A finite nonempty $Q \subseteq \mathcal{Y}^D$ is a *pseudo-cube on $D$* if $|Q_x(q)| \geq 2$ for every $q \in Q$ and every $x \in D$; equivalently, every point has at every coordinate a neighbour in $Q$ that differs there and agrees elsewhere. The class $\mathcal{H}$ *DS-shatters* $D$ when $\mathcal{H}|_D$ contains a pseudo-cube on $D$, and $\mathrm{DSdim}(\mathcal{H})$ is the largest size of a DS-shattered set. The notion is due to Daniely and Shalev-Shwartz [3], and Brukhim, Carmon, Dinur, Moran and Yehudayoff [2] proved that DSdim characterises multiclass PAC learnability, resolving a problem open since the late 1980s.

Li [5] introduces an $\mathbb{F}_2$ refinement. Call a nonzero vector $\eta \in \mathbb{F}_2^V$ an *even certificate* for a finite nonempty $V \subseteq \mathcal{Y}^D$ if

$$\sum_{w \in V_x(v)} \eta(w) = 0 \qquad \text{for every } x \in D \text{ and every } v \in V, \tag{1}$$

and call $V$ an *even pseudo-cube* when such an $\eta$ exists [5, Definition 3.1]. The *parity dimension* $\mathrm{ParDim}(\mathcal{H})$ is the supremum of the $d = |D|$ for which some finite nonempty $\mathcal{H}' \subseteq \mathcal{H}$ has $\mathcal{H}'|_D$ an even pseudo-cube [5, Definition 3.2].[1] Li's chain [5, Corollary 3.4] reads

$$\tfrac{1}{2}\,\mathrm{DSdim}(\mathcal{H}) \;\le\; \sup_{n\ge 1}\mu_{\mathcal{H}}(n) \;\le\; \mathrm{ParDim}(\mathcal{H}) \;\le\; \mathrm{DSdim}(\mathcal{H}), \tag{2}$$

$\mu$ being one-inclusion density, with ParDim = DSdim = VCdim for binary classes. The right-hand inequality holds because the support of an even certificate is a pseudo-cube. The point of the refinement is dynamic: ParDim does not increase under autoregressive rollout [5, Theorem 3.5] whereas DSdim can [5, Theorem 3.6], and this is what yields a PAC bound with no dependence on rollout length [5, Theorem 1.1].

**What is left open.** Since an even pseudo-cube is a pseudo-cube, the interesting direction of (2) is whether ParDim < DSdim is possible at all, and where. Li answers the first question in [5, Corollary 3.7]: for the rollout class of the gadget of [5, Appendix D], ParDim = 2 while DSdim $\ge 3$. The proof is indirect. It runs $\mathrm{ParDim}(\mathcal{H}) \le \mathrm{DSdim}(\mathcal{H}) = 2$, then applies the rollout non-increase theorem [5, Theorem 3.5] to get $\mathrm{ParDim}(\mathrm{Roll}_{\mathtt{halt}}(\mathcal{H})) \le 2$, then uses the left-hand inequality of (2) for the matching lower bound. In particular no pseudo-cube carrying no even certificate is ever displayed there — the existence of such an object is a consequence of a deferred theorem, and the same deferred theorem carries the paper's headline sample-complexity bound. As for the second question, [5, Corollary 3.4] asserts ParDim = DSdim for binary classes, proved in [5, Appendix A] by an argument specific to the binary cube (every coordinate-deletion fibre of $\{0,1\}^D$ has exactly two points, so the all-one vector is an even certificate), which says nothing over a larger alphabet — and Li's own separating example lives over a four-letter alphabet. Nothing in [5], in [2], or in the $k$-ary Sauer lemma of Hanneke, Meng, Moran and Shaeiri [4] bounds the dimension at which ParDim < DSdim can begin.

**Results.** Our main theorem settles the threshold.

**Theorem 1.1** (Theorem 4.1)**.** *Let $\mathcal{Y}$ be an arbitrary label set, $D$ a set of at most two coordinates, and $V \subseteq \mathcal{Y}^D$ a finite nonempty pseudo-cube on $D$. Then $V$ carries an even certificate. Consequently* $\mathrm{ParDim}(\mathcal{H}) \ge \min(\mathrm{DSdim}(\mathcal{H}), 2)$ *for every class $\mathcal{H}$, and* $\mathrm{ParDim}(\mathcal{H}) = \mathrm{DSdim}(\mathcal{H})$ *whenever* $\mathrm{DSdim}(\mathcal{H}) \le 2$.

At dimension three the conclusion fails, and it fails on an object considerably smaller than the one Li's construction produces.

**Theorem 1.2** (Theorem 5.1)**.** *The 24-point set $S$ of Table 1 inside the grid $\{0,1,2\}\times\{0,1,2\}\times\{0,1,2,3\}$ is a pseudo-cube on its three coordinates, it is minimal (deleting any one of its points leaves no pseudo-cube), and no nonempty subset of $S$ carries an even certificate.*

Together the two theorems say that *three is the least dimension at which parity dimension and DS dimension can differ*. We also verify Li's separation directly, on the 28-row table of [5, Appendix D] itself: that table is a 3-dimensional pseudo-cube no nonempty subclass of which carries an even certificate (Theorems 6.1 and 6.2), so, given only that the rollout class is constant off its three root states (Remark 6.4), [5, Corollary 3.7] holds for a reason that does not pass through [5, Theorem 3.5]. We record the matching data for the base class as well (§6).

