Induced Lines in Hales-Jewett Cubes

Demetres Christofides*

Abstract

A line in $[n]^d$ is a set $\{x^{(1)}, \ldots, x^{(n)}\}$ of n elements of $[n]^d$ such that for each $1 \leqslant i \leqslant d$, the sequence $x_i^{(1)}, \ldots, x_i^{(n)}$ is either strictly increasing from 1 to n, or strictly decreasing from n to 1, or constant. How many lines can a set $S \subseteq [n]^d$ of a given size contain?

One of our aims in this paper is to give a counterexample to the Ratio Conjecture of Patashnik, which states that the greatest average degree is attained when $S = [n]^d$. Our other main aim is to prove the result (which would have been strongly suggested by the Ratio Conjecture) that the number of lines contained in S is at most $|S|^{2-\varepsilon}$ for some $\varepsilon > 0$.

We also prove similar results for combinatorial, or Hales-Jewett, lines, i.e. lines such that only strictly increasing or constant sequences are allowed.

1 Introduction

Although this paper is not concerned with game theory, we start with a brief discussion of Tic-Tac-Toe, since this motivates several of the conjectures that we discuss.

The n-in-a-row d-dimensional Tic-Tac-Toe, or just n^d-game, is defined as follows. The board X is $[n]^d$ where, as usual, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. Two players alternately pick points from X, with no point chosen more than once. The winner is the first player to choose all n points of a line, where a line is a set $\{x^{(1)}, \ldots, x^{(n)}\}$ of n elements of $[n]^d$ such that for each $1 \leqslant i \leqslant d$, the sequence $x_i^{(1)}, \ldots, x_i^{(n)}$ is either strictly increasing from 1 to n, or strictly decreasing from n to 1, or constant. (Note that, since we demand the line to have n elements, not all such sequences can be constant.) If all the points of X are chosen and no player can claim a line, then the game is a draw. For example, the 3^2-game is the traditional Tic-Tac-Toe (or Noughts and Crosses).

Trivially, the number of points in X is n^d. To count the number of lines, observe that every line is determined from the two points that extend the line linearly to $X' = \{0, 1, \ldots, n+1\}^d$. Since for every point of $X' \setminus X$, there exists a unique line

*Supported by grants from the Engineering and Physical Sciences Research Council and from the Cambridge Commonwealth Trust.

extending to it, and $|X' \setminus X| = (n+2)^d - n^d$, we deduce that the total number of lines is

$$\frac{(n+2)^d - n^d}{2}.$$

The strategy stealing argument (see e.g. [1]) shows that the first player has at least a drawing strategy. So the best thing the second player can do against a perfect opponent is to try not to lose. A way to achieve this would be via a 'pairing strategy'. Suppose that a subset of the board can be partitioned into disjoint pairs of points, such that every line contains at least one such pair. Then, whenever the first player picks one point from a pair, the second player replies by picking the other one. Otherwise, the second player picks any point he likes. This certainly guarantees a draw for the second player. Hales and Jewett [4] noted that, by Hall's theorem, a pairing strategy exists if and only if for every set of m lines, the total number of points contained in those lines is at least $2m$. Trivially, if the second player can force a draw via a pairing strategy, then there are at least twice as many points as lines. They conjectured that this condition is also sufficient:

Conjecture 1 (Pairing Conjecture [4]). *If $(n+2)^d \leqslant 2n^d$, then the second player can force a draw in the n^d-game by a pairing strategy.*

Note that for a pairing strategy to exist, it is necessary that $d \leqslant \frac{\log 2}{\log(1+2/n)} \sim \frac{log2}{2}n$. However, it is worth mentioning that Beck showed that the second player can force a draw for d much larger than the above bound. To be more precise, he has shown (see e.g. [2]) that the second player can force a draw whenever $d \leqslant c\frac{n^2}{\log n}$, for some constant $c > 0$.

Since the second player can force a draw even when a pairing strategy cannot exist, the following beautiful generalization of the Pairing Conjecture made by Patashnik [6], and popularized by Beck [3], is perhaps of more interest than the Pairing Conjecture itself.

Conjecture 2 (Ratio Conjecture [6]). *The ratio of the number of lines spanned by a set S of points in $[n]^d$ to the size of S is maximized when $S = [n]^d$.*

We begin Section 2 by considering some highly believable statements concerning the maximum number of lines contained in a set S. These are statements that would imply the Ratio Conjecture. Unfortunately, they turn out to be false. We then proceed to disprove the Ratio Conjecture in $[n]^d$ for every $n \geqslant 3$.

We also disprove the corresponding conjecture for Hales-Jewett lines (i.e. lines all of whose sequences are either strictly increasing or constant) in two ways; first by adapting our disproof of the Ratio Conjecture, and then by noting an interesting relationship between the Hales-Jewett lines in $[n]^d$ and the Tic-Tac-Toe lines in $[2n]^d$.

