Combinatorics, Probability and Computing (2017) 26, 720–745. © Cambridge University Press 2017
doi:10.1017/S0963548317000098

Sets in Almost General Position

LUKA MILIĆEVIĆ

Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK
(e-mail: lm497@cam.ac.uk)

Received 4 May 2015; revised 1 August 2016; first published online 18 April 2017

Erdős asked the following question: given n points in the plane in almost general position (no four collinear), how large a set can we guarantee to find that is in general position (no three collinear)? Füredi constructed a set of n points in almost general position with no more than $o(n)$ points in general position. Cardinal, Tóth and Wood extended this result to $\mathbb{R}^3$, finding sets of n points with no five in a plane whose subsets with no four points in a plane have size $o(n)$, and asked the question for higher dimensions: for given n, is it still true that the largest subset in general position we can guarantee to find has size $o(n)$? We answer their question for all d and derive improved bounds for certain dimensions.

2010 *Mathematics subject classification*: Primary 52C35
Secondary 52C10

1. Introduction

A set of points in the plane is said to be *in general position* if it has no three collinear points, and *in almost general position* if there are no four collinear points. Let $\alpha(n)$ be the maximum k such that any set of n points in the plane in almost general position has k points in general position. Erdős [2] asked for an improvement of the (easy) bounds $\sqrt{2n-1} \leqslant \alpha(n) \leqslant n$ (see equation (13) in the paper). This was done by Füredi [3], who proved $\Omega(\sqrt{n \log n}) \leqslant \alpha(n) \leqslant o(n)$.

Cardinal, Tóth and Wood [1] considered the problem in $\mathbb{R}^3$. First, let us generalize the notion of general position. A set of points in $\mathbb{R}^d$ is said to be *in general position* if there are no $d+1$ points in the same hyperplane, and *in almost general position* if there are no $d+2$ points in the same hyperplane. Let $\alpha(n,d)$ be the maximum integer k such that every set of n points in $\mathbb{R}^d$ in almost general position contains a subset of k points in general position. Cardinal, Tóth and Wood proved that $\alpha(n,3) = o(n)$. They noted that for a fixed $d \geqslant 4$, only $\alpha(n,d) \leqslant Cn$ is known, for a constant $C \in (0,1)$, and they asked whether $\alpha(n,d) = o(n)$. The goal of this paper is to answer their question in all dimensions. In particular we prove the following.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Theorem 1.1. *For a fixed integer $d \geqslant 2$, we have $\alpha(n,d) = o(n)$.*

In fact, we are able to get better bounds for certain dimensions. This is the content of the next theorem.

Theorem 1.2. *Suppose that $d, m \in \mathbb{N}$ satisfy $2^{m+1} - 1 \leqslant d \leqslant 3 \cdot 2^m - 3$. Let $N \geqslant 1$. Then*

$$\alpha(2^N, d) \leqslant \left(\frac{25}{N}\right)^{1/2^{m+1}} 2^N.$$

It is worth noting the lower bound $\alpha(n,d) = \Omega_d((n \log n)^{1/d})$ due to Cardinal, Tóth and Wood [1]. However, we do not try to improve their bound in this paper.

Füredi [3] used the density Hales–Jewett theorem [4, 5] to establish $\alpha(n) = \alpha(n,2) = o(n)$. Here we reproduce his argument. By the density Hales–Jewett theorem, for a given $\epsilon > 0$, there is a positive integer N such that all subsets of $[3]^N$ of density ϵ contain a combinatorial line. Map the set $[3]^N$ to $\mathbb{R}^2$ using a generic linear map f to obtain a set $X = f([3]^N) \subseteq \mathbb{R}^2$. By the choice of f, collinear points in X correspond to collinear points in $[3]^N$, and f restricted to $[3]^N$ is injective. Therefore, X has no four points on a line, and so is in almost general position, but if $S \subseteq X$ has size at least $\epsilon|X|$, the set $f^{-1}(S) \subseteq [3]^N$ has density at least ϵ in $[3]^N$. Therefore, $f^{-1}(S)$ contains a line, hence $S = f(f^{-1}(S))$ has three collinear points. Since $\epsilon > 0$ was arbitrary, this proves that $\alpha(n,2) = o(n)$.

If one tries to generalize this argument to higher dimensions, by mapping $[m]^N$ to $\mathbb{R}^d$, then there will be m^{d-1} cohyperplanar points, and we would need $m^{d-1} = d+1$ to get almost general position. But the only positive integers that have this property are $(m,d) \in \{(3,2),(2,3)\}$. Taking $m = 2, d = 3$ gives $\alpha(n,3) = o(n)$, as observed by Cardinal, Tóth and Wood [1]. For other choices of (m,d) we have too many cohyperplanar points, as $m^{d-1} > d+1$. Overcoming this obstacle is the main goal of our paper.

Notation. Throughout this paper we write $[k]$ for the set $\{1,2,\ldots,k\}$. By a *d-cube of edge length k* we mean a set of the form

$$\{v_0 + \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_d v_d : \lambda_1, \lambda_2, \ldots, \lambda_d \in \{0,1,2,\ldots,k\}\},$$

where $v_1, v_2, \ldots, v_d$ are (not necessarily independent) vectors in a real vector space. A *d-dimensional combinatorial subspace* of $[N]^k$ is a set that consists of all $x \in [N]^k$ such that $x_i = a_i$ when $i \in I_0$, and x_i does not change when i ranges over I_j, for $j = 1,\ldots,d$, where $I_0, I_1 \ldots, I_d$ are some fixed sets that partition $[k]$, $I_1, I_2, \ldots, I_d$ are non-empty and a_i for $i \in I_0$ are some fixed elements of $[N]$. Given vectors $v_0, v_1, \ldots, v_r \in \mathbb{R}^d$, we say that they are *affinely dependent* if there are $\lambda_0, \lambda_1, \ldots, \lambda_r \in \mathbb{R}$ not all zero but adding up to zero, such that $\sum_{i=0}^r \lambda_i v_i = 0$. Finally, for a real vector space $\mathbb{R}^n$ we use the usual inner product $\langle x, y\rangle$ given by $\sum_{i=1}^n x_i y_i$, where x_i and y_i are coordinates of x and y with respect to the standard basis.

1.1. Organization of the paper

Section 2 is devoted to some motivation for the arguments of this paper, and to an overview of the approach taken in the proofs of the main results. In Section 3 we

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

introduce the key notion of this paper, $\mathcal{F}$-*incident sets*, where $\mathcal{F}$ is an arbitrary family of maps from $\mathbb{R}^N$ to $\mathbb{R}^d$. Roughly, these are the sets that stay cohyperplanar under all maps in $\mathcal{F}$. In the same section we prove Proposition 3.1, which provides us with an incidence removal function, meaning a single function which makes all the non-$\mathcal{F}$-incident sets non-cohyperplanar.

In Section 4 we focus on the study of $\mathcal{F}_{N,d,m}$-incident sets, where $\mathcal{F}_{N,d,m}$ is a family of maps from $\mathbb{R}^N$ to $\mathbb{R}^d$ resembling polynomials of degree m. In particular, Lemma 4.1 shows that the combinatorial subspaces and lines of a d-cube give rise to span $\mathcal{F}_{N,d,m}$-incident sets. The rest of the section is devoted to deriving a characterization of $\mathcal{F}_{N,d,m}$-incident sets in terms of vectors given by products of coordinates. The proof of $\alpha(n,d) = o_d(n)$ is the result of work carried out in Section 5. That section also contains Lemma 5.1, which is the main tool used in the analysis of $\mathcal{F}_{N,d,m}$-incident sets. Finally, in Section 6, we improve the bounds for certain dimensions, using Lemma 6.3 in the analysis of $\mathcal{F}_{N,d,m}$-incident sets.

2. Motivation and outline of the proof

Recall that the main obstacle to generalizing Füredi's argument to higher dimensions is that $(d-1)$-cubes have too many cohyperplanar points. A (seemingly) possible way to get around this issue would be to modify the initial set $[m]^N$ to a subset X, which does not have too many incidences, and yet some form of the Hales–Jewett theorem may still be applied to X. The desired set would once again be the image of X under a generic linear map from $\mathbb{R}^N \to \mathbb{R}^d$. It is tempting to try to remove certain points from each $(d-1)$-cube inside $[m]^N$, so that precisely $d+1$ out of the original m^{d-1} points remain. However, this is impossible for sufficiently large N, as the set $X \subseteq [m]^N$ gives a 2-colouring of $[m]^N$ (a point is blue if it is in X, red otherwise), and thus the Hales–Jewett theorem provides us with a monochromatic $(d-1)$-cube. Therefore, such an approach at least needs further modification, if it can be made to work at all.

Having abandoned this first idea, it is natural to try to map $[m]^N$ using a map f which is more general than a linear map. Previously we used a generic linear map, in other words, a map f with the property that the only sets of size $d+1$ whose images under f are cohyperplanar in $\mathbb{R}^d$ are precisely the affinely dependent subsets of $[m]^N$ of size $d+1$. This leads us to the key notion of this paper, namely that of $\mathcal{F}$-*incident sets*, which we now define. Let $\mathcal{F}$ be a family of functions from $\mathbb{R}^N$ to $\mathbb{R}^d$ (that we shall use instead of just linear maps). We say that a set $S \subseteq \mathbb{R}^N$ is $\mathcal{F}$-*incident* if the multiset $f(S)$ is affinely dependent for all $f \in \mathcal{F}$. Crucially, as for linear maps, we can a find a 'generic' map $f \in \operatorname{span} \mathcal{F}$, such that if $f(S)$ is affinely dependent then S is $\mathcal{F}$-incident. This is the content of Proposition 3.1. We refer to such a map as an 'incidence removal function'.