The tool behind all of the finite verifications is an elementary normal form (Lemma 3.1): a finite class carries an even certificate if and only if it has a nonempty subclass all of whose *own* coordinate-deletion fibres have even cardinality. Over $\mathbb{F}_2$ a vector is the indicator of its support, and the support of a certificate meets every fibre in an even number of points. Because "all my own fibres are even" does not mention the ambient class, one computation on a table settles every subclass of it at once — which is exactly the quantifier in the definition of ParDim — and it makes the condition purely combinatorial: no linear algebra, and no quantification over $\mathbb{F}_2$ vectors. The normal form is not stated in [5], whose proof of ParDim $\le$ DSdim passes through the same computation and extracts from it only that "that fiber contains another point $q' \in Q$".

[1] Throughout, statement numbers of [5] refer to the arXiv v1 e-print, as compiled from its LaTeX source; we do not cite numbers read off unresolved cross-references.

Objects of the kind produced here are rare and large, which is why an explicit one is worth recording: among the 3 189 335 pseudo-cubes contained in the $3\times3\times4$ grid only 216 carry no even certificate, and, by the exhaustive searches of Remark 5.3, which lie outside the formal development, no pseudo-cube with at most 23 points, over any alphabet, fails to carry one. A random search among small pseudo-cubes therefore finds nothing, and the absence of a witness is easy to mistake for the absence of the phenomenon. It is not: Li's separation is correct, and the two witnesses below make it explicit.

**Organisation.** §2 fixes notation. §3 proves the normal form and its two immediate consequences. §4 proves the threshold theorem. §5 gives the 24-point witness, the control showing that failure is not a property of dimension three as such, and the record of the exhaustive searches. §6 verifies the data of Li's example. §7 places the results against earlier work and §8 describes the formal verification; Appendix A reproduces the formal statement of every result.

## 2. Preliminaries

We work with one finite table at a time. Fix a finite index set $D$ (the coordinates), a label set $\mathcal{Y}$, and a finite $V \subseteq \mathcal{Y}^D$; all definitions below are the ones of [5] read on such a $V$. Recall $V_x(v)$ from §1. Note that $V_x(v)$ is nonempty ($v$ lies in it) and that the fibres $\{V_x(v) : v \in V\}$ partition $V$ for each fixed $x$.

**Definition 2.1** ([5, Definition 2.9], after [3])**.** $V$ is a *pseudo-cube on* $D$ if $V \neq \emptyset$ and, for every $v \in V$ and every $x \in D$, there is $w \in V$ with $w \neq v$ and $w|_{D\setminus\{x\}} = v|_{D\setminus\{x\}}$; equivalently $|V_x(v)| \geq 2$ for all $v \in V$, $x \in D$. A class $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ DS-shatters a finite $D \subseteq \mathcal{X}$ if $\mathcal{H}|_D$ contains a pseudo-cube on $D$, and $\mathrm{DSdim}(\mathcal{H})$ is the supremum of the sizes of DS-shattered sets.

**Definition 2.2** ([5, Definition 3.1])**.** An *even certificate* for $V$ is a vector $\eta \in \mathbb{F}_2^V$, not identically zero, satisfying (1). If one exists we say $V$ *carries* an even certificate, and call $V$ an even pseudo-cube on $D$.

Li's definition quantifies the marginal condition over the restricted patterns $u \in \mathcal{H}'|_{D\setminus\{x\}}$; since the fibres of $x$ are indexed by exactly those patterns, this is the form (1) used here. The parity dimension $\mathrm{ParDim}(\mathcal{H})$ is then the supremum of the $|D|$ for which some finite nonempty $\mathcal{H}' \subseteq \mathcal{H}$ makes $\mathcal{H}'|_D$ an even pseudo-cube [5, Definition 3.2], with the convention $\mathrm{ParDim}(\mathcal{H}) = 0$ when no positive-dimensional even pseudo-cube exists.

**Definition 2.3.** $V$ is an *even set* on $D$ if $|V_x(v)|$ is even for every $v \in V$ and every $x \in D$.

An even set need not be a pseudo-cube for a trivial reason and is one for a trivial reason: the empty set is even, and a *nonempty* even set has all fibres of even cardinality at least 1, hence at least 2. We use this repeatedly.

## 3. A normal form for even certificates

**Lemma 3.1.** *A finite $V \subseteq \mathcal{Y}^D$ carries an even certificate if and only if some nonempty $S \subseteq V$ is an even set on $D$.*

*Proof.* Suppose $\eta$ is an even certificate for $V$ and put $S = \mathrm{supp}(\eta) = \{v \in V : \eta(v) = 1\}$, which is nonempty. Fix $x \in D$ and $v \in S$. Over $\mathbb{F}_2$ the sum in (1) is the parity of $|V_x(v) \cap S|$, so that intersection has even size; and $V_x(v) \cap S = S_x(v)$, because the fibre of $v$ inside $S$ consists of the points of $S$ agreeing with $v$ off $x$. Hence $S$ is even.