Trivially, a set S in $[n]^d$ cannot induce more than $|S|^2$ lines. On the other hand, our counterexamples to the Ratio Conjecture give sets S with at least $|S|^\lambda$ lines, for some $\lambda > \frac{\log(n+2)}{\log n}$. Our main aim in Section 3 is to bound this power away from 2.

We show that there is an $\varepsilon > 0$ such that a set S in $[n]^d$ cannot induce more than $|S|^{2-\varepsilon+o(1)}$ lines. In fact, for $n = 3$, we show that a set S in $[3]^d$ cannot induce more than $|S|^{1.8+o(1)}$ lines, while in the other direction we exhibit arbitrarily large sets S inducing at least $|S|^{1.7729\ldots+o(1)}$ lines. We prove similar results for Hales-Jewett lines.

2 Disproving the Ratio Conjecture

It will be convenient to consider elements of $[n]^d$ as lines, called constant lines, and also count each non-constant line twice, once for each of the two points of $X' \setminus X$ it determines. In this way, there is a natural one to one correspondence between elements of X' and lines in $[n]^d$. Denoting by $L(S)$ the number of lines under this convention spanned by the set S, we have $L(X) = (n+2)^d$. The Ratio Conjecture then says that for a set S in $[n]^d$ we have

$$\frac{L(S)}{|S|} \leqslant \left(1 + \frac{2}{n}\right)^d.$$

Firstly we discuss some stronger results that would imply the conjecture. For simplicity, we consider only the case $n = 3$; we will return to general n later. It seems reasonable to hope that for every set $S \subseteq [3]^d$, we have

$$L(S) \leqslant |S|^\lambda. \tag{1}$$

where $\lambda = \frac{\log 5}{\log 3}$. Note that this is tight when S is a ***complete subcube***, i.e. there is a partition

$$[d] = I_1 \cup \ldots \cup I_{d_1} \cup J_1 \cup \ldots \cup J_{d_2} \cup K$$

and elements $a_k \in [3]$ for each $k \in K$ such that

$$S = \left\{ (x_1, \ldots, x_n) \in [3]^d : \begin{array}{c} x_k = a_k \text{ if } k \in K, x_i = x_{i'} \text{ whenever } i, i' \in I_l \text{ for some } l \\ \text{and } x_i = 3 - x_j \text{ whenever } i \in I_l \text{ and } j \in J_l \text{ for some } l \end{array} \right\}.$$

If (1) were true we would have

$$\frac{L(S)}{|S|} \leqslant |S|^{\lambda-1} \leqslant 3^{d(\lambda-1)} \leqslant \left(\frac{5}{3}\right)^d,$$

implying the Ratio Conjecture. However, (1) is false in $[3]^4$. Figure 1 shows an example of 39 points in $[3]^4$ with $L(S) = 217 > 39^\lambda$.

The next matrix will hopefully help the reader verify our claim.

$$\begin{pmatrix} 17 & 7 & 1 & 7 & 17 \\ 7 & 17 & 1 & 17 & 7 \\ 1 & 1 & 17 & 1 & 1 \\ 7 & 17 & 1 & 17 & 7 \\ 17 & 7 & 1 & 7 & 17 \end{pmatrix}$$

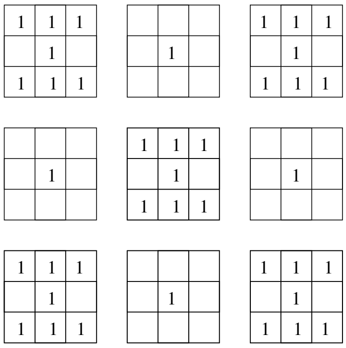


Figure 1: $|S| = 39$; $L(S) = 217 > 39^\lambda$

Each entry in the central 3×3 submatrix corresponds to the number of lines in one of the 9 squares. For example, the entry 17 in the second column of the fourth row of the matrix, says that there are exactly 17 lines in the bottom left square. The rest of the entries in the matrix, count 'cross-lines'. For example the entry 7 in the first column of the second row, says that there are exactly 7 lines having one element in each of the squares of the first row. Of course, since we count each line twice, there are 14 such lines. The other 7 are recorded in the fifth column of the second row. (The reader might wish to think of 'left to right' lines etc.)

A weaker result which would still imply the Ratio Conjecture would be that $L(S) \leqslant f(|S|)$ for some increasing function $f : [1, \infty) \to \mathbb{R}$ satisfying $f(3^d) = 5^d$. But remarkably, it is not even true that if $|S| = 3^k$, where k is an integer, then $L(S) \leqslant 5^k$. Figure 2 shows an example of $243 = 3^5$ points in $[3]^6$ with $L(S) = 3177 > 5^5$.