Once we have constructed an incidence removal function, the next aim is to study $\mathcal{F}$-incident sets for suitable $\mathcal{F}$. Our goal now is essentially the following: we want each dense subset of $[m]^N$ to contain an $\mathcal{F}$-incident set of size $d+1$ (which gives cohyperplanar sets with $d+1$ elements), but at the same time we want the image of $[m]^N$ under an incidence removal function not to contain $d+2$ cohyperplanar points. An easy way to fulfil the second requirement is to make sure that $\mathcal{F}$-incident sets of size $d+1$ cannot have intersection of size d. On the other hand, as in the case of linear maps, we use the density

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Hales–Jewett theorem to show that dense subsets contain the desired $\mathcal{F}$-incident sets, so we want the combinatorial subspaces to be (span $\mathcal{F}$)-incident; note that here we need a stronger property of being (span $\mathcal{F}$)-incident instead of just $\mathcal{F}$-incident, as the incidence removal function belongs to a bigger family span $\mathcal{F}$.

To give an idea of how we choose the family of functions $\mathcal{F}$ making the combinatorial lines span $\mathcal{F}$-incident, observe the following identities that hold for arbitrary a, b:

$$\begin{aligned}\mathbf{1}\cdot 1+(-\mathbf{3})\cdot 1+\mathbf{3}\cdot 1+(-\mathbf{1})\cdot 1&=0,\\ \mathbf{1}\cdot a+(-\mathbf{3})\cdot(a+b)+\mathbf{3}\cdot(a+2b)+(-\mathbf{1})\cdot(a+3b)&=0,\\ \mathbf{1}\cdot a^2+(-\mathbf{3})\cdot(a+b)^2+\mathbf{3}\cdot(a+2b)^2+(-\mathbf{1})\cdot(a+3b)^2&=0.\end{aligned}$$

What is crucial here is that we have the same coefficients appearing in the three linear combinations above. Hence, if one looks at a function $f:\mathbb{R}^N\to\mathbb{R}^3$ of the form

$$f(x)=\begin{pmatrix}(\langle x,v_1\rangle+c_1)^2\\ (\langle x,v_2\rangle+c_2)^2\\ (\langle x,v_3\rangle+c_3)^2\end{pmatrix} \tag{2.1}$$

for some $v_1,v_2,v_3\in\mathbb{R}^N$ and reals c_1,c_2,c_3, then $f(x),f(x+y),f(x+2y),f(x+3y)$ are necessarily coplanar, as

$$\mathbf{1}\cdot f(x)+(-\mathbf{3})\cdot f(x+y)+\mathbf{3}\cdot f(x+2y)+(-\mathbf{1})\cdot f(x+3y)=0$$

and the sum of coefficients is zero. Moreover, if g is any linear combination of functions of the form described above, then $g(x),g(x+y),g(x+2y),g(x+3y)$ are coplanar, owing to the fact that the same coefficients appear in the identities above.

In the case of $d=2$, we used linear maps only, and we had that the image of $[3]^N$ to the plane under a generic linear map is the desired set. In that case, the combinatorial lines in $[3]^N$ gave us collinear sets of points in the plane. Moving to the functions constructed from polynomials of degree 2, the image of $[4]^N$ under a 'generic degree 2 function' to $\mathbb{R}^3$ has cohyperplanar sets of four points that are also images of combinatorial lines. After some analysis of the $\mathcal{F}$-incident sets for $\mathcal{F}$ given by equation (2.1), we are able to show that those sets have intersection of size at most 1, provided the size of sets is at most 4. The motivation for this step comes from the heuristics that we expect our non-trivial $\mathcal{F}$-incident sets to be precisely the relevant combinatorial subspaces (in this case the lines), and as such they cannot have large intersection (in the case of lines, they cannot share more than one point). This was the second requirement that we had, and completes our sketch of the proof that $\alpha(n,3)=o(n)$. This naturally extends to larger values of d.

Using different identities, we are able to get better bounds on $\alpha(n,d)$. For example, the fact that

$$\begin{aligned}&x^2+(x+a+b)^2+(x+a+c)^2+(x+b+c)^2\\ &\quad=(x+a)^2+(x+b)^2+(x+c)^2+(x+a+b+c)^2\end{aligned}$$

holds for all x,a,b,c enables us to use three-dimensional combinatorial subspaces of $\{0,1\}^N$ as sources of cohyperplanar sets. Generalizing this identity to higher degrees, we can use the higher-dimensional combinatorial subspaces as well. The better bounds on $\alpha(n,d)$

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

when these subspaces are used come from the better bounds for density Hales–Jewett theorem [5] in the case of $\{0,1\}^N$ (which is the generalized Sperner's theorem).

When it comes to the analysis of $\mathcal{F}$-incident sets, let us first define precisely the families of functions that we shall consider in this paper. For given $N,d,m \in \mathbb{N}$ we define the family $\mathcal{F}_{N,d,m}$ of functions $f : \mathbb{R}^N \to \mathbb{R}^d$ of the form $f_i(x) = (\langle x, u_i\rangle + c_i)^l$ for $i = 1,2,\dots,d$, for any $u_1,u_2,\dots,u_d \in \mathbb{R}^N$, $c_1,c_2,\dots,c_d \in \mathbb{R}$ and $1 \leqslant l \leqslant m$. The bulk of this paper consists of studying the $\mathcal{F}_{N,d,m}$-incident sets.

2.1. Analysis and properties of $\mathcal{F}_{N,d,m}$-incident sets

The first important claim regarding $\mathcal{F}_{N,d,m}$-incident sets is the characterization given by Proposition 4.3, which we explain here. To simplify the notation, we introduce the following notion. The terminology ($\leqslant m$)-*function to* S stands for any function $\varphi : A \to S$, where A has size at most m. Given a vector $x \in \mathbb{R}^N$ and a ($\leqslant m$)-function φ to $[N]$, we define $\varphi(x) = \prod_{a\in A} x_{\varphi(a)}$. Proposition 4.3 tells us that $\{x_0,x_1,\dots,x_r\}$ for $r \leqslant d$ is $\mathcal{F}_{N,d,m}$-incident if and only if the vectors

$$\begin{pmatrix}\varphi_1(x_0)\\ \varphi_2(x_0)\\ \vdots\\ \varphi_r(x_0)\end{pmatrix}, \quad \begin{pmatrix}\varphi_1(x_1)\\ \varphi_2(x_1)\\ \vdots\\ \varphi_r(x_1)\end{pmatrix}, \dots, \begin{pmatrix}\varphi_1(x_r)\\ \varphi_2(x_r)\\ \vdots\\ \varphi_r(x_r)\end{pmatrix}$$

are affinely dependent for all ($\leqslant m$)-functions $\varphi_1,\varphi_2,\dots,\varphi_r$. Then, in order to prove that our $\mathcal{F}_{N,d,m}$-incident sets cannot have large intersections, we use Lemmas 5.1 and 6.3. First we state Lemma 5.1 to illustrate its contrast to Proposition 4.3.

Lemma 2.1 (Lemma 5.1). *Let $m,r,N \in \mathbb{N}$. Suppose that $y_1,y_2,\dots,y_r \in \mathbb{R}^N$ are vectors such that* $\operatorname{rank}\{y_1,y_2,\dots,y_r\} + m - 1 \geqslant r$. *Suppose further that $y_1,y_2,\dots,y_r$ are distinct and have non-zero coordinates. Then we may find ($\leqslant m$)-functions $\varphi_1,\varphi_2,\dots,\varphi_r$ for which the vectors*

$$\begin{pmatrix}\varphi_1(y_1)\\ \varphi_2(y_1)\\ \vdots\\ \varphi_r(y_1)\end{pmatrix}, \quad \begin{pmatrix}\varphi_1(y_2)\\ \varphi_2(y_2)\\ \vdots\\ \varphi_r(y_2)\end{pmatrix}, \dots, \begin{pmatrix}\varphi_1(y_r)\\ \varphi_2(y_r)\\ \vdots\\ \varphi_r(y_r)\end{pmatrix}$$

are linearly independent.

The conclusions of Lemma 5.1 and Proposition 4.3 are almost exactly opposite. We only need to take care to pass from affine dependence to linear dependence, which is easy as a set $\{v_1,v_2,\dots,v_r\}$ is linearly dependent if and only if $\{0,v_1,v_2,\dots,v_r\}$ is affinely dependent. Hence, we use these two results along with some combinatorial arguments to deduce structural information about $\mathcal{F}_{N,d,m}$-incident sets, and in particular to show that such sets exhibit behaviour similar to combinatorial subspaces, as expected.

We now state Lemma 6.3, another crucial result in the study of $\mathcal{F}_{N,d,m}$-incidence. Note that it has a more combinatorial flavour than Lemma 5.1.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Lemma 2.2 (Lemma 6.3). *Let $m,k \in \mathbb{N}$. Given any distinct sets $X_1, X_2, \dots, X_r \in \mathbb{N}^{(<\omega)}$, we can find sets $S_1, S_2, \dots, S_r \subseteq \mathbb{N}^{(\leqslant m)}$ such that the matrix*

$$I = \begin{pmatrix} \mathbb{1}_{S_1 \subseteq X_1} & \mathbb{1}_{S_1 \subseteq X_2} & \cdots & \mathbb{1}_{S_1 \subseteq X_r} \\ \mathbb{1}_{S_2 \subseteq X_1} & \mathbb{1}_{S_2 \subseteq X_2} & \cdots & \mathbb{1}_{S_2 \subseteq X_r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{1}_{S_r \subseteq X_1} & \mathbb{1}_{S_r \subseteq X_2} & \cdots & \mathbb{1}_{S_r \subseteq X_r} \end{pmatrix}$$

has $\dim \ker I = 0$ *if* $r < 2^{m+1}$, *and* $\dim \ker I \leqslant 1$ *if* $r < 3 \cdot 2^m$. *Here* $\mathbb{1}_{A \subseteq B}$ *takes the value* 1 *if* A *is a subset of* B, *and* 0 *otherwise.*

The subset relation here is actually quite natural. This comes from considering the $\mathcal{F}_{N,d,m}$-incident subsets of $\{0,1\}^N$. As we have seen above, when analysing $\mathcal{F}_{N,d,m}$-incidence, we are interested in the values of $\varphi(x)$ for $x \in \{0,1\}^N$ and $(\leqslant m)$-function φ to $[N]$. If we set $S = \operatorname{Im} \varphi$, then $\varphi(x) = \prod_{s \in S} x_s = \mathbb{1}_{S \subseteq X}$, where $X = \{i : x_i = 1\}$.

Finally, we note that studying the algebraic properties of such matrices may be of independent interest.