Conversely, let $S \subseteq V$ be nonempty and even, and let $\eta = 1_S$. It is nonzero. For $x \in D$ and $v \in V$, the sum in (1) is the parity of $|V_x(v) \cap S| = |S_x(v)|$. If $v \in S$ this is even because $S$ is even. If $v \notin S$ and $S_x(v) = \emptyset$ the sum is 0; and if some $w \in S_x(v)$ exists then $S_x(w) = S_x(v)$, which is even because $w \in S$. $\square$

Two consequences are used throughout. The first is that the property is monotone in the class, which the raw Definition 2.2 does not make evident, since enlarging $V$ changes every fibre.

**Corollary 3.2.** *If $S \subseteq V$ and $S$ carries an even certificate, then so does $V$.*

*Proof.* By Lemma 3.1 some nonempty $T \subseteq S$ is even; $T \subseteq V$ is then a nonempty even subset of $V$, and Lemma 3.1 applies again. ☐

Corollary 3.2 is what makes a single computation decide a whole supremum: to know that no finite nonempty subclass of a class has an even pseudo-cube on $D$, it suffices to refute an even certificate for the one ambient table $\mathcal{H}|_D$.

**Corollary 3.3** (ParDim $\leq$ DSdim, [5, Corollary 3.4])**.** *If $V$ carries an even certificate then some nonempty $S \subseteq V$ is a pseudo-cube on $D$.*

*Proof.* Take the nonempty even $S \subseteq V$ of Lemma 3.1. Each of its fibres is nonempty and of even size, hence of size at least 2, which is Definition 2.1. ☐

## 4. Below dimension three the two dimensions agree

**Theorem 4.1.** *Let $\mathcal{Y}$ be any label set, let $D$ have at most two elements, and let $V \subseteq \mathcal{Y}^D$ be a finite nonempty pseudo-cube on $D$. Then $V$ carries an even certificate.*

*Proof.* Write $D = \{x_0, x_1\}$, allowing $x_0 = x_1$, which covers $|D| \leq 1$ as well. Let $R = \{v|_{D\setminus\{x_0\}} : v \in V\}$ and $C = \{v|_{D\setminus\{x_1\}} : v \in V\}$ be the value sets of the two coordinate-deletion maps. For each $x \in D$ the fibres $\{V_x(v)\}$ partition $V$ and each has at least two elements, so

$$2|R| \leq |V|, \qquad 2|C| \leq |V|.$$

Fix $v_0 \in V$ and let $c_0 \in C$ be its $x_1$-deletion value. Consider the $\mathbb{F}_2$-linear map

$$\Phi : \mathbb{F}_2^V \longrightarrow \mathbb{F}_2^J, \qquad J = (\{x_0\} \times R) \ \cup \ (\{x_1\} \times (C \setminus \{c_0\})),$$

sending $\eta$ to the tuple of its fibre sums over the fibres indexed by $J$; that is, all $x_0$-fibres, and all $x_1$-fibres except the one indexed by $c_0$. Then

$$|J| \leq |R| + |C| - 1 \leq \tfrac{1}{2}|V| + \tfrac{1}{2}|V| - 1 = |V| - 1 < |V|,$$

so $|\mathbb{F}_2^J| < |\mathbb{F}_2^V|$ and $\Phi$ is not injective. Let $\eta$ be a nonzero vector in its kernel. All $x_0$-fibre sums of $\eta$ vanish, hence so does the total sum $\sum_{v\in V} \eta(v)$, being the sum of the $x_0$-fibre sums. The total sum is equally the sum of all $x_1$-fibre sums; all but the $c_0$-one vanish by construction, so the $c_0$-one vanishes too. Thus every fibre sum of $\eta$ vanishes and $\eta$ is an even certificate. ☐

**Corollary 4.2.** *For every class $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ one has* $\mathrm{ParDim}(\mathcal{H}) \geq \min(\mathrm{DSdim}(\mathcal{H}), 2)$*. In particular* $\mathrm{ParDim}(\mathcal{H}) = \mathrm{DSdim}(\mathcal{H})$ *whenever* $\mathrm{DSdim}(\mathcal{H}) \leq 2$.

*Proof.* Let $d = \min(\mathrm{DSdim}(\mathcal{H}), 2)$; we may assume $d \geq 1$. Choose $D$ with $|D| = d$ that is DS-shattered, and a pseudo-cube $Q \subseteq \mathcal{H}|_D$ on $D$. Choosing one preimage in $\mathcal{H}$ for each element of $Q$ gives a finite nonempty $\mathcal{H}' \subseteq \mathcal{H}$ with $\mathcal{H}'|_D = Q$. By Theorem 4.1, $Q$ carries an even certificate, so $\mathcal{H}'|_D$ is an even pseudo-cube on $D$ and $\mathrm{ParDim}(\mathcal{H}) \geq d$. The last sentence combines this with ParDim $\leq$ DSdim from (2). ☐

Theorem 4.1 is what the binary argument of [5, Corollary 3.4] does not give: there every fibre of the full cube has exactly two points and the all-one vector is a certificate, an argument that has no analogue once a fibre may have three points. The counting above replaces it, at the cost of restricting the number of coordinates rather than the alphabet. Corollary 4.2 also closes off the smallest conceivable gap in the chain (2), which by itself permits $\mathrm{ParDim}(\mathcal{H}) = 1$ with $\mathrm{DSdim}(\mathcal{H}) = 2$.