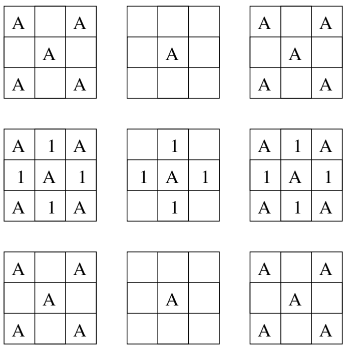
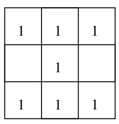


Figure 2: $|S| = 3^5$; $L(S) = 3177 > 5^5$

Here, A denotes the following set in $[3]^2$



and 1 denotes the set in $[3]^2$ containing just $(2,2)$. The corresponding matrix for Figure 2 is given below

$$\begin{pmatrix} 141 & 209 & 17 & 209 & 141 \\ 85 & 153 & 17 & 153 & 85 \\ 149 & 217 & 25 & 217 & 149 \\ 85 & 153 & 17 & 153 & 85 \\ 141 & 209 & 17 & 209 & 141 \end{pmatrix}$$

The above examples are not good enough to disprove the Ratio Conjecture. However, they suggest that the conjecture might be false, and in fact it is.

We begin by presenting a counterexample to the conjecture for the case $n = 3$. We remove all points from $[3]^d$ which have exactly k 1's and 3's, and $d - k$ 2's, for some k to be determined later. Clearly, the number of points removed is

$$\binom{d}{k} 2^k.$$

The number of lines removed is

$$\binom{d}{k} 2^k 3^{d-k} + \binom{d}{k} (4^k - 2^k).$$

Indeed, the first summand counts the number of lines removed with exactly k sequences of the form 1,1,1 and 3,3,3, while the second summand counts the number of lines removed with exactly k sequences of the form 1,1,1, 3,3,3, 1,2,3 and 3,2,1, with at least one of them being non-constant.

So to disprove the Ratio Conjecture it is enough to find an integer k such that

$$3^{d-k} + 2^k - 1 < \left(\frac{5}{3}\right)^d.$$

Note that it is enough to choose an $\alpha \in (0,1)$ such that $\frac{5}{3} > 2^\alpha, 3^{1-\alpha}$. Then, we will be done by putting $k = \lfloor \alpha d \rfloor$ for some large enough d. But any α in the interval $\left(\frac{\log(9/5)}{\log 3}, \frac{\log(5/3)}{\log 2}\right)$ will do. (Note that this is a non-empty interval!)

Note also that once we have a counterexample in dimension d, then we can get a counterexample in any dimension greater than d. Indeed given a set S in $[3]^d$ with $L(S)/|S| > (5/3)^d$, then $S' = [3] \times S$ satisfies $|S'| = 3|S|$ and $L(S') = 5L(S)$, whence $L(S')/|S'| > (5/3)^{d+1}$.

We do not know the smallest dimension d for which there is a counterexample. Our method gives $d \leqslant 8$. We did not try to optimize this. (Removing all points with the number of 1's and 3's in the interval $[k, k']$ can give even larger ratios, but does not give a counterexample for $d < 8$.) It is tedious but routine to check that d must be at least 4.

We now return to general n. Similar counterexamples work for every $n \geqslant 3$ provided d is large enough. Indeed, remove all points from $[n]^d$ with exactly k 1's and n's. The number of points removed is

$$\binom{d}{k} 2^k (n-2)^{d-k}$$

and the number of lines removed is

$$\binom{d}{k} 2^k n^{d-k} + \binom{d}{k} (n-2)^{d-k} (4^k - 2^k).$$

As before, it is enough to find an integer k such that

$$\left(1 + \frac{2}{n-2}\right)^{d-k} + 2^k - 1 < \left(1 + \frac{2}{n}\right)^d. \tag{2}$$

Working as above, one just needs to check that

$$1 - \frac{\log\left(1 + \frac{2}{n}\right)}{\log\left(1 + \frac{2}{n-2}\right)} < \frac{\log\left(1 + \frac{2}{n}\right)}{\log 2}.$$

This can be easily checked for $3 \leqslant n \leqslant 6$. For $n > 6$, one uses the fact that for $x > 0$,

$$x - \frac{x^2}{2} < \log(1 + x) < x,$$

to get

$$1 - \frac{\log\left(1 + \frac{2}{n}\right)}{\log\left(1 + \frac{2}{n-2}\right)} < \frac{\frac{4}{n^2-4}}{\frac{2}{n-2} - \frac{2}{(n-2)^2}} = \frac{2(n-2)}{n^2 - n - 6} < \frac{2}{n}$$

and

$$\frac{\log\left(1 + \frac{2}{n}\right)}{\log 2} > \frac{1}{\log 2}\left(\frac{2}{n} - \frac{2}{n^2}\right) = \frac{2}{n \log 2}\left(1 - \frac{1}{n}\right) > \frac{2}{n}.$$

We define $d(n)$ to be the smallest integer d for which there is a counterexample to the Ratio Conjecture in $[n]^d$. It would be interesting to know the asymptotic behaviour of $d(n)$. Refining the above counterexample we have:

Theorem 3. $d(n) = O(n \log n)$

Proof. It is enough to show that if $d = Cn \log n$ for some large enough C, and if $k = \alpha d$ for some $\alpha = \alpha(n)$ satisfying

$$1 - \frac{\log\left(1 + \frac{2}{n}\right)}{\log\left(1 + \frac{2}{n-2}\right)} < \alpha < \frac{\log\left(1 + \frac{2}{n}\right)}{\log 2},$$

then (2) holds. Note that, for example, we can take $\alpha = \frac{5}{2n}$ provided n is large enough. We will show that

$$\left(1 + \frac{2}{n}\right)^d \geqslant x 2^k, \; y \left(1 + \frac{2}{n-2}\right)^{d-k},$$

for some $x = x(n), y = y(n)$ satisfying $\frac{1}{x} + \frac{1}{y} = 1$. In fact, to simplify the calculations we will take in advance $x = n$ and $y = 1 + \frac{1}{n-1}$. But then, we may take any d satisfying

$$d \geqslant \frac{\log x}{\log\left(1 + \frac{2}{n}\right) - \alpha \log 2} \sim C_1 n \log n$$

and

$$d \geqslant \frac{\log y}{\log\left(1 + \frac{2}{n}\right) - (1-\alpha)\log\left(1 + \frac{2}{n-2}\right)} \sim C_2 n \qquad \square$$

We remark that to disprove the Pairing Conjecture, it would be enough to disprove the Ratio Conjecture in $[n]^d$ for some $d \leqslant \frac{\log 2}{\log(1+2/n)} \sim \frac{n \log 2}{2}$. In particular, if $d(n) = o(n)$, then the Pairing Conjecture would be false for every large enough n. On the other hand, if $d(n)$ grows faster than linearly, then not only would the Pairing Conjecture be true for large enough n, but moreover for every fixed integer m, there would be only finitely many n for which the Ratio Conjecture is false with the ratio of lines to points being at most m.

Finally, we consider the Ratio Conjecture for a slightly different definition of lines. In proving their famous theorem, Hales and Jewett [4] considered only ***combinatorial lines***; non-constant lines with all their coordinate sequences either constant, or strictly increasing from 1 to n. For example, in $[3]^2$, there are only 7 combinatorial lines, since we are not considering $\{(1,3),(2,2),(3,1)\}$ as a line. The corresponding conjecture for combinatorial lines is also false. Note that there are exactly $(n+1)^d - n^d$ such lines in $[n]^d$. As before, it will be simpler to also count the constant lines, giving us a total of $(n+1)^d$ lines. By removing all points with exactly k 1's, we remove

$$\binom{d}{k}(n-1)^{d-k}$$

points and

$$\binom{d}{k}n^{d-k} + \binom{d}{k}(n-1)^{d-k}(2^k - 1)$$

lines. To disprove the conjecture we need to choose a k such that

$$\left(1 + \frac{1}{n-1}\right)^{d-k} + 2^k - 1 < \left(1 + \frac{1}{n}\right)^d.$$

To finish the counterexample, one may proceed exactly as before, or just note that this is (2) with n replaced by $2n$.

Given this argument, one might wonder whether the fact that the Ratio Conjecture for the usual Tic-Tac-Toe lines is false for $2n$ directly implies that the Ratio Conjecture for combinatorial lines is false for n. This is actually 'almost' true. Firstly, note that the ratio of lines to points in $[2n]^d$ is equal to the ratio of combinatorial lines to points in $[n]^d$. Note also that there is a one to one correspondence between subsets S of $[n]^d$ and subsets S' of $[2n]^d$ which are invariant under reflections along each hyperplane of the form $x_i = n$. Indeed, to get from S to S' one takes the union of all reflections of S along those hyperplanes, while to get from S' to S one just restricts to $[n]^d$. For example, the set of all elements with exactly k 1's in $[n]^d$, corresponds to the set of all elements with exactly k 1's and $2n$'s in $[2n]^d$. Moreover, what is important is that every point of S corresponds to 2^d points of S', and every combinatorial line of S corresponds to 2^d Tic-Tac-Toe lines of S'. This is immediate for the points and the 'non-diagonal' lines and it is easy to check for the 'diagonal lines'. (Note that mere translations would not work for the diagonal lines.) The fact that our counterexample for Tic-Tac-Toe lines in $[2n]^d$ is 'symmetric' is now enough to complete this argument.

3 Upper Bounds on the Number of Induced Lines

What is the maximum number of lines a set S in $[n]^d$ of a given size can contain? Observe that every ordered pair of points of S determines at most one line, whence $L(S) \leqslant |S|^2$. How small can one hope to make the exponent here?

For simplicity, we begin by considering the case $n = 3$. Recall that in Section 2, we have exhibited sets S with $L(S) > |S|^{\log 5/\log 3}$. Moreover, as we have already shown, if S provides a counterexample to the Ratio Conjecture, then $L(S) > |S|^{\log 5/\log 3}$. However, observing that in our counterexamples of the Ratio Conjecture, we have removed only an asymptotically small fraction of the number of points, it is natural to ask whether $L(S) \leqslant |S|^{\frac{\log 5}{\log 3}+o(1)}$ holds for every set S.