3. Definition and basic properties of $\mathcal{F}$-incidences

Throughout this section, $\mathcal{F}$ will stand for a family of maps from $\mathbb{R}^N$ to $\mathbb{R}^d$. Given such a family of functions $\mathcal{F}$, our goal is to understand the non-trivial affinely dependent sets of points in the images of $f \in \mathcal{F}$.

We say that points $s_1, s_2, \dots, s_k \in \mathbb{R}^d$ (not necessarily distinct) are *affinely dependent* if there are $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ not all zero such that $\sum_{i=1}^k \lambda_i = 0$ and $\sum_{i=1}^k \lambda_i s_i = 0$. A k-tuple $S = (s_1, s_2, \dots, s_k)$ of points in $\mathbb{R}^N$ is said to be *$\mathcal{F}$-incident* if for all $f \in \mathcal{F}$ we have $f(s_1), f(s_2), \dots, f(s_k)$ affinely dependent. A set $S = \{s_1, s_2, \dots, s_k\}$ of points in $\mathbb{R}^N$ is *$\mathcal{F}$-incident* if a corresponding k-tuple $(s_1, s_2, \dots, s_k)$ is. Note that the order does not play a role in this definition, nor is the order of points important at any point in this paper. The only reason why we use k-tuples is the possibility that some of $f(s_i)$ might overlap. (This issue could also be resolved by using multisets.) Further, S is *minimal $\mathcal{F}$-incident* if it is $\mathcal{F}$-incident and no proper subset of S is $\mathcal{F}$-incident.

Proposition 3.1 (incidence removal function). *Let $X \subseteq \mathbb{R}^N$ be a finite set and let $\mathcal{F}$ be a family of functions from $\mathbb{R}^N$ to $\mathbb{R}^d$. Then there is $f \in \operatorname{span} \mathcal{F}$ with the property that*

if $\{s_1, s_2, \dots, s_k\}$ is not $\mathcal{F}$-incident, then $f(s_1), f(s_2), \dots, f(s_k)$ are affinely independent. (†)

Furthermore, if $\mathcal{F}$ separates the points of X (i.e. for distinct $x, y \in X$ there is $f \in \mathcal{F}$ such that $f(x) \neq f(y)$), then there is $f \in \operatorname{span} \mathcal{F}$ which is injective on X, with the property (†).

The proof of the proposition is based on simple linear algebra and some easy facts regarding the vanishing of polynomials. It could be skipped if the reader wishes to: the reader need only be aware of the existence of the incidence removal function and its properties.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Proof. Throughout this proof, for a function f and a set $S = \{s_1, s_2, \ldots, s_k\}$, we regard $f(S)$ as a multiset of elements $f(s_1), \ldots, f(s_k)$. So, if we say that $f(S)$ is affinely dependent, we mean that $f(s_1), f(s_2), \ldots, f(s_k)$ are affinely dependent.

We start by establishing the existence of a map f with property (†). The second part of the proposition will follow from a simple argument later. Let $T_1, T_2, \ldots, T_m$ be the list of all subsets of X which are not $\mathcal{F}$-incident. Thus, for each index i we have a function $f_i \in \mathcal{F}$ such that $f_i(T_i)$ is affinely independent. We shall inductively construct functions $F_i \in \operatorname{span} \mathcal{F}$ such that all of $F_i(T_1), F_i(T_2), \ldots, F_i(T_i)$ are affinely independent. Start by taking $F_1 = f_1$ for the case $i = 1$.

Suppose that we have $i \geqslant 1$ and a function $F_i \in \operatorname{span} \mathcal{F}$ such that each of

$$F_i(T_1), F_i(T_2), \ldots, F_i(T_i)$$

is affinely independent. Assume that $i < m$, otherwise we are done. Also, if $F_i(T_{i+1})$ is already affinely independent, simply take $F_{i+1} = F_i$. Hence, without loss of generality $F_i(T_{i+1})$ is affinely dependent. We shall construct F_{i+1} as a linear combination $F_i + \lambda f_{i+1}$, where $\lambda > 0$ is a sufficiently small real, chosen so that it does not introduce new dependencies.

Let $u_1, u_2, \ldots, u_k \in \mathbb{R}^N$. Let $F^{(\lambda)} = F_i + \lambda f_{i+1}$ and suppose that

$$F^{(0)}(u_1), F^{(0)}(u_2), \ldots, F^{(0)}(u_k)$$

are affinely independent. Then

$$F^{(0)}(u_2) - F^{(0)}(u_1), \ldots, F^{(0)}(u_k) - F^{(0)}(u_1)$$

are linearly independent.

Lemma 3.2. *Suppose that $v_1, \ldots, v_l \in \mathbb{R}^d$ are linearly independent. Then, we can find $I \subseteq [d]$ of size l such that $v_1, \ldots, v_l$ are still linearly independent when restricted in coordinates in I.*

Proof. Look at the $d \times l$ matrix $A = (v_1 v_2 \cdots v_l)$. Since $v_1, v_2, \ldots, v_l$ are linearly independent, the column rank of A is l. But the column rank is the same as the row rank, so we can find l linearly independent rows with indices $r_1, \ldots, r_l$. Take $I = \{r_1, \ldots, r_l\}$ and let A' be the matrix A restricted to rows in I. Then, the row rank of A' is l, so its column rank is l, as desired. □

By Lemma 3.2 we can find a set of coordinates I of size $k-1$ such that $F^{(0)}(u_2) - F^{(0)}(u_1)$, $F^{(0)}(u_3) - F^{(0)}(u_1), \ldots, F^{(0)}(u_k) - F^{(0)}(u_1)$ are linearly independent after restriction to I. We restrict our attention to these coordinates only. Then we can define

$$p(\lambda) = \det(F^{(\lambda)}(u_2) - F^{(\lambda)}(u_1) \cdots F^{(\lambda)}(u_k) - F^{(\lambda)}(u_1)),$$

which is a polynomial in λ. Since $p(0) \neq 0$, by continuity we have $\delta > 0$ such that if $|\lambda| < \delta$ then $p(\lambda) \neq 0$. Therefore, $F^{(\lambda)}(u_1), F^{(\lambda)}(u_2), \ldots, F^{(\lambda)}(u_k)$ are affinely independent if $|\lambda| < \delta$. Note that we can now remove the restriction to coordinates of I, as this will not affect the affine independence.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

We can apply this argument to all $T_1, \dots, T_i$, to get $\delta > 0$ such that if $|\lambda| < \delta$ then $(F_i + \lambda f_{i+1})(T_j)$ is affinely independent for all $j = 1, \dots, i$.

Now suppose that the choice $F_i + \lambda f_{i+1}$ does not work for us as F_{i+1}. Then we must have that $(F_i + \lambda f_{i+1})(T_{i+1})$ is affinely dependent for all $|\lambda| < \delta$. Thus if $\lambda > \delta^{-1}$ then $(\lambda F_i + f_{i+1})(T_{i+1})$ is affinely dependent. List the elements of T_{i+1} as $t_1, t_2, \dots, t_r$. Now, apply Lemma 3.2 to $f_{i+1}(T_{i+1})$ to get a set of coordinates of size $r-1$, on which this set is still affinely independent, and use a similar polynomial to the one we had before, this time

$$q(\lambda) = \det((\lambda F_i + f_{i+1})(t_2 - t_1) \cdots (\lambda F_i + f_{i+1})(t_r - t_1)).$$

Then $q(0) \neq 0$, but $q(\lambda) = 0$ if $\lambda > \delta^{-1}$, which is a contradiction, and thus the first part of the proposition is proved.

For the last part, if $\mathcal{F}$ separates the points of X, observe that there are no two-element sets which are $\mathcal{F}$-incident. Hence, whenever $x, y \in X$ are distinct, their images $f(x)$ and $f(y)$ are affinely independent by the first part, so f is injective, as desired. □

4. Families of higher-degree maps and the resulting incident sets

Throughout the rest of the paper we will focus on the family $\mathcal{F}_{N,d,m}$ of functions $f : \mathbb{R}^N \to \mathbb{R}^d$ of the form $f_i(x) = (\langle x, u_i \rangle + c_i)^l$ for $i = 1, 2, \dots, d$, for any $u_1, u_2, \dots, u_d \in \mathbb{R}^N$, $c_1, c_2, \dots, c_d \in \mathbb{R}$ and $1 \leqslant l \leqslant m$.

We start by giving the examples of non-trivial span $\mathcal{F}_{N,d,m}$-incident sets. The proofs are based on algebraic identities, as described in the Introduction. For the case of lines, we use the rank-nullity theorem to prove that there is an identity we are looking for, and in the case of combinatorial subspaces, we prove the identity explicitly.

Lemma 4.1 (examples of non-trivial span $\mathcal{F}_{N,d,m}$-incident sets).

(1) (*Lines.*) *For* $x, y \in \mathbb{R}^N$, *the* $(m+2)$*-tuple* $(x + iy : \ i = 0, 1, \dots, m+1)$ *is* span $\mathcal{F}_{N,d,m}$*-incident.*

(2) (*Combinatorial subspaces.*) *For* $x_0, x_1, \dots, x_{m+1} \in \mathbb{R}^N$, *the* 2^{m+1}*-tuple*

$$\left(x_0 + \sum_{i \in I} x_i : I \subseteq [m+1] \right)$$

is span $\mathcal{F}_{N,d,m}$*-incident.*

Proof. *Lines.* We show that there are $\lambda_0, \dots, \lambda_{m+1}$, not all zero, such that for all $f \in \mathcal{F}_{N,d,m}$ we have $\sum_{i=0}^{m+1} \lambda_i f(x + iy) = 0$ and $\sum_{i=0}^{m+1} \lambda_i = 0$. Then, the same linear combination shows that $f(x), f(x+y), \dots, f(x + (m+1)y)$ are affinely dependent for $f \in \operatorname{span} \mathcal{F}_{N,d,m}$. Before we proceed with the proof, observe that if $y = 0$, our line becomes degenerate and the $(m+2)$-tuple we consider is immediately $\mathcal{F}_{N,d,m}$-incident. Thus, we assume that $y \neq 0$.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Thus, we want non-trivial λ_i to add up to zero, such that for all $u \in \mathbb{R}^N, c \in \mathbb{R}, l \in [m]$ we have

$$\sum_{i=0}^{m+1} \lambda_i (\langle x + iy, u \rangle + c)^l = 0.$$

Expanding this expression using the binomial theorem and treating it as a polynomial in c, it becomes equivalent to

$$\sum_{i=0}^{m+1} \lambda_i \langle x + iy, u \rangle^l = 0$$

for all $u \in \mathbb{R}$ and $l = 0, 1, \ldots, m$. Expanding further, this is equivalent to

$$\sum_{k=0}^{l} \binom{l}{k} \langle x, u \rangle^{l-k} \langle y, u \rangle^k \left(\sum_{i=0}^{m+1} \lambda_i i^k \right) = 0$$

for all $u \in \mathbb{R}$ and $l = 0, 1, \ldots, m$. Hence, if $\lambda_0, \ldots, \lambda_{m+1}$ satisfy

$$\sum_{i=0}^{m+1} \lambda_i i^l = 0$$

for all $l = 0, 1, \ldots, m$, we are done. But by the rank-nullity theorem ('more variables than equations'), we must have a non-trivial solution to these equations, giving us the desired λ_i.