*Remark* 4.3. The same count, kept general, gives the following inequality, which we state as a remark because it is proved on paper and is not part of the formal development below. For a finite $V \subseteq \mathcal{Y}^D$ with $|D| = d$, let $f_x$ be the number of distinct $x$-fibres of $V$. Recording, for each coordinate, all fibre sums except one dropped fibre per coordinate after the first, and recovering each dropped sum from the total, shows that the space of vectors satisfying (1) has dimension at least $|V| - \sum_{x\in D} f_x + (d-1)$. Two consequences at $d = 3$: a pseudo-cube all of whose fibres have at least three points carries an even certificate (then $f_x \leq |V|/3$), and one carrying none has at least six fibres of size exactly two. Both witnesses below are consistent with the second: Li's 28-row table has 32 fibres of size two and the 24-point table of §5 has 28. Only the case $d \le 2$ of the inequality — Theorem 4.1 — is formally verified.

## 5. A 24-point witness at dimension three

Write $S$ for the following subset of $\{0,1,2\} \times \{0,1,2\} \times \{0,1,2,3\}$, given by listing for each pair $(a,b)$ of first two coordinates the set of third coordinates present.

| | $b=0$ | $b=1$ | $b=2$ |
|---|---|---|---|
| $a=0$ | $\{0,2\}$ | $\{0,1,3\}$ | $\{1,2,3\}$ |
| $a=1$ | $\{1,3\}$ | $\{0,1,2\}$ | $\{0,2,3\}$ |
| $a=2$ | $\{0,1,2,3\}$ | $\{2,3\}$ | $\{0,1\}$ |

Table 1. The 24-point set $S$: the cell $(a,b)$ lists the $c$ with $(a,b,c) \in S$.

**Theorem 5.1.** *$|S| = 24$, and $S$ is a pseudo-cube on its three coordinates. It is minimal: for every $v \in S$ the set $S \setminus \{v\}$ is not a pseudo-cube. No nonempty subset of $S$ carries an even certificate; in particular $S$ is a 3-dimensional pseudo-cube that is not an even pseudo-cube, and neither is any of its subclasses.*

*Proof.* The cardinality and the pseudo-cube property are read off Table 1: every cell $(a,b)$ has at least two entries, which supplies the third-coordinate flip at each of its points, while a first-coordinate flip at $(a,b,c)$ asks that $c$ occur in one of the two other cells of the column $b$, and a second-coordinate flip that it occur in one of the two other cells of the row $a$. Minimality is the verification, for each of the 24 points, that its deletion leaves some point without a flip. All of this is a finite check and was carried out exhaustively.

For the refutation we use Lemma 3.1: it suffices to show that the only even subset of $S$ is empty. Among the 33 fibres of $S$ there are 28 of size exactly 2, four of size 3 and one of size 4. A fibre $\{v,w\}$ of size two forces $v \in T \iff w \in T$ for every even $T \subseteq S$, since otherwise that fibre of $T$ would be a singleton. So an even subset is a union of connected components of the graph on $S$ whose edges are the size-two fibres. That graph is connected — a spanning tree of 23 edges is exhibited in the formal development — so the only candidates are $\emptyset$ and $S$ itself, and $S$ is not even because it has a fibre of size 3. By Lemma 3.1 no subset of $S$ carries an even certificate, and by Corollary 3.2 the same holds for every subclass. ☐

The refutation just given is what keeps the verification linear in the size of the table: it replaces a sweep over the $2^{24}$ subsets, or a rank computation, by 23 fibre identities and one parity. It also shows that the failure is not inherited from a smaller sub-object, since $S$ is a minimal pseudo-cube. The complementary control is that failure is not a property of dimension three either.

**Proposition 5.2.** *The binary cube $\{0,1\}^3$ is a 3-dimensional pseudo-cube and it carries an even certificate.*

*Proof.* Every coordinate-deletion fibre of $\{0,1\}^3$ has exactly two points, so the cube is an even set and the all-one vector is a certificate; being nonempty and even it is a pseudo-cube by the argument of Corollary 3.3. This is the binary case of [5, Corollary 3.4]. ☐

*Remark* 5.3 (Exhaustive searches; outside the formal development). The searches recorded here are exhaustive computations in C and are *not* part of the machine-checked development: what is checked by the proof kernel is the two witnesses themselves, not the exhaustion. They were implemented twice, independently, when the two witnesses were found, and every count below was reproduced once more by a third implementation written for the refereeing of this paper. For a grid $[k_1] \times [k_2] \times [k_3]$ a subset is even exactly when it meets every axis line in an even number of points, so "carries an even certificate" becomes "contains a nonempty even set", and the grid can be swept subset by subset. The measured counts are: 4 pseudo-cubes in the $2 \times 2 \times 3$ grid, 49 in $2 \times 3 \times 3$, 392 in $2 \times 3 \times 4$ and 21 739 in $3 \times 3 \times 3$, all of them carrying an even certificate; and 3 189 335 pseudo-cubes in $3 \times 3 \times 4$, of which exactly 216 carry none, the smallest having 24 points.