However, a moment's thought shows that this is false. Indeed, let T be any set in $[3]^d$ with $L(T) = |T|^\lambda$ for some $\lambda > \frac{\log 5}{\log 3}$, and consider $T^k \subseteq [3]^{kd}$. The fact that $|T^k| = |T|^k$ and $L(T^k) = L(T)^k$, shows that there are arbitrarily large sets S with $L(S) = |S|^\lambda$. For example, our counterexample to the Ratio Conjecture (with $d = 8$ and $k = 5$) is a set with 4769 elements inducing 286689 lines, thus showing that the exponent cannot be reduced to anything less than $\frac{\log 286689}{\log 4769} = 1.4836\ldots$.

We start this section by showing that in fact, a much larger exponent is needed.

Theorem 4. *For every $\varepsilon > 0$, there is a positive integer d depending on ε and a set $S \subseteq [3]^d$, such that $L(S) \geqslant |S|^{1+\alpha-\varepsilon}$ where $\alpha = 0.7729078\ldots$ is the unique solution of*

$$\alpha \log 2\alpha + (1-\alpha)\log(1-\alpha) = 0$$

in $(0, 1)$.

Proof. Consider $S \subseteq [3]^d$ consisting of all points which contain only 1's and 3's and all points with at least k 2's, where $k = k(d)$ is the largest integer such that

$$\binom{d}{k}\frac{1}{2^k} \geqslant 1.$$

The number of points in S is

$$2^d + \sum_{i=k}^{d}\binom{d}{i}2^{d-i}.$$

Note that $k \geqslant d/2$, so for $k \leqslant i \leqslant d$ we have

$$\frac{\binom{d}{i+1}2^{d-i-1}}{\binom{d}{i}2^{d-i}} = \frac{d-i}{2(i+1)} \leqslant \frac{d-k}{2k} \leqslant \frac{1}{2}.$$

It follows that

$$|S| \leqslant 2^d + 2\binom{d}{k}2^{d-k} < 4\binom{d}{k}2^{d-k}.$$

Note also that, (ignoring the constant lines,)

$$L(S) > 2^d \sum_{i=k}^{d} \binom{d}{i} > 2^d \binom{d}{k} \geqslant 2^{d+k}.$$

We claim that, for any $\varepsilon' > 0$, $(\alpha - \varepsilon')d < k < (\alpha + \varepsilon')d$. Indeed suppose $k \geqslant \alpha' d$, for some $\alpha' > \alpha$ independent of d. By Stirling's approximation we have

$$\begin{aligned}\binom{d}{k}\frac{1}{2^k} &< \left(\frac{1}{2\alpha'}\right)^{\alpha' d}\left(\frac{1}{1-\alpha'}\right)^{(1-\alpha')d} \\ &= \exp\left\{-d(\alpha' \log 2\alpha' + (1-\alpha')\log(1-\alpha'))\right\} < 1,\end{aligned}$$

if d is large enough, a contradiction. Similarly, if $k \leqslant \alpha' d$ for some $\alpha' < \alpha$ independent of d, then

$$\binom{n}{k+1}\frac{1}{2^{k+1}} \geqslant \frac{c}{\sqrt{d}}\left(\frac{1}{2\alpha'}\right)^{\alpha' d}\left(\frac{1}{1-\alpha'}\right)^{(1-\alpha')d} > 1,$$

if d is large enough, contradicting the maximality of k. We deduce that for any $\varepsilon'' > 0$ and any large enough d, that $L(S)^{1+\varepsilon''} \geqslant |S|^{(1+\alpha-\varepsilon'')}$, as required. □

Given this result, it is very natural to ask whether there is an $\varepsilon > 0$ such that $L(S) \leqslant |S|^{2-\varepsilon+o(1)}$. We now prove this fact. We first study the related question for combinatorial lines, since the idea of the argument is the same but the proof is much cleaner. We write $L'(S)$ for the number of combinatorial lines spanned by the set S.

Theorem 5. *For every set $S \subseteq [3]^d$, the number of combinatorial lines spanned by S is at most $|S|^{3/2}$.*

We wish to prove Theorem 5 by induction. However, a direct approach by induction does not seem to work. The key is to generalise the statement to a stronger result which is more amenable to an inductive approach. To this end, given sets S_1, S_2, S_3 in $[3]^d$ we denote by $L'(S_1, S_2, S_3)$ the number of (combinatorial) lines in $[3]^{d+1}$ which contain exactly one point from each of $\{1\} \times S_1$, $\{2\} \times S_2$ and $\{3\} \times S_3$. Note that $L'(S) = L'(S, S, S)$. Hence Theorem 5 will follow from:

Theorem 5′. Given sets S_1, S_2, S_3 in $[3]^d$ of sizes s_1, s_2 and s_3 we have

$$L'(S_1, S_2, S_3) \leqslant (s_1 s_2 s_3)^{1/2}.$$

Proof. Induction on the dimension, the result being immediate for $d = 1$. For the induction step, let us be given $S_1, S_2, S_3 \subseteq [3]^d$. Partition each S_i into $(\{1\} \times S_{i1}) \cup (\{2\} \times S_{i2}) \cup (\{3\} \times S_{i3})$ and let s_{ij} denote the size of S_{ij}. Since

$$L'(S_1, S_2, S_3) = L'(S_{11}, S_{21}, S_{31}) + L'(S_{12}, S_{22}, S_{32}) + L'(S_{13}, S_{23}, S_{33}) + L'(S_{11}, S_{22}, S_{33}),$$

by the induction hypothesis it is enough to prove that

$$((s_{11} + s_{12} + s_{13})(s_{21} + s_{22} + s_{23})(s_{31} + s_{32} + s_{33}))^{1/2} \geqslant (s_{11}s_{21}s_{31})^{1/2} + (s_{12}s_{22}s_{32})^{1/2} + (s_{13}s_{23}s_{33})^{1/2} + (s_{11}s_{22}s_{33})^{1/2}. \quad (3)$$

We claim that we may assume $s_{21} = s_{31} = 0$. Indeed, by replacing s_{21} and s_{31} by 0, and s_{22} and s_{33} by $s_{21} + s_{22}$ and $s_{31} + s_{33}$ respectively, then the left hand side of (3) remains the same, while the right hand side increases by at least

$$(s_{11})^{1/2}((s_{21} + s_{22})^{1/2}(s_{31} + s_{33})^{1/2} - (s_{21}s_{31})^{1/2} - (s_{22}s_{33})^{1/2})$$

which is non-negative by Cauchy-Schwarz. Similarly, we may assume that $s_{12} = s_{32} = 0$ and $s_{13} = s_{23} = 0$. But then the inequality becomes trivial. □

We now prove upper bounds for general $n \geqslant 3$, but still for combinatorial lines. The proof is essentially the same as above. Instead of the Cauchy-Schwarz inequality, we will need to make use of the following extension of Hölder's inequality, see e.g. [5, Theorem 2.8.3].

Lemma 6. *Let* a_{ij} $(1 \leqslant i \leqslant n;\ 1 \leqslant j \leqslant 2)$ *be positive numbers and let* $p_1, \ldots, p_n$ *be positive numbers such that* $\frac{1}{p_1} + \ldots + \frac{1}{p_n} \geqslant 1$. *Then*

$$(a_{11}^{p_1} + a_{12}^{p_1})^{\frac{1}{p_1}} \cdots (a_{n1}^{p_n} + a_{n2}^{p_n})^{\frac{1}{p_n}} \geqslant a_{11} \cdots a_{n1} + a_{12} \cdots a_{n2}.$$

Theorem 7. *For every set* $S \subseteq [n]^d$, *the number of combinatorial lines spanned by* S *is at most* $|S|^{n/(n-1)}$.

Proof. With the obvious extension of notation, given sets $S_1, \ldots, S_n$ in $[n]^d$, it is enough to prove that

$$L'(S_1, \ldots, S_n) \leqslant (s_1 \cdots s_n)^{1/(n-1)}.$$

By the induction hypothesis this reduces to

$$\prod_{i=1}^{n} \left(\sum_{j=1}^{n} s_{ij} \right)^{\frac{1}{n-1}} \geqslant \sum_{j=1}^{n} \left(\prod_{i=1}^{n} s_{ij} \right)^{\frac{1}{n-1}} + \left(\prod_{i=1}^{n} s_{ii} \right)^{\frac{1}{n-1}}. \quad (4)$$

Fix a j, and replace each s_{ij} $(i \neq j)$ by 0 and each s_{ii} $(i \neq j)$ by $s_{ii} + s_{ij}$. Then the left hand side of (4) remains the same, while the right hand side increases by at least

$$(s_{jj})^{\frac{1}{n-1}} \left(\left(\prod_{i \neq j} (s_{ii} + s_{ij}) \right)^{\frac{1}{n-1}} - \left(\prod_{i \neq j} s_{ii} \right)^{\frac{1}{n-1}} - \left(\prod_{i \neq j} s_{ij} \right)^{\frac{1}{n-1}} \right),$$

which by Lemma 6 is non-negative. Hence, we may assume that s_{ij} is 0 for every $i \neq j$. This completes the proof of the inequality and hence of the theorem. □

Note that by similar reasoning as in the beginning of this section, we cannot reduce the exponent $n/(n-1)$ in Theorem 7 to anything lower than $\frac{\log(n+1)}{\log n}$.

We now turn our attention back to the usual Tic-Tac-Toe lines and our main aim of showing that $L(S) \leqslant |S|^{2-\varepsilon+o(1)}$ for some $\varepsilon > 0$. We begin yet again with the case $n = 3$. The idea of the proof is similar to the proof for combinatorial lines, however the analogue of (3) does not seem that easy to prove.