Combinatorial subspaces. As in the case of lines, we show that there are coefficients λ_I, for set-valued indices $I \subseteq [m+1]$ (including the $I = \emptyset$), not all zero but adding up to zero, such that

$$\sum_{I \subseteq [m+1]} \lambda_I f\left(x_0 + \sum_{i \in I} x_i \right) = 0, \quad \text{for all } f \in \mathcal{F}_{N,d,m},$$

which suffices to prove the claim in full generality. In this case, we can actually set $\lambda_I = (-1)^{|I|}$.

It is enough to show that for any $u \in \mathbb{R}^N, c \in \mathbb{R}, l \in [m]$ we have

$$\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(\langle x_0 + \sum_{i \in I} x_i, u \rangle + c \right)^l = 0$$

(in these sums, $I = \emptyset$ is included). But writing $a_0 = \langle x_0, u \rangle + c, a_i = \langle x_i, u \rangle$ for $i = 1, \ldots, m+1$, we see that it is sufficient to show

$$\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(a_0 + \sum_{i \in I} a_i \right)^l = 0$$

for all $a_0, a_1, \ldots, a_{m+1} \in \mathbb{R}, l \in [m+1]$. This is the content of the next lemma.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Lemma 4.2. *Let $l, m \in \mathbb{N}, l \leqslant m$ and $a_0, a_1, \ldots, a_{m+1} \in \mathbb{R}$. Then*

$$\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(a_0 + \sum_{i \in I} a_i \right)^l = 0.$$

Proof of Lemma 4.2. Note that

$$\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(a_0 + \sum_{i \in I} a_i \right)^l = \sum_{k=0}^{l} \binom{l}{k} a_0^{l-k} \left(\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(\sum_{i \in I} a_i \right)^k \right),$$

so we only need to consider the case $a_0 = 0$. Proving the lemma in this case would tell us that

$$\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(\sum_{i \in I} a_i \right)^k = 0$$

holds for all k, so the whole expression with arbitrary a_0 above would vanish.

Consider the expression

$$\sum_{I \subseteq [m+1]} (-1)^{|I|} \left(\sum_{i \in I} a_i \right)^l$$

as a polynomial of degree l in $a_1, \ldots, a_{m+1}$. The coefficient of $a_1^{d_1} a_2^{d_2} \cdots a_{m+1}^{d_{m+1}}$ is

$$\binom{l}{d_1, d_2, \ldots, d_{m+1}} \sum_{S \subseteq I \subseteq [m+1]} (-1)^{|I|},$$

where S is the set of indices i such that $d_i > 0$. Since $|S| \leqslant m$, the sum $\sum_{S \subseteq I \subseteq [m+1]} (-1)^{|I|}$ is zero, which finishes the proof. □

Applying Lemma 4.2 completes the proof. □

Before coming to a key proposition which describes the $\mathcal{F}_{N,d,m}$-incident sets, we introduce a couple of pieces of notation. If φ is a function from a set of size at most m to a set X, we say that φ is a $(\leqslant m)$-*function* to X. Given a $(\leqslant m)$-function $\varphi : A \to [N]$ and $x \in \mathbb{R}^N$, we write $\varphi(x) = \prod_{a \in A} x_{\varphi(a)}$. Here we allow an 'empty' function, that is, a function φ from an empty set to $[N]$, defined by $\varphi(x) = 1$, for all $x \in \mathbb{R}^N$.

Proposition 4.3. *Let $r, d, m, N \in \mathbb{N}$, suppose $r \leqslant d$ and let $X = \{x_0, x_1, \ldots, x_r\} \subseteq \mathbb{R}^N$. The following are equivalent.*

(i) *X is $\mathcal{F}_{N,d,m}$-incident.*

(ii) *Given any $(\leqslant m)$-functions $\varphi_1, \varphi_2, \ldots, \varphi_r$ to $[N]$, the vectors*

$$\begin{pmatrix} \varphi_1(x_0) \\ \varphi_2(x_0) \\ \vdots \\ \varphi_r(x_0) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(x_1) \\ \varphi_2(x_1) \\ \vdots \\ \varphi_r(x_1) \end{pmatrix}, \ldots, \begin{pmatrix} \varphi_1(x_r) \\ \varphi_2(x_r) \\ \vdots \\ \varphi_r(x_r) \end{pmatrix}$$

are affinely dependent.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

The proof of the proposition is a straightforward algebraic manipulation, mostly based on the fact that if a polynomial over the reals vanishes everywhere, then its coefficients are zero. The reader could skip the proof if desired.

Proof. Start from the definition, (i) is equivalent to the vectors

$$\begin{pmatrix}(\langle x_0,u_1\rangle+c_1)^l\\(\langle x_0,u_2\rangle+c_2)^l\\\vdots\\(\langle x_0,u_d\rangle+c_d)^l\end{pmatrix},\quad\begin{pmatrix}(\langle x_1,u_1\rangle+c_1)^l\\(\langle x_1,u_2\rangle+c_2)^l\\\vdots\\(\langle x_1,u_d\rangle+c_d)^l\end{pmatrix},\ldots,\begin{pmatrix}(\langle x_r,u_1\rangle+c_1)^l\\(\langle x_r,u_2\rangle+c_2)^l\\\vdots\\(\langle x_r,u_d\rangle+c_d)^l\end{pmatrix}$$

being affinely dependent for any choice of parameters $c_1,c_2,\ldots,c_d\in\mathbb{R}$, $u_1,u_2,\ldots,u_d\in\mathbb{R}^N$ and $1\leqslant l\leqslant m$. In particular, as $r\leqslant d$, this is further equivalent to vectors

$$\begin{pmatrix}(\langle x_1,u_1\rangle+c_1)^l-(\langle x_0,u_1\rangle+c_1)^l\\(\langle x_1,u_2\rangle+c_2)^l-(\langle x_0,u_2\rangle+c_2)^l\\\vdots\\(\langle x_1,u_r\rangle+c_r)^l-(\langle x_0,u_r\rangle+c_r)^l\end{pmatrix},\quad\begin{pmatrix}(\langle x_2,u_1\rangle+c_1)^l-(\langle x_0,u_1\rangle+c_1)^l\\(\langle x_2,u_2\rangle+c_2)^l-(\langle x_0,u_2\rangle+c_2)^l\\\vdots\\(\langle x_2,u_r\rangle+c_r)^l-(\langle x_0,u_r\rangle+c_r)^l\end{pmatrix},\ldots,$$
$$\begin{pmatrix}(\langle x_r,u_1\rangle+c_1)^l-(\langle x_0,u_1\rangle+c_1)^l\\(\langle x_r,u_2\rangle+c_2)^l-(\langle x_0,u_2\rangle+c_2)^l\\\vdots\\(\langle x_r,u_r\rangle+c_r)^l-(\langle x_0,u_r\rangle+c_r)^l\end{pmatrix}$$

being linearly dependent for all the choices of parameters. Hence, taking the determinant, (i) is the same as

$$\det((\langle x_i,u_j\rangle+c_j)^l-(\langle x_0,u_j\rangle+c_j)^l)=0$$

for any choice of $u_1,\ldots,u_r$, $c_1,\ldots,c_r,l$. Expanding using the binomial theorem, and writing S_r for the symmetric group on $[r]$, we obtain

$$\begin{aligned}0&=\sum_{\pi\in S_r}\operatorname{sgn}(\pi)\prod_{i=1}^{r}((\langle x_{\pi(i)},u_i\rangle+c_i)^l-(\langle x_0,u_i\rangle+c_i)^l)\\&=\sum_{\pi\in S_r}\operatorname{sgn}(\pi)\prod_{i=1}^{r}\left(\sum_{k=0}^{l}c_i^k\binom{l}{k}(\langle x_{\pi(i)},u_i\rangle^{l-k}-\langle x_0,u_i\rangle^{l-k})\right)\\&=\sum_{0\leqslant k_1,k_2,\ldots,k_r\leqslant l}c_1^{k_1}c_2^{k_2}\cdots c_r^{k_r}\prod_{i=1}^{r}\binom{l}{k_i}\left(\sum_{\pi\in S_r}\operatorname{sgn}(\pi)\prod_{i=1}^{r}(\langle x_{\pi(i)},u_i\rangle^{l-k_i}-\langle x_0,u_i\rangle^{l-k_i})\right).\end{aligned}$$

However, this holds for any choice of $c_1,c_2,\ldots,c_r\in\mathbb{R}$, so when the expression above is viewed as a polynomial in variables $c_1,c_2,\ldots,c_r$, we conclude that all its coefficients are zero. In other words, (i) is equivalent to the following. For any $0\leqslant k_1,k_2,\ldots,k_r\leqslant m$, and