A second, alphabet-free search is a depth-first enumeration that grows a table from a single point: at each step it takes the first (point, coordinate) pair lacking a flip and adds the point obtained by changing that coordinate to each value already in use, or to one fresh value. Such an enumeration reaches every minimal pseudo-cube of at most the cap size, up to relabelling the values of each coordinate: if $Q$ is a minimal pseudo-cube then, after relabelling, the table built so far always lies inside $Q$, since the flip missing at the deficient pair exists in $Q$ and its value is either already in use or fresh; the process stops at a pseudo-cube contained in $Q$, which by minimality is $Q$ itself. As a pseudo-cube without an even certificate contains a minimal pseudo-cube without one (Corollary 3.2), a size-capped run settles all alphabets at once. With size caps $8, 12, 16, 18, 20, 22, 23$ the enumeration reaches respectively 1, 12, 339, 3 522, 33 712, 439 697 and 1 518 575 pseudo-cubes, with no exception at any cap. The cap-23 run took 7 612 seconds on one core in its original form, and under a minute once branches are cut by the bound that a table with $k$ points lacking a flip at one coordinate needs at least $k$ further points; the counts agree. Taken together with Theorem 5.1 these measurements say that the least size of a 3-dimensional pseudo-cube carrying no even certificate is exactly 24, over any alphabet. The completeness argument just given concerns isomorphism classes of tables and is not machine-checked.

Combining Theorem 4.1 with Theorem 5.1 gives the statement announced in the introduction: pseudo-cubes on one or two coordinates are always even, and there is a pseudo-cube on three coordinates that is not, so three is the least dimension at which parity dimension and DS dimension can differ.

## 6. Li's separating example, verified directly

We now record what the same tools say about the example of [5, Appendix D]. Everything in this section is the mathematics of [5]; what is added is that the verification is direct and exhaustive, and independent of [5, Theorem 3.5].

The construction uses tokens $\mathsf{P}_1, \mathsf{P}_2, \mathsf{N}_1, \mathsf{N}_2$ with signs $\beta(\mathsf{P}_i) = \mathsf{P}$, $\beta(\mathsf{N}_i) = \mathsf{N}$, rows $r \in \{0, 1, 2\}$, columns $s \in \{0, 1, 2, 3\}$, and the table $T_{r,s}$ of Table 2. The class is $\mathcal{H} = \{h_{\ell,r,s} : \ell \in T_{r,s}\}$, which has 28 hypotheses.

| | $s=0$ | $s=1$ | $s=2$ | $s=3$ |
|---|---|---|---|---|
| $r=0$ | $\mathsf{N}_1\mathsf{N}_2$ | $\mathsf{P}_1\mathsf{P}_2\mathsf{N}_1\mathsf{N}_2$ | $\mathsf{P}_1\mathsf{P}_2$ | $\mathsf{P}_2\mathsf{N}_2$ |
| $r=1$ | $\mathsf{P}_1\mathsf{P}_2$ | $\mathsf{N}_1\mathsf{N}_2$ | $\mathsf{P}_1\mathsf{N}_2$ | $\mathsf{P}_2\mathsf{N}_1$ |
| $r=2$ | $\mathsf{P}_1\mathsf{P}_2\mathsf{N}_1\mathsf{N}_2$ | $\mathsf{P}_1\mathsf{P}_2$ | $\mathsf{P}_2\mathsf{N}_2$ | $\mathsf{N}_1\mathsf{N}_2$ |

TABLE 2. The index table $T_{r,s} \subseteq \{\mathsf{P}_1, \mathsf{P}_2, \mathsf{N}_1, \mathsf{N}_2\}$ of [5, Appendix D].

Under the stopping rule of [5, Appendix D] the rollout of $h_{\ell,r,s}$ from the three root states $x_1 = (1, \varepsilon)$, $x_2 = (2, \varepsilon)$, $x_3 = (3, \varepsilon)$ returns $\beta(\ell)\ell$, $R_r$ and $C_s$ respectively, so after the injective relabelling of [5] the trace of the rollout class on these three states is

$$P = \{(\ell, r, s) : \ell \in T_{r,s}\} \subseteq \{\mathsf{P}_1, \mathsf{P}_2, \mathsf{N}_1, \mathsf{N}_2\} \times \{0, 1, 2\} \times \{0, 1, 2, 3\}.$$

Our transcription of $T_{r,s}$ reproduces all four of the derived supports $G_\ell = \{(r, s) : \ell \in T_{r,s}\}$ printed in [5, Appendix D], its three obstruction sets $E$, $F_{\mathsf{P}}$, $F_{\mathsf{N}}$, and the count $|\mathcal{H}| = 28$.