Theorem 8. *For every set $S \subseteq [3]^d$ we have $L(S) \leqslant 2|S|^{9/5}$.*

Given sets S_1, S_2, S_3 in $[3]^d$ we denote by $L(S_1, S_2, S_3)$ the number of (Tic-Tac-Toe) lines in $[3]^{d+1}$ which contain exactly one point from each of $\{1\} \times S_1$, $\{2\} \times S_2$ and $\{3\} \times S_3$. Note that this time $L(S) = 2L(S, S, S)$. Hence, Theorem 5 will follow from:

Theorem 8′. Given sets S_1, S_2, S_3 in $[3]^d$ of sizes s_1, s_2 and s_3 we have

$$L(S_1, S_2, S_3) \leqslant (s_1 s_2 s_3)^{3/5}.$$

Proof. Induction on the dimension, the result being immediate for $d = 1$. Partition S_1, S_2 and S_3 as before. By the induction hypothesis, it is enough to prove that

$$((s_{11} + s_{12} + s_{13})(s_{21} + s_{22} + s_{23})(s_{31} + s_{32} + s_{33}))^{3/5} \geqslant \\ (s_{11}s_{21}s_{31})^{3/5} + (s_{12}s_{22}s_{32})^{3/5} + (s_{13}s_{23}s_{33})^{3/5} + (s_{11}s_{22}s_{33})^{3/5} + (s_{13}s_{22}s_{31})^{3/5}.$$

Using Lemma 6 we may assume that $s_{12} = s_{32} = 0$. By scaling we may also assume that $s_{11} + s_{13} = s_{21} + s_{22} + s_{23} = s_{31} + s_{33} = 1$. It remains to prove that

$$(s_{11}s_{21}s_{31})^{3/5} + (s_{13}s_{23}s_{33})^{3/5} + (s_{11}s_{22}s_{33})^{3/5} + (s_{13}s_{22}s_{31})^{3/5} \leqslant 1. \tag{5}$$

To handle this, we make use of the following inequality.

Lemma 9. *Let a, b, c be non-negative real numbers satisfying $a + b + c = 1$ and let $0 < \lambda < 1$. Suppose A, B, C are non-negative. Then,*

$$Aa^\lambda + Bb^\lambda + Cc^\lambda \leqslant \left(A^{\frac{1}{1-\lambda}} + B^{\frac{1}{1-\lambda}} + C^{\frac{1}{1-\lambda}}\right)^{1-\lambda}.$$

Proof. We may assume that at least one of A, B, C is non-zero. Consider the Lagrangian

$$L(a, b, c; \mu) = Aa^\lambda + Bb^\lambda + Cc^\lambda - \mu(a + b + c - 1).$$

For every fixed μ, this is maximized when

$$\frac{A}{a^{1-\lambda}} = \frac{B}{b^{1-\lambda}} = \frac{C}{c^{1-\lambda}} = \mu.$$

If for example $a = 0$, this should be understood as saying that $A = 0$, etc. Choosing

$$\mu = \left(A^{\frac{1}{1-\lambda}} + B^{\frac{1}{1-\lambda}} + C^{\frac{1}{1-\lambda}}\right)^{1-\lambda},$$

we deduce that L is maximized when

$$a = \frac{A^{\frac{1}{1-\lambda}}}{A^{\frac{1}{1-\lambda}} + B^{\frac{1}{1-\lambda}} + C^{\frac{1}{1-\lambda}}};\ b = \frac{B^{\frac{1}{1-\lambda}}}{A^{\frac{1}{1-\lambda}} + B^{\frac{1}{1-\lambda}} + C^{\frac{1}{1-\lambda}}};$$

$$\text{and } c = \frac{C^{\frac{1}{1-\lambda}}}{A^{\frac{1}{1-\lambda}} + B^{\frac{1}{1-\lambda}} + C^{\frac{1}{1-\lambda}}}.$$

Since these values of a, b, c satisfy the constraint, the original expression is maximized there. The lemma follows. □

Applying the lemma to (5), we deduce that it is enough to show

$$(s_{11}s_{31})^{3/2} + (s_{13}s_{33})^{3/2} + ((s_{11}s_{33})^{3/5} + (s_{13}s_{31})^{3/5})^{5/2} \leqslant 1.$$

Note that this holds if any of the variables is 0. By putting $x = \frac{s_{13}}{s_{11}} = \frac{1-s_{11}}{s_{11}}$, $y = \frac{s_{33}}{s_{31}} = \frac{1-s_{31}}{s_{31}}$ and noting that $s_{11} = \frac{1}{x+1}$ and $s_{31} = \frac{1}{y+1}$, it is enough to show that

$$1 + (xy)^{3/2} + (x^{3/5} + y^{3/5})^{5/2} \leqslant ((x+1)(y+1))^{3/2},$$

where x, y are positive reals. Putting $z = y/x$ (note that we have assumed $x \neq 0$), and then replacing x by x^2 and z by z^{10} to get rid of (most of) the rational powers we are left with

$$(1 + z^{15}x^6 + (z^6+1)^{5/2}x^3)^2 \leqslant (z^{10}x^4 + (z^{10}+1)x^2 + 1)^3$$

Subtracting the left hand side from the right hand side and dividing by x^2 we are left with proving that

$$(3z^{30} + 3z^{20})x^8 - 2(z^6+1)^{5/2}z^{15}x^7 + (3z^{30} + 9z^{20} + 3z^{10})x^6 + (9z^{20} + 9z^{10})x^4$$
$$-(5z^{24}+10z^{18}+2z^{15}+10z^{12}+5z^6)x^4+(3z^{20}+9z^{10}+3)x^2-2(z^6+1)^{5/2}x+(3z^{10}+3) \geqslant 0.$$