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

any $u_1, u_2, \dots, u_r \in \mathbb{R}^N$, we have

$$\begin{aligned}
0 &= \sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} (\langle x_{\pi(i)}, u_i\rangle^{k_i} - \langle x_0, u_i\rangle^{k_i}) \\
&= \sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} \left(\left(\sum_{j=1}^{N} x_{\pi(i)j} u_{ij} \right)^{k_i} - \left(\sum_{j=1}^{N} x_{0j} u_{ij} \right)^{k_i} \right) \\
&= \sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} \left(\sum_{\varphi_i : [k_i]\to[N]} \left(\prod_{j=1}^{k_i} x_{\pi(i)\varphi_i(j)} u_{i\varphi_i(j)} - \prod_{j=1}^{k_i} x_{0\varphi_i(j)} u_{i\varphi_i(j)} \right) \right) \\
&= \sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} \left(\sum_{\varphi_i : [k_i]\to[N]} \left(\prod_{j=1}^{k_i} u_{i\varphi_i(j)} \right) \left(\prod_{j=1}^{k_i} x_{\pi(i)\varphi_i(j)} - \prod_{j=1}^{k_i} x_{0\varphi_i(j)} \right) \right) \\
&= \sum_{\varphi_1 : [k_1]\to[N], \dots, \varphi_r : [k_r]\to[N]} \left(\prod_{i=1}^{r} \prod_{j=1}^{k_i} u_{i\varphi_i(j)} \right) \left(\sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} \left(\prod_{j=1}^{k_i} x_{\pi(i)\varphi_i(j)} - \prod_{j=1}^{k_i} x_{0\varphi_i(j)} \right) \right) \\
&= \sum_{\varphi_1 : [k_1]\to[N], \dots, \varphi_r : [k_r]\to[N]} \left(\prod_{i=1}^{r} \varphi_i(u_i) \right) \left(\sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} (\varphi_i(x_{\pi(i)}) - \varphi_i(x_0)) \right).
\end{aligned}$$

Now, look at the expression above as a polynomial in variables u_{ij}. Observe that if $\varphi_1, \varphi_2, \dots, \varphi_r$, $\psi_1, \psi_2, \dots, \psi_r$ are such that $\prod_{i=1}^{r} \varphi_i(u_i) = \prod_{i=1}^{r} \psi_i(u_i)$ as formal expressions, then we must have

$$\sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} (\varphi_i(x_{\pi(i)}) - \varphi_i(x_0)) = \sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} (\psi_i(x_{\pi(i)}) - \psi_i(x_0))$$

as well. This tells us that the coefficients of our polynomial are positive integer multiples of

$$\sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} (\varphi_i(x_{\pi(i)}) - \varphi_i(x_0)).$$

Once again, the polynomial over $\mathbb{R}$ vanishes everywhere if and only if its coefficients are zero, so we deduce that (i) holds if and only if, for all ($\leqslant m$)-functions $\varphi_1, \varphi_2, \dots, \varphi_r$ to $[N]$, we have

$$0 = \sum_{\pi\in S_r} \operatorname{sgn}(\pi) \prod_{i=1}^{r} (\varphi_i(x_{\pi(i)}) - \varphi_i(x_0)) = \det_{1\leqslant i,j\leqslant r} (\varphi_i(x_j) - \varphi_i(x_0)),$$

which says precisely that the vectors

$$\begin{pmatrix} \varphi_1(x_1) - \varphi_1(x_0) \\ \varphi_2(x_1) - \varphi_2(x_0) \\ \vdots \\ \varphi_r(x_1) - \varphi_r(x_0) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(x_2) - \varphi_1(x_0) \\ \varphi_2(x_2) - \varphi_2(x_0) \\ \vdots \\ \varphi_r(x_2) - \varphi_r(x_0) \end{pmatrix}, \dots, \begin{pmatrix} \varphi_1(x_r) - \varphi_1(x_0) \\ \varphi_2(x_r) - \varphi_2(x_0) \\ \vdots \\ \varphi_r(x_r) - \varphi_r(x_0) \end{pmatrix}$$

are linearly dependent, which is equivalent to (ii), as desired. □

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Proposition 4.4. *Let $r, d, m, N \in \mathbb{N}$ be given, and suppose that $r \leqslant d$ holds. Suppose also that the set $\{x_0, x_1, \ldots, x_r\} \subseteq \mathbb{R}^N$ is $\mathcal{F}_{N,d,m}$-incident. Then, given any affine map $\alpha : \mathbb{R}^N \to \mathbb{R}^N$ and any ($\leqslant m$)-functions $\varphi_1, \varphi_2, \ldots, \varphi_r$ to $[N]$, the vectors*

$$\begin{pmatrix} \varphi_1(\alpha(x_0)) \\ \varphi_2(\alpha(x_0)) \\ \vdots \\ \varphi_r(\alpha(x_0)) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(\alpha(x_1)) \\ \varphi_2(\alpha(x_1)) \\ \vdots \\ \varphi_r(\alpha(x_1)) \end{pmatrix}, \ldots, \begin{pmatrix} \varphi_1(\alpha(x_r)) \\ \varphi_2(\alpha(x_r)) \\ \vdots \\ \varphi_r(\alpha(x_r)) \end{pmatrix}$$

are affinely dependent.

On the other hand, if the set $\{x_0, x_1, \ldots, x_r\} \subseteq \mathbb{R}^N$ is not $\mathcal{F}_{N,d,m}$-incident, then, given any affine isomorphism $\alpha : \mathbb{R}^N \to \mathbb{R}^N$, we may find ($\leqslant m$)-functions $\varphi_1, \varphi_2, \ldots, \varphi_r$ to $[N]$, so that the vectors

$$\begin{pmatrix} \varphi_1(\alpha(x_0)) \\ \varphi_2(\alpha(x_0)) \\ \vdots \\ \varphi_r(\alpha(x_0)) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(\alpha(x_1)) \\ \varphi_2(\alpha(x_1)) \\ \vdots \\ \varphi_r(\alpha(x_1)) \end{pmatrix}, \ldots, \begin{pmatrix} \varphi_1(\alpha(x_r)) \\ \varphi_2(\alpha(x_r)) \\ \vdots \\ \varphi_r(\alpha(x_r)) \end{pmatrix}$$

are affinely independent.

Proof. Suppose that an affine map $\alpha : \mathbb{R}^N \to \mathbb{R}^N$ is given. We may write it in the form $\alpha(x) = Ax + v$ for $N \times N$ matrix A and a vector $v \in \mathbb{R}^N$. Given vectors $u_1, u_2, \ldots, u_r \in \mathbb{R}^N$, constants $c_1, c_2, \ldots, c_r \in \mathbb{R}$ and $1 \leqslant l \leqslant m$, we have

$$(\langle \alpha(x), u_i \rangle + c_i)^l = (\langle Ax + v, u_i \rangle + c_i)^l = (\langle x, A^T u \rangle + (\langle v, u_i \rangle + c_i))^l.$$

But then, since $x_0, x_1, \ldots, x_r$ is $\mathcal{F}_{N,d,m}$-incident, it follows that so is $\alpha(x_0), \alpha(x_1), \ldots, \alpha(x_r)$. Apply Proposition 4.3 to $\alpha(x_0), \alpha(x_1), \ldots, \alpha(x_r)$, from which the first claim in the proposition follows.

For the second part, observe that if $\alpha(x_0), \alpha(x_1), \ldots, \alpha(x_r)$ is $\mathcal{F}_{N,d,m}$-incident, then by the previous arguments, so is $x_0 = \alpha^{-1}(\alpha(x_0)), x_1 = \alpha^{-1}(\alpha(x_1)), \ldots, x_r = \alpha^{-1}(\alpha(x_r))$. Therefore, the points $\alpha(x_0), \alpha(x_1), \ldots, \alpha(x_r)$ are not $\mathcal{F}_{N,d,m}$-incident. Proposition 4.3 applies, and gives the desired ($\leqslant m$)-functions. □

5. Proof of $\alpha(n, d) = o_d(n)$

Lemma 5.1. *Let $m, r, N \in \mathbb{N}$. Suppose that $y_1, y_2, \ldots, y_r \in \mathbb{R}^N$ are some vectors with the property that $\operatorname{rank}\{y_1, y_2, \ldots, y_r\} + m - 1 \geqslant r$. Suppose further that $y_1, y_2, \ldots, y_r$ are distinct and have non-zero coordinates. Then we may find ($\leqslant m$)-functions $\varphi_1, \varphi_2, \ldots, \varphi_r$ for which the vectors*

$$\begin{pmatrix} \varphi_1(y_1) \\ \varphi_2(y_1) \\ \vdots \\ \varphi_r(y_1) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(y_2) \\ \varphi_2(y_2) \\ \vdots \\ \varphi_r(y_2) \end{pmatrix}, \ldots, \begin{pmatrix} \varphi_1(y_r) \\ \varphi_2(y_r) \\ \vdots \\ \varphi_r(y_r) \end{pmatrix}$$

are linearly independent.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Proof. We prove the lemma by induction, first on m, then on r. Observe that in the case when $m = 1$, for a $(\leqslant m)$-function φ, the resulting function $\varphi(x)$ is just evaluation of x at a chosen coordinate. Hence, for $m = 1$ we are actually asked to find a set of coordinates I of size r, such that y_i are still linearly independent when restricted to I. Applying Lemma 3.2 proves the claim in this case.

Suppose now that the claim holds for some $m' \geqslant 1$. Write $m = m' + 1$. For fixed m, we prove the lemma by induction on $r \geqslant 1$. If $r = 1$, then, take $\varphi : [1] \to [N]$, given by $\varphi(1) = 1$, so the vector $\varphi(y_1) = (y_1)_1$ is non-zero.

Suppose that the claim holds for some $r \geqslant 1$, and that $\{y_1, y_2, \ldots, y_{r+1}\}$ satisfy the assumptions of the lemma.