**Theorem 6.1** (the claim of [5, Appendix D], recomputed)**.** $|P| = 28$ *and $P$ is a pseudo-cube on its three coordinates. Moreover $P$ is minimal: for every $v \in P$ the set $P \setminus \{v\}$ is not a pseudo-cube.*

*Proof.* Exhaustive check on the 28 rows: for each row and each of the three coordinates, a flip is located; for minimality, each of the 28 deletions is checked to destroy the property. The first assertion is [5]'s own ("We claim that $P$ is a 3-pseudo-cube"), proved there by observing that every cell of $T_{r,s}$ has size at least two and that every row and column occurring in a $G_\ell$ has degree at least two. Minimality is not claimed there. □

**Theorem 6.2.** *No nonempty subset of $P$ carries an even certificate. Hence $P$ is a 3-dimensional pseudo-cube that is not an even pseudo-cube, and neither is any of its subsets.*

*Proof.* As in Theorem 5.1, via Lemma 3.1 and Corollary 3.2. The 38 fibres of $P$ have sizes 2 (thirty-two of them), 3 (four) and 4 (two); the size-two fibres connect all 28 rows, a spanning tree of 27 of them being exhibited in the formal development, so the only even subsets are $\emptyset$ and $P$, and $P$ has a fibre of odd size. □

This is the object [5, Corollary 3.7] asserts to exist and does not exhibit. Note the route: Theorem 6.2 uses neither [5, Theorem 3.5] nor the density bound of (2), so, with Remark 6.4, it is an independent confirmation of a corollary that [5] derives from a theorem whose proof also carries its main sample-complexity bound. The matching lower bound is equally explicit.

**Proposition 6.3.** *The restriction of $P$ to its first two coordinates carries an even certificate. Hence the largest number of coordinates on which some subset of $P$ is an even pseudo-cube is exactly 2, while $P$ itself is a pseudo-cube on all three of its coordinates.*

*Proof.* The four patterns $\{\mathsf{P}_1, \mathsf{N}_2\} \times \{0, 2\}$ occur in that restriction and form a $2 \times 2$ box, all of whose fibres have exactly two points; by Lemma 3.1 they give an even certificate, so the value is at least 2, and Theorem 6.2 gives at most 2. The last assertion is Theorem 6.1. □

*Remark* 6.4 (From the trace to the rollout class). Theorems 6.1 and 6.2 and Proposition 6.3 are statements about the finite table $P$, and they are what is machine-checked. They carry over to $\mathrm{Roll}_{\mathtt{halt}}(\mathcal{H})$ itself through one observation about the construction: the rollout class is constant at every state other than the three root states, because under the stopping rule of [5, Appendix D] "only rollouts started from root states $w = \varepsilon$ emit new actions; non-root start states halt immediately". A positive-dimensional pseudo-cube — hence also the support of an even certificate — cannot use a constant coordinate, so a pseudo-cube of the rollout class on three or more states, even or not, lives on the three root states, where every finite subclass restricts to a subset of $P$. Therefore $\mathrm{DSdim}(\mathrm{Roll}_{\mathtt{halt}}(\mathcal{H})) = 3$, where [5, Theorem 3.6] records $\geq 3$, and $\mathrm{ParDim}(\mathrm{Roll}_{\mathtt{halt}}(\mathcal{H})) = 2$, which is [5, Corollary 3.7] obtained without [5, Theorem 3.5]. This transfer is elementary, but it is not part of the formal development, which verifies only the statements about $P$.

For the base class $\mathcal{H}$ itself, [5, Appendix D] shows $\mathrm{DSdim}(\mathcal{H}) = 2$: the upper half by a case analysis proving that three row–column obstruction graphs are forests, the lower half by exhibiting four hypotheses. We verify both by different routes and add the parity half.

**Theorem 6.5.** *Let $x_0 = (1, \varepsilon)$, $x_{\mathsf{P}} = (1, \mathsf{P})$, $x_{\mathsf{N}} = (1, \mathsf{N})$, $x_R = (2, \varepsilon)$, $x_C = (3, \varepsilon)$ be the five states at which $\mathcal{H}$ is not constant. For every one of the 16 subsets of $\{x_0, x_{\mathsf{P}}, x_{\mathsf{N}}, x_R, x_C\}$ of size at least 3, the trace of $\mathcal{H}$ on that subset contains no nonempty pseudo-cube; consequently it carries no even certificate either.*

*Proof.* For each of the 16 subsets we exhibit an *elimination order* for the restricted class: a listing of its patterns in which every entry is, at the moment it is listed, alone in one of its fibres among the entries still to come. A class with an elimination order contains no nonempty pseudo-cube, by induction on the list: the first entry cannot lie in a pseudo-cube, since the neighbour its distinguished coordinate demands would have to come later and would share that fibre. The certificates are finite data and were checked exhaustively. The second assertion follows from Corollary 3.3. This is the upper half of [5, Theorem 3.6] verified by a route independent of its forest argument; the reduction of arbitrary base witnesses to three-state ones is [5]'s, using that a positive-dimensional pseudo-cube cannot use a constant coordinate and that projections of pseudo-cubes are pseudo-cubes [5, Lemma D.1(i)]. With that reduction, $\mathrm{DSdim}(\mathcal{H}) \leq 2$ and $\mathrm{ParDim}(\mathcal{H}) \leq 2$. □

**Proposition 6.6.** *The four hypotheses* $h_{\mathsf{P}_1,0,1}$, $h_{\mathsf{P}_1,1,0}$, $h_{\mathsf{N}_1,0,0}$, $h_{\mathsf{N}_1,1,1}$ *realise on* $(x_0, x_R)$ *the four pairs* $(\mathsf{P}, R_0)$, $(\mathsf{P}, R_1)$, $(\mathsf{N}, R_0)$, $(\mathsf{N}, R_1)$*; this* $2\times 2$ *box is a pseudo-cube and an even set. Hence* $\mathrm{DSdim}(\mathcal{H}) \geq 2$ *and* $\mathrm{ParDim}(\mathcal{H}) \geq 2$*, so with Theorem 6.5,* $\mathrm{ParDim}(\mathcal{H}) = \mathrm{DSdim}(\mathcal{H}) = 2$*.*

*Proof.* The four hypotheses and the four pairs are [5, Appendix D]'s; the box has all fibres of size exactly two, so it is even, and Lemma 3.1 and Corollary 3.3 give both halves. ☐

So on this example, with Remark 6.4, DS dimension jumps from 2 to 3 under rollout while parity dimension stays at 2 — the behaviour [5, Theorems 3.5 and 3.6] predict, here computed on both sides rather than deduced.