We will be finished with several applications of the Arithmetic-Geometric Mean inequality. Firstly, we claim that

$$(3z^{30} + 3z^{20})x^8 + (z^{30} + 4z^{20} + z^{10})x^6 \geqslant 2(z^6+1)^{5/2}z^{15}x^7.$$

Indeed by AM-GM it is enough to show that

$$3z^{30} + 15z^{25} + 15z^{20} + 3z^{15} \geqslant z^{30} + 5z^{27} + 10z^{24} + 10z^{21} + 5z^{18} + z^{15}.$$

The following four inequalities show that this holds:

$2z^{30} + 3z^{25} \geqslant 5z^{27}$; $8z^{25} + 2z^{20} \geqslant 10z^{24}$; $2z^{25} + 8z^{20} \geqslant 10z^{21}$ and $3z^{20} + 2z^{15} \geqslant 5z^{18}$.

Similarly,

$$(3z^{10} + 3) + (z^{20} + 4z^{10} + z)x^2 \geqslant 2(z^6+1)^{5/2}x.$$

Finally, it remains to prove that

$$(2z^{30}+5z^{20}+2z^{10})x^6+(2z^{20}+5z^{10}+2)x^2+(9z^{20}+9z^{10})x^4 \geqslant \\ (5z^{24}+10z^{18}+2z^{15}+10z^{12}+5z^6)x^4$$

But the left hand side is at least

$$(4z^{25}+9z^{20}+10z^{15}+9z^{10}+4z^5)x^4,$$

hence the following inequalities

$$4z^{25}+z^{20}\geqslant 5z^{24};\ 6z^{20}+4z^{15}\geqslant 10z^{18};\ 4z^{15}+6z^{10}\geqslant 10z^{12} \text{ and } z^{10}+4z^5\geqslant 5z^6$$

complete the proof of Theorem 8. □

We have shown that for $S\subseteq [3]^d$, we have $L(S)\leqslant |S|^{1.8+o(1)}$, while the result is not true if 1.8 is replaced by $1.7729078\ldots$. We actually believe that 1.8 can be replaced by $3x = 1.7755024\ldots$, where

$$x=\frac{\log(1+\sqrt{3})}{\log(2+2\sqrt{3})}$$

is the real root of the equation

$$2+2^{\frac{1}{1-x}}=4^{\frac{x}{1-x}}.$$

Unfortunately, we have not been able to do so. This would be proved if one could show that the left hand side in (5) with 3/5 replaced by x, is maximized when $s_{11}=s_{13}=s_{31}=s_{33}=1/2$. However, we see no analogous reason to believe what the 'right' value for the exponent in Theorem 8 should be.

Given all this trouble to find an upper bound for the case $n=3$, finding upper bounds for general n might seem a difficult task. However it is intuitively believable that one should have even better bounds for general n. It is therefore not surprising that by adapting the above proof one can show that $L(S)\leqslant |S|^{1.8+o(1)}$ for every $S\subseteq [n]^d$ and every $n\geqslant 3$.

Theorem 10. *Let $n\geqslant 3$ be a positive integer. Then for every set S in $[n]^d$ we have $L(S)\leqslant 2|S|^{9/5}$.*

Proof. This will follow by proving the following two facts:

1. $L(S_1,\ldots,S_{2m+1})\leqslant (s_1 s_{m+1} s_{2m+1})^{3/5}$ if $n=2m+1$ is odd;
2. $L(S_1,\ldots,S_{2m})\leqslant (\min\{s_1 s_m s_{2m}, s_1 s_{m+1} s_{2m}\})^{3/5}$ if $n=2m$ is even.

The proofs of these statements are very similar to the proof of Theorem 8, and so we omit them. □

Acknowledgements

The author wishes to thank Prof. Imre Leader for many helpful discussions concerning the results in this paper and also for commenting on earlier drafts of the paper.

References

[1] J. Beck, Foundations of positional games, Random Structures Algorithms **9** (1996), 15–47.

[2] J. Beck, Tic-Tac-Toe, in: *Contemporary combinatorics*, János Bolyai Math. Soc., Budapest, (2002), 93–137.

[3] J. Beck, Positional games, Combin. Probab. Comput. **14** (2005), 649–696.

[4] A. W. Hales and R. I. Jewett, Regularity and positional games, Trans. Amer. Math. Soc. **106** (1963), 222–229.

[5] D. S. Mitrinović, *Analytic inequalities*, Springer, New York, 1970.

[6] O. Patashnik, Qubic: $4 \times 4 \times 4$ tic-tac-toe, Math. Mag. **53** (1980), 202–216.

Demetres Christofides

University of Cambridge
Department of Pure Mathematics and Mathematical Statistics
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WB
United Kingdom

D.Christofides@dpmms.cam.ac.uk