Case 1. $y_{r+1} \notin \operatorname{span}\{y_1, y_2, \ldots, y_r\}$. Then

$$r + 1 \leqslant \operatorname{rank}\{y_1, y_2, \ldots, y_{r+1}\} + m - 1 = \operatorname{rank}\{y_1, y_2, \ldots, y_r\} + m,$$

hence $\operatorname{rank}\{y_1, y_2, \ldots, y_r\} + m - 1 \geqslant r$. By the induction hypothesis, we have $(\leqslant m)$-functions $\varphi_1, \varphi_2, \ldots, \varphi_r$ such that

$$\begin{pmatrix} \varphi_1(y_1) \\ \varphi_2(y_1) \\ \vdots \\ \varphi_r(y_1) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(y_2) \\ \varphi_2(y_2) \\ \vdots \\ \varphi_r(y_2) \end{pmatrix}, \ldots, \begin{pmatrix} \varphi_1(y_r) \\ \varphi_2(y_r) \\ \vdots \\ \varphi_r(y_r) \end{pmatrix}$$

are linearly independent. Hence, there are unique scalars $\lambda_1, \lambda_2, \ldots, \lambda_r \in \mathbb{R}$ such that

$$\varphi_i(y_{r+1}) = \sum_{j=1}^{r} \lambda_j \varphi_i(y_j), \quad \text{for all } i = 1, \ldots, r.$$

However, the vector y_{r+1} is not in the $\operatorname{span}\{y_1, \ldots, y_r\}$, and so $y_{r+1} \neq \sum_{j=1}^{r} \lambda_j y_j$. Hence, we can pick $\varphi_{r+1} : [1] \to [N]$ to be $\varphi(1) = c$, where c is the coordinate such that $(y_{r+1})_c \neq \sum_{j=1}^{r} \lambda_j (y_j)_c$, finishing the proof in this case.

Case 2. $y_{r+1} \in \operatorname{span}\{y_1, y_2, \ldots, y_r\}$. Then

$$r + 1 \leqslant \operatorname{rank}\{y_1, y_2, \ldots, y_{r+1}\} + m - 1 = \operatorname{rank}\{y_1, y_2, \ldots, y_r\} + m - 1,$$

so

$$r \leqslant \operatorname{rank}\{y_1, y_2, \ldots, y_r\} + m - 2.$$

By the induction hypothesis, we have $(\leqslant m - 1)$-functions $\varphi_1, \ldots, \varphi_r$ for which

$$\begin{pmatrix} \varphi_1(y_1) \\ \varphi_2(y_1) \\ \vdots \\ \varphi_r(y_1) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(y_2) \\ \varphi_2(y_2) \\ \vdots \\ \varphi_r(y_2) \end{pmatrix}, \ldots, \begin{pmatrix} \varphi_1(y_r) \\ \varphi_2(y_r) \\ \vdots \\ \varphi_r(y_r) \end{pmatrix}$$

are linearly independent. As before, there are unique scalars $\lambda_1, \lambda_2, \ldots, \lambda_r \in \mathbb{R}$ such that

$$\varphi_i(y_{r+1}) = \sum_{j=1}^{r} \lambda_j \varphi_i(y_j), \quad \text{for all } i = 1, \ldots, r.$$

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

We try to take φ_{r+1} to be some φ_i with additional elements in the domain, mapped to $c \in [N]$. If this works, we are done. Otherwise, for all $i = 1, \dots, r$ and $c \in [N]$, we have $\varphi_i(y_{r+1})(y_{r+1})_c = \sum_{j=1}^{r} \lambda_j \varphi_i(y_j)(y_j)_c$. Since the coordinates are non-zero, we get

$$\varphi_i(y_{r+1}) = \sum_{j=1}^{r} (\lambda_j \cdot (y_j)_c/(y_{r+1})_c)\varphi_i(y_j).$$

However, by uniqueness of λ_j, we must have $\lambda_j \cdot (y_j)_c/(y_{r+1})_c = \lambda_j$ for all j, c. If some $\lambda_j \neq 0$, then for all c we get $(y_j)_c/(y_{r+1})_c = 1$, that is, $y_{r+1} = y_j$, which is a contradiction, as our vectors are distinct. Otherwise, all the coefficients $\lambda_j = 0$, so $\varphi_1(y_{r+1}) = 0$, but coordinates of y_{r+1} are non-zero, resulting in contradiction once again. □

The next result is a corollary of the algebraic lemma we have just proved. It is consistent with the intuition we described in the Introduction. There we said that we expected lines in $[m+1]^N$ to be the sources of non-trivial $\mathcal{F}_{N,m+1,m}$-incident sets. In other words, a $\mathcal{F}_{N,m+1,m}$-incident set is either of size larger than $m+2$, and thus its image must be affinely dependent (by looking at dimension of the target space), or the set is on a line.

Corollary 5.2. *Suppose that $S \subseteq \mathbb{R}^N$ is $\mathcal{F}_{N,m+1,m}$-incident. Then $|S| \geqslant m+2$, and if $|S| = m+2$ then S is a subset of a line.*

Proof. If $|S| \geqslant m+3$, we are done. Suppose now that $|S| \leqslant m+2$. Let $s_0, s_1, \dots, s_{m+1}$ be the elements of S. We can find an affine isomorphism $\alpha : \mathbb{R}^N \to \mathbb{R}^N$ such that $\alpha(s_0) = 0$, and $y_i = \alpha(s_i)$, for $i = 1, 2, \dots, m+1$, are distinct and have non-zero coordinates. By Proposition 4.4 (note that we may apply it because $|S| - 1 \leqslant m+1$, and $m+1$ is the dimension of the target space), the vectors

$$\begin{pmatrix} \varphi_1(y_1) \\ \varphi_2(y_1) \\ \vdots \\ \varphi_{m+1}(y_1) \end{pmatrix}, \begin{pmatrix} \varphi_1(y_2) \\ \varphi_2(y_2) \\ \vdots \\ \varphi_{m+1}(y_2) \end{pmatrix}, \dots, \begin{pmatrix} \varphi_1(y_{m+1}) \\ \varphi_2(y_{m+1}) \\ \vdots \\ \varphi_{m+1}(y_{m+1}) \end{pmatrix}$$

are linearly dependent, for any choice of ($\leqslant m$)-functions $\varphi_1, \varphi_2, \dots, \varphi_{m+1}$ to $[N]$. Therefore Lemma 5.1 would give us a contradiction, unless

$$\operatorname{rank}\{y_1, y_2, \dots, y_{m+1}\} + m - 1 \leqslant m.$$

So $\operatorname{rank}\{y_1, y_2, \dots, y_{m+1}\} \leqslant 1$, and as $y_1 \neq 0$, there are scalars $\lambda_1, \dots, \lambda_{m+1}$ such that $y_i = \lambda_i y_1$ holds for all $i = 1, \dots, m+1$. But, since α is an affine isomorphism, the points $s_0 = \alpha^{-1}(0), s_1 = \alpha^{-1}(y_1), \dots, s_{m+1} = \alpha^{-1}(y_{m+1})$ are on a line, as desired. □

Theorem 5.3. *For $d, n \in \mathbb{N}$, $d \geqslant 2$, we have $\alpha(n, d) = o_d(n)$.*

Proof. For the notational consistency with previous results, we set $m = d - 1$. Let $\epsilon > 0$ be arbitrary and let N be sufficiently large that the ϵ-density Hales–Jewett theorem holds for the combinatorial lines in $[m+2]^N$. Let $X = [m+2]^N$, and let f be a function given

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

by Proposition 3.1 applied to X and $\mathcal{F}_{N,m+1,m}$. Since $\mathcal{F}_{N,m+1,m}$ separates the points of X, we may assume that f is injective on X. Finally, let $Y = f(X) \subseteq \mathbb{R}^{m+1}$. We claim that Y has no more than $m+2$ points in a hyperplane, and that all the subsets of Y of size at least $\epsilon|Y|$ have a hyperplane containing $m+2$ points.

There are no more than $m+2$ points of Y on a hyperplane. Consider a hyperplane H and suppose that Y has $m+3$ points $y_1, \ldots, y_{m+3}$ inside H. Take a maximal affinely independent subset of $y_1, \ldots, y_{m+3}$: without loss of generality, this is $y_1, y_2, \ldots, y_r$ for some r. Since H is an m-dimensional affine subspace, we have $r \leqslant m+1$. So $S_1 = \{y_1, y_2, \ldots, y_r, y_{m+2}\}$ is affinely dependent, and has size at most $m+2$. Then, by the definition of f and Proposition 3.1, $T_1 = f^{-1}(S_1)$ is $\mathcal{F}_{N,m+1,m}$-incident. Since f is a bijection from X onto its image, T_1 has size at most $m+2$, so by Corollary 5.2, T_1 is a subset of a line, and $|T_1| = m+2$ and $r = m+1$. Applying the same arguments to $S_2 = \{y_1, \ldots, y_r, y_{m+3}\}$ and $T_2 = f^{-1}(S_2)$, we have that T_2 is also a subset of a line and has size $m+2$ and also $|T_1 \cap T_2| = m+1$. But, as $T_1, T_2 \subseteq [m+2]^N$, this is impossible and we have a contradiction, so Y has no more than $m+2$ points on a hyperplane.

Dense subsets of Y are not in general position. Let $S \subseteq Y$ have size at least $\epsilon|Y|$. Then $T = f^{-1}(S)$ has a combinatorial line L by the density Hales–Jewett theorem. Hence, $f(L) \subseteq S$ and $f(S)$ has $m+2$ points that lie on the same hyperplane, by Lemma 4.1. This finishes the proof. □

6. Better bounds for certain dimensions

In this section we provide better bounds on $\alpha(n,d)$ for certain dimensions d. The key difference in this approach is the use of a more efficient version of the density Hales–Jewett theorem, which we now state.