## 7. Relation to earlier work

The pseudo-cube and the DS dimension are due to Daniely and Shalev-Shwartz [3]. Brukhim, Carmon, Dinur, Moran and Yehudayoff [2] characterised multiclass PAC learnability through the DS dimension, resolving a question open since the pioneering work on multiclass PAC learning of the late 1980s; their definition of a pseudo-cube is the one used in [5] and here. Hanneke, Meng, Moran and Shaeiri [4] establish a sharp Sauer-type inequality for multiclass and list prediction expressed in terms of the DS dimension and its list extension, tight for every alphabet size; Pabbaraju [6] bounds the maximum hypergraph density of a multiclass class by its DS dimension, settling a conjecture of [3] and the optimal dependence of multiclass sample complexity on the DS dimension. None of these develops an $\mathbb{F}_2$ refinement of the notion: [2] discusses pseudo-cubes throughout and contains no occurrence of parity, $\mathbb{F}_2$, $GF(2)$, mod 2, homology or linear algebra; [4] likewise.

The even pseudo-cube and the parity dimension of a function class appear, as far as we can determine, only in [5], which is eight weeks old at the time of writing. A full-text search of a local mirror of the arXiv (373 shards, about 86 gigabytes of text, current to 3 September 2026) finds "even pseudo-cube" in that paper alone. The phrase "parity dimension" occurs in ten other papers of the mirror, in every case in an unrelated sense — chiefly the parity dimension of a graph, an $\mathbb{F}_2$-nullity invariant introduced by Amin, Clark and Slater [1] in connection with odd dominating sets, besides isolated uses in physics and representation theory — and "ParDim" occurs elsewhere only as a LaTeX macro name for unrelated quantities; none of these papers mentions pseudo-cubes or the DS dimension, which occur, in any sense, in 312 papers of the mirror. Two web searches for the notion return the same single source; the OpenAlex record of [5] shows no citing works; and a search of the OEIS, in both directions, for the counts of §5 and for the terms of this subject returns nothing. We therefore state the results of §4 and §5 as new, while noting that their proofs are elementary: what was missing was the question and the value, not the technique.

## 8. Verification

Every theorem, proposition, lemma and corollary stated above with a formal counterpart has been checked by the Lean 4 proof assistant against Mathlib; the development is organised in five modules — definitions and the normal form, the two witnesses, the base class, and the threshold theorem — and Appendix A reproduces the formal statement of each. Patterns are encoded as lists of natural numbers and coordinate deletion at $x$ as the substitution of a fixed value at position $x$; a separate finite check verifies, on the tables in question, that this really is the relation "agree at every coordinate other than $x$", so the encoding is not taken on trust. Every finite verification uses the kernel's own evaluator, and none uses compiled native evaluation: `#print axioms` on each of the twenty-seven exported results returns exactly `[propext, Classical.choice, Quot.sound]`, with no `sorryAx` and no native-evaluation axiom. The exhaustive searches of Remark 5.3, by contrast, are ordinary C programs and are outside this guarantee, as are the general counting inequality of Remark 4.3 and the transfer from the finite table $P$ to Li's rollout class in Remark 6.4; the repository is private, and a repository reference is to be added at submission.

## Appendix A. The formal statements

We reproduce the Lean 4 statements verbatim. The general theory is parameterised by a *key family* `key : ι → α → κ`: two patterns lie in the same fibre at the coordinate `x` exactly when their keys agree there, so that restrictions to a subset of the coordinates are instances of the same theorems. For the concrete tables the key is `dropAt x v = v.set x 0` and `restr` is the trace map onto a list of coordinates. Definitions first.

```
def fib (key : ι → α → κ) (V : Finset α) (x : ι) (v : α) : Finset α :=
  V.filter fun w => key x w = key x v
def IsPseudoCube (key : ι → α → κ) (K : Finset ι) (V : Finset α) : Prop :=
  ∀ v ∈ V, ∀ x ∈ K, ∃ w ∈ V, key x w = key x v ∧ w ≠ v
def HasEvenCert (key : ι → α → κ) (K : Finset ι) (V : Finset α) : Prop :=
  ∃ η : α → ZMod 2, (∃ v ∈ V, η v ≠ 0) ∧ ∀ x ∈ K, ∀ v ∈ V, ∑ w ∈ fib key V x v, η w = 0
def IsEvenSet (key : ι → α → κ) (K : Finset ι) (S : Finset α) : Prop :=
  ∀ x ∈ K, ∀ v ∈ S, Even (fib key S x v).card
def dropAt : ℕ → List ℕ → List ℕ := fun x v => v.set x 0
```