Theorem 6.1 (generalized Sperner's theorem: [5], Theorem 2.3). *Let $\mathcal{A}$ be a collection of subsets of $[n]$ that contains no d-dimensional combinatorial subspace. Then the size of $\mathcal{A}$ is at most $(25/n)^{1/2^d} 2^n$.*

Here, we consider the points in $\{0,1\}^N \subseteq \mathbb{R}^N$, which we also interpret as subsets of $[N]$. Observe that, given an ($\leqslant m$)-function φ to $[N]$, with image $S \subseteq [N]$ and a point $x \in \{0,1\}^N$ corresponding to $X \subseteq [N]$ (by setting $X := \{i \in [N] : x_i = 1\}$), we have

$$f(x) = \mathbb{1}_{S \subseteq X},$$

where, for a general property P, the value of $\mathbb{1}_P$ is 1, when P holds, and zero otherwise. Hence, we can reinterpret Proposition 4.4 in the language of sets as follows. Suppose that $\emptyset, X_1, X_2, \ldots, X_r$ correspond to $r+1$ points in $\{0,1\}^N$ which are not $\mathcal{F}_{N,d,m}$-incident (so the first point is 0). Then there are sets $S_1, S_2, \ldots, S_r \subseteq [N]$ of size at most m, for which the

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

vectors

$$\begin{pmatrix} \mathbb{1}_{S_1\subseteq\emptyset} \\ \mathbb{1}_{S_2\subseteq\emptyset} \\ \vdots \\ \mathbb{1}_{S_r\subseteq\emptyset} \end{pmatrix}, \quad \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_1} \\ \mathbb{1}_{S_2\subseteq X_1} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_1} \end{pmatrix}, \quad \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_2} \\ \mathbb{1}_{S_2\subseteq X_2} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_2} \end{pmatrix}, \ldots, \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_r} \\ \mathbb{1}_{S_2\subseteq X_r} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_r} \end{pmatrix}$$

are affinely independent. If all the sets S_i are non-empty, then the vectors

$$\begin{pmatrix} \mathbb{1}_{S_1\subseteq X_1} \\ \mathbb{1}_{S_2\subseteq X_1} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_1} \end{pmatrix}, \quad \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_2} \\ \mathbb{1}_{S_2\subseteq X_2} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_2} \end{pmatrix}, \ldots, \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_r} \\ \mathbb{1}_{S_2\subseteq X_r} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_r} \end{pmatrix}$$

are linearly independent. Now we show that actually no set S_i is empty. Otherwise, without loss of generality $S_1 = S_2 = \cdots = S_k = \emptyset$ and the others are non-empty, so after subtracting the vector $(\mathbb{1}_{S_1\subseteq\emptyset}\mathbb{1}_{S_2\subseteq\emptyset}\cdots\mathbb{1}_{S_r\subseteq\emptyset})^T$ from each vector $(\mathbb{1}_{S_1\subseteq X_i}\mathbb{1}_{S_2\subseteq X_i}\cdots\mathbb{1}_{S_r\subseteq X_i})^T$ with $i \geqslant 1$, we obtain that

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mathbb{1}_{S_{k+1}\subseteq X_1} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_1} \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mathbb{1}_{S_{k+1}\subseteq X_2} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_2} \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mathbb{1}_{S_{k+1}\subseteq X_r} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_r} \end{pmatrix}$$

are linearly independent, which is not possible (when viewed as a matrix, the row rank is less than r). We sum up this discussion in the following observation.

Observation 6.2. *Suppose that the sets $\emptyset, X_1, X_2, \ldots, X_r \subseteq [N]$ correspond to $r+1$ points that are not $\mathcal{F}_{N,d,m}$-incident. Then there are non-empty sets $S_1, S_2, \ldots, S_r \subseteq \mathbb{N}$ of size at most m such that the vectors*

$$\begin{pmatrix} \mathbb{1}_{S_1\subseteq X_1} \\ \mathbb{1}_{S_2\subseteq X_1} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_1} \end{pmatrix}, \quad \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_2} \\ \mathbb{1}_{S_2\subseteq X_2} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_2} \end{pmatrix}, \ldots, \begin{pmatrix} \mathbb{1}_{S_1\subseteq X_r} \\ \mathbb{1}_{S_2\subseteq X_r} \\ \vdots \\ \mathbb{1}_{S_r\subseteq X_r} \end{pmatrix}$$

are linearly independent.

Viewing these vectors together as an $r \times r$ matrix, we have found that the nullity of this matrix is related to the notion of $\mathcal{F}_{N,d,m}$-incidence. This motivates the study of the nullity of such matrices in general. Before stating the lemma, which contains some basic results regarding this problem, we introduce some notation.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Given sets $A_1, A_2, \ldots, A_r, B_1, B_2, \ldots, B_s \in \mathbb{N}^{(<\omega)}$, we write

$$I(A_1, A_2, \ldots, A_r; B_1, B_2, \ldots, B_s)$$

for the $s \times r$ matrix with entries $I_{ij} = \mathbb{1}_{B_i \subseteq A_j}$. Further, we define

$$K(A_1, A_2, \ldots, A_r; B_1, B_2, \ldots, B_s)$$

as the kernel of I and

$$n(A_1, A_2, \ldots, A_r; B_1, B_2, \ldots, B_s)$$

as the nullity of I. Also, if A, B are finite sequences of finite sets, of lengths r and s, we write $I(A, B) = I(A_1, A_2, \ldots, A_r; B_1, B_2, \ldots, B_s)$, and similarly we define $K(A, B), n(A, B)$.

Lemma 6.3. *Let $m, k \in \mathbb{N}$. Given any distinct sets $X_1, X_2, \ldots, X_r \in \mathbb{N}^{(<\omega)}$, we can find sets $S_1, S_2, \ldots, S_r \subseteq \mathbb{N}^{(\leqslant m)}$ which enjoy the following properties:*

(i) $n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r) = 0$, *provided* $r < 2^{m+1}$,
(ii) $n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r) \leqslant 1$, *provided* $r < 3 \cdot 2^m$.

We prove the lemma by induction and (ideas related to) compressions, and in fact use part (i) in order to deduce part (ii). As will be stressed in the proof, there is a subtlety in proving $n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r) \leqslant 1$: that is, naive application of induction could only give us $n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r) \leqslant 2$, and we actually use the first part of the lemma to obtain the required saving of 1 on the right-hand side.

Proof. *Part (i).* We prove the claim by induction on $\sum_{i=1}^{r} |X_i|$. If this is zero, then we have $r = 1$ and $X_1 = \emptyset$, so just take $S_1 = \emptyset$.

Suppose that the lemma holds for smaller values of $\sum_{i=1}^{r} |X_i|$. Let $x \in \mathbb{N}$ be any element that is contained in at least one of the sets X_i. Let $\{Y_1, Y_2, \ldots, Y_u\}$ denote the collection of sets given by $\{X_i \setminus \{x\} : i = 1, \ldots, r\}$, and further let $\{Z_1, \ldots, Z_v\}$ be the set

$$\{X_i : x \notin X_i, X_i \cup \{x\} = X_j \text{ for some } j\}.$$

Thus $v \leqslant u$ and $u + v = r$. By the induction hypothesis, there are relevant sets $S_1, \ldots, S_u \in \mathbb{N}^{(\leqslant m)}$ for $Y_1, \ldots, Y_u$. Also, since $v \leqslant r/2 < 2^m$, we have relevant sets $S'_{u+1}, \ldots, S'_r \in \mathbb{N}^{(\leqslant m-1)}$ for $Z_1, Z_2, \ldots, Z_v$, and note that without loss of generality none of $S_1, S_2, \ldots, S_u, S'_{u+1}, \ldots, S'_r$ contains x. Set $S_{u+i} = S'_{u+i} \cup \{x\}$ for all $i = 1, \ldots, v$. We claim that the sets $S_1, S_2, \ldots, S_r$ have the desired property. So far, we know that for all i, $|S_i| \leqslant m$ holds.

Let $(\lambda_1 \lambda_2 \cdots \lambda_r)^T$ be an element of the kernel $K(X, S)$. We can rewrite this as

$$\sum_{j\,:\, S_i \subseteq X_j} \lambda_j = 0, \quad \text{for all } i = 1, 2, \ldots, r.$$

Define $\mu_i = \sum_{j\,:\, Y_i = X_j \setminus \{x\}} \lambda_j$, for each $i = 1, \ldots, u$. Then, recalling that $x \notin S_i$ for $i \leqslant u$, we have

$$\sum_{j\,:\, S_i \subseteq Y_j} \mu_j = \sum_{j\,:\, S_i \subseteq X_l} \lambda_l = 0, \quad \text{for all } i = 1, 2, \ldots, u.$$

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Since $n(Y_1, Y_2, \dots, Y_u; S_1, S_2, \dots, S_u) = 0$, we infer $\mu_j = 0$ for all j. Returning to the definition of μ_i, we have that whenever j is such that $x \in X_j$, but $X_j \setminus \{x\}$ is not any of the sets X_l, then there is i such that $\lambda_j = \mu_i$, so λ_j must vanish.

Hence, for $i > u$, we have that $x \in S_i$ and $\sum_j \lambda_j \mathbb{1}_{S_i \subseteq X_j} = 0$, when the sum is taken over all j such that $x \in X_j$, but $X_j \setminus \{x\}$ is not any of the sets X_l. Set $v_i = -\lambda_j$ for $Z_i = X_j$, which by previous work is the same as $v_i = \lambda_l$ for l such that $X_l \setminus \{x\} = Z_i$ and $x \in X_l$. Therefore, for $i > u$, we have

$$\sum_{j\,:\,S_i' \subseteq Z_j} v_j = \sum_{j\,:\,S_i \subseteq Z_j \cup \{x\}} v_j = \sum_{j\,:\,S_i \subseteq X_j} \lambda_j = 0$$

with no additional restrictions on j apart from those written in the sums. Hence the vector $(v_1 v_2 \cdots v_v)^T$ must lie inside $K(Z_1, Z_2, \dots, Z_v; S'_{u+1}, S'_{u+2}, \dots, S'_r)$. But this kernel is trivial, so all $v_i = 0$. Finally, this implies that $\lambda_i = 0$ holds for the remaining i, as desired.

Part (ii). We follow similar steps to the previous part. However, we have to be slightly careful, since the previous argument unchanged would give us

$$K(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r)$$

essentially as a sum of kernels of similar matrices for $Y_1, Y_2, \dots, Y_u$ and $Z_{u+1}, Z_{u+2}, \dots, Z_r$. That way, we could be one dimension short of the desired goal, as this argument only allows us to deduce $n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) \leqslant 2$, so we have to be more efficient. In order to overcome this issue, we shall apply part (i) of the lemma.

We prove the claim by induction on $\sum_{i=1}^r |X_i|$. If this is zero, then we have $r = 1$ and $X_1 = \emptyset$, so just take $S_1 = \emptyset$.

Suppose that the lemma holds for smaller values of $\sum_{i=1}^r |X_i|$. Let $x \in \mathbb{N}$ be any element that is contained in at least one of the sets X_i. Let $\{Y_1, Y_2, \dots, Y_u\}$ denote the collection of sets given by $\{X_i \setminus \{x\} : i = 1, \dots, r\}$, and further let $\{Z_1, \dots, Z_v\}$ be the set

$$\{X_i : x \notin X_i, X_i \cup \{x\} = X_j \text{ for some } j\}.$$

Thus $v \leqslant u$ and $u + v = r$. Pick the sets $S_1, S_2, \dots, S_u \in \mathbb{N}^{(\leqslant m)}$ such that

$$U = K(Y_1, Y_2, \dots, Y_u; S_1, S_2, \dots, S_u)$$

is of minimum dimension. Further, pick the sets $S'_{u+1}, S'_{u+2}, \dots, S'_r \in \mathbb{N}^{(\leqslant m-1)}$ such that

$$V = K(Z_1, Z_2, \dots, Z_v; S'_{u+1}, S'_{u+2}, \dots, S'_r)$$

is of minimum dimension. Without loss of generality we may assume that $x \notin S_i, S'_j$. Finally, set $S_{u+i} = S'_{u+i} \cup \{x\}$ for $i = 1, \dots, v$. All S_i have size at most m.

By the induction hypothesis, we have $\dim U \leqslant 1$ and, since $v \leqslant r/2 < 3 \cdot 2^{m-1}$, we also have $\dim V \leqslant 1$. However, we can make a saving of one dimension as promised. Suppose that $\dim U = \dim V = 1$. Then, by part (i), since U, V are of the minimum possible dimension, we must have $u \geqslant 2^{m+1}$ and $v \geqslant 2^m$, which is a contradiction as $u + v = r < 3 \cdot 2^m$. Therefore, $\dim U + \dim V \leqslant 1$. We shall now finish the proof by similar arguments as in the previous case, but we need to treat the cases when $\dim U = 1$ and $\dim V = 1$ separately.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

We may reorder $X_1, X_2, \ldots, X_r$, if necessary, to have $Y_i = X_i \setminus \{x\}$, for $i = 1, 2, \ldots, u$, and $Z_i = X_{u+i} \setminus \{x\}$ with $x \in X_{u+i}$ for $i = 1, 2, \ldots, v$. Furthermore, we may also assume that $Z_i = X_{u-v+i}$ with $x \notin X_{u-v+i}$ for $i = 1, 2, \ldots, v$. Now proceed as in part (i), with the argument modified to deal with the possibly non-trivial kernels. Suppose that $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ are such that $\sum_{j\colon S_i \subseteq X_j} \lambda_j = 0$ for all $i = 1, 2, \ldots, r$. Define $\mu_i = \sum_{j\colon Y_i = X_j \setminus \{x\}} \lambda_j$, for each $i = 1, \ldots, u$, thus

$$\mu_i = \begin{cases} \lambda_i & \text{if } i \leqslant u - v, \\ \lambda_i + \lambda_{i+v} & \text{if } u - v < i \leqslant u. \end{cases}$$

Then we have $\sum_{j\colon S_i \subseteq Y_j} \mu_j = 0$ for all $i = 1, 2, \ldots, u$. This gives $\mu \in U$. Next, set $v_i = \lambda_j$ for $Z_i = X_j$, that is, $v_i = \lambda_{u+i}$ for $i = 1, 2, \ldots, v$. We have

$$0 = \sum_{j\colon S_i \subseteq X_j} \lambda_j = \sum_{j\colon S_i' \cup \{x\} \subseteq Z_j \cup \{x\}} v_j + \sum_{j \in [u-v]\colon S_i \subseteq X_j} \lambda_j = \sum_{j\colon S_i' \subseteq Z_j} v_j + \sum_{j \in [u-v]\colon S_i \subseteq X_j} \lambda_j \tag{6.1}$$

for all $i = u + 1, \ldots, r$ (note the restriction on the index $j \in [u - v]$ in some of the sums). Note also that we can express λ_i in terms of μ_i and v_i as follows:

$$\lambda_i = \begin{cases} \mu_i & \text{if } i \leqslant u - v, \\ \mu_i - v_{i+v-u} & \text{if } u - v < i \leqslant u, \\ v_{i-u} & \text{if } u < i. \end{cases} \tag{6.2}$$

Case 1. $\dim U = 0, \dim V = 1$. As $U = \{0\}$, we must have all $\mu_j = 0$. Therefore, $\lambda_j = 0$ for $j \leqslant u - v$, which implies that

$$\sum_{j\colon S_i' \subseteq Z_j} v_j = 0$$

for all $i = 1, 2, \ldots, u$. Therefore, $v \in V$. Using the expressions (6.2), we see that we can express any given $\lambda \in K(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r)$ as a sum of vectors in two subspaces of $\mathbb{R}^r$, isomorphic to U and V, so $K(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r)$ is a subset of at most one-dimensional subspace, as desired.

Case 2. $\dim V = 0, \dim U = 1$. Now, the scalars μ_j might be non-trivial. On the other hand, due to equation (6.1) we have

$$\sum_{j\colon S_i' \subseteq Z_j} v_j = - \sum_{j \in [u-v]\colon S_i \subseteq X_j} \lambda_j,$$

which, combined with the fact that $V = \{0\}$, implies that v_j are uniquely determined by the choice of μ_j. Finally, relationships in (6.2) show once again that λ_j lies in an (at most) one-dimensional subspace, as desired. □

The following corollary is just a modification of the lemma in the case when all the sets are non-empty. The need for this additional technicality comes from Observation 6.2.

Corollary 6.4. *Let $m, k \in \mathbb{N}$. Given any distinct non-empty sets $X_1, X_2, \ldots, X_r \in \mathbb{N}^{(<\omega)}$, we can find non-empty sets $S_1, S_2, \ldots, S_r \subseteq \mathbb{N}^{(\leqslant m)}$ which enjoy the following properties:*

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

(i) $n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) = 0$, *provided* $r < 2^{m+1} - 1$,
(ii) $n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) \leqslant 1$, *provided* $r < 3 \cdot 2^m - 1$.

Proof. In both cases, we apply Lemma 6.3 to the distinct sets $\emptyset, X_1, X_2, \dots, X_r$ to find sets $S_0, S_1, \dots, S_r$ of size at most m such that

$$n(\emptyset, X_1, X_2, \dots, X_r; S_0, S_1, S_2, \dots, S_r) \leqslant q,$$

where $q = 0$ if $r < 2^{m+1} - 1$, and $q = 1$ if $r < 3 \cdot 2^m - 1$. We now show that, starting from

$$n(\emptyset, X_1, X_2, \dots, X_r; S_0, S_1, S_2, \dots, S_r) \leqslant q,$$

we can reorder the sets S_i so that

$$n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) \leqslant q,$$

which finishes the proof.

Let I be the matrix $I(\emptyset, X_1, X_2, \dots, X_r; S_0, S_1, S_2, \dots, S_r)$. By the rank-nullity theorem, the rank of I (which is also column rank) is at least $r + 1 - q$. If all the sets S_i are non-empty, then the first column of I is zero. Removing the first row from I, we get a matrix with column rank also $\geqslant r + 1 - q$, thus having row rank also $\geqslant r + 1 - q$. Remove the first row; the remaining matrix is $I(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r)$ and it has row rank at least $r - q$. Thus its rank is at least $r - q$, so by the rank-nullity theorem, $n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) \leqslant q$, as desired.

On the other hand, if $S_0 = \emptyset$ (after reordering if necessary), remove the first row from I, to get a matrix with row rank at least $r - q$, and whose first column is zero. But removing the first column does not change the column rank, and we end up with the matrix $I(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r)$ of column rank $\geqslant r - q$, which by the rank-nullity theorem gives

$$n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) \leqslant q$$

as desired. □

The next corollary is tailored to the analysis of the $\mathcal{F}_{N,d,m}$-incident sets (see Corollary 6.7).

Corollary 6.5. *Suppose that* $X_1, X_2, \dots, X_r \in \mathbb{N}^{(<\omega)}$ *are distinct,* $t \leqslant r$, *and* $S_1, S_2, \dots, S_t \in \mathbb{N}^{(<\omega)}$ *satisfy*

$$n(X_1, X_2, \dots, X_t; S_1, S_2, \dots, S_t) = 0.$$

Provided $r < 3 \cdot 2^m$, *we can find sets* $S_{t+1}, S_{t+2}, \dots, S_r \in \mathbb{N}^{(\leqslant m)}$ *such that*

$$n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) \leqslant 1.$$

If $r < 3 \cdot 2^m - 1$ *and the sets* X_i *are non-empty, then additionally the sets* $S_{t+1}, S_{t+2}, \dots, S_r$ *can be chosen to be non-empty.*

Proof. Apply Lemma 6.3(ii) to get sets $T_1, T_2, \dots, T_r \in \mathbb{N}^{(\leqslant m)}$ such that

$$n(X_1, X_2, \dots, X_r; T_1, T_2, \dots, T_r) \leqslant 1,$$

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

or Corollary 6.4 to make the sets T_i non-empty, provided $r < 3 \cdot 2^m - 1$ holds and the sets X_i are non-empty. Look at the $(t+r) \times r$ matrix

$$I(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_t, T_1, T_2, \ldots, T_r).$$

We shall remove t rows from those corresponding to $T_1, T_2, \ldots, T_r$ to get the desired matrix. The following row-removal lemma does this for us.

Lemma 6.6. *Suppose that A is an $(r+t) \times r$ matrix with the first t rows linearly independent and $t \leqslant r$. Then we can remove t rows from the last r rows of A, so that the kernel of A is not affected.*

Proof. If $I \subseteq [r+t]$, let A_I stand for the matrix formed from the rows of A with indices in I. Starting from the set $I = [r+s]$, we shall iteratively remove elements greater than t from I, so that at each step we have $\ker A_I = \ker A$.

Suppose that we have $I \subseteq [r+t]$ with $[t] \subseteq I$, but $|I| > r$, such that $\ker A_I = \ker A$ holds. If we can pick an element $x \in I \setminus [t]$, so that $\ker A_{I \setminus \{x\}} = \ker A_I$, we are done. Otherwise, no such x works. Observe that if a row v^T of A_I is a linear combination of the other rows, then it can be removed from