Lemma 3.1 and Corollaries 3.2 and 3.3:

```
theorem hasEvenCert_iff :
    HasEvenCert key K V ↔ ∃ S ⊆ V, S.Nonempty ∧ IsEvenSet key K S
theorem hasEvenCert_mono {S : Finset α} (hSV : S ⊆ V) (h : HasEvenCert key K S) :
    HasEvenCert key K V
theorem exists_pseudoCube_of_hasEvenCert (h : HasEvenCert key K V) :
    ∃ S ⊆ V, S.Nonempty ∧ IsPseudoCube key K S
```

Theorem 4.1 and Corollary 4.2, the latter in its finite-table form:

```
theorem hasEvenCert_of_dim_two [DecidableEq ι] (key : ι → α → κ) (x0 x1 : ι) (V : Finset α)
    (hpc : IsPseudoCube key {x0, x1} V) (hV : V.Nonempty) :
    HasEvenCert key {x0, x1} V
theorem hasEvenCert_of_ds_pair [DecidableEq ι] (key : ι → α → κ) (x0 x1 : ι) {V : Finset α}
    (h : ∃ S ⊆ V, S.Nonempty ∧ IsPseudoCube key {x0, x1} S) :
    ∃ S ⊆ V, S.Nonempty ∧ HasEvenCert key {x0, x1} S
```

Theorem 5.1, where `Sset` is the table of Table 1 and `K3 = {0, 1, 2}`, and Proposition 5.2, where `Box` is the binary cube:

```
theorem small_card : Sset.card = 24
theorem small_isPseudoCube : IsPseudoCube dropAt K3 Sset
theorem small_no_evenCert : ¬ HasEvenCert dropAt K3 Sset
theorem small_no_evenCert_subclass : ∀ V ⊆ Sset, ¬ HasEvenCert dropAt K3 V
theorem small_minimal :
    ∀ v ∈ Sset, ¬ IsPseudoCube dropAt K3 (Sset.erase v)
theorem box_hasEvenCert : HasEvenCert dropAt K3 Box ∧ IsPseudoCube dropAt K3 Box
```

Theorems 6.1 and 6.2 and Proposition 6.3, where `Pset` is the table $P$ and `Pset01` its restriction to the first two coordinates:

```
theorem paper_table_card : Pset.card = 28
theorem paper_isPseudoCube : IsPseudoCube key K3 Pset
theorem paper_minimal : ∀ v ∈ Pset, ¬ IsPseudoCube key K3 (Pset.erase v)
theorem paper_no_evenCert : ¬ HasEvenCert key K3 Pset
theorem paper_no_evenCert_subclass : ∀ V ⊆ Pset, ¬ HasEvenCert key K3 V
theorem exists_pseudoCube_without_evenCert :
    ∃ (V : Finset (List ℕ)), IsPseudoCube key K3 V ∧ V.card = 28 ∧ ¬ HasEvenCert key K3 V
theorem paper_pardim_ge_two : HasEvenCert key {0, 1} Pset01
```

Theorem 6.5 and Proposition 6.6, where `Bset` is the $28 \times 5$ table of the base class, `cases` the sixteen coordinate subsets of size at least three paired with their elimination orders, `T03` the $2 \times 2$ box and `B03` the trace on $(x_0, x_R)$:

```
theorem base_card : Bset.card = 28
theorem base_no_pseudoCube :
    ∀ p ∈ cases, ∀ S ⊆ Bset.image (restr p.1), S.Nonempty →
      ¬ IsPseudoCube dropAt (Finset.range p.1.length) S
theorem base_no_evenCert :
    ∀ p ∈ cases, ¬ HasEvenCert dropAt (Finset.range p.1.length) (Bset.image (restr p.1))
theorem base_dsdim_ge_two : IsPseudoCube dropAt {0, 1} T03
theorem base_pardim_ge_two : HasEvenCert dropAt {0, 1} B03
```

## References


[1] A. T. Amin, L. H. Clark and P. J. Slater, *Parity dimension for graphs*, Discrete Math. 187 (1998), no. 1–3, 1–17, doi:10.1016/S0012-365X(97)00242-2.

[2] N. Brukhim, D. Carmon, I. Dinur, S. Moran and A. Yehudayoff, *A characterization of multiclass learnability*, in: 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 943–955, doi:10.1109/FOCS54457.2022.00093; arXiv:2203.01550.

[3] A. Daniely and S. Shalev-Shwartz, *Optimal learners for multiclass problems*, in: Proceedings of The 27th Conference on Learning Theory (COLT 2014), Proceedings of Machine Learning Research 35, PMLR, 2014, pp. 287–316; arXiv:1405.2420.

[4] S. Hanneke, Q. Meng, S. Moran and A. Shaeiri, *An optimal Sauer lemma over $k$-ary alphabets*, arXiv:2604.12952 (v1, 14 April 2026).

[5] Z. Li, *The optimal sample complexity of learning autoregressive chain-of-thought*, arXiv:2607.07423 (v1, 8 July 2026).

[6] C. Pabbaraju, *The optimal sample complexity of multiclass and list learning*, arXiv:2604.24749 (v4, 18 August 2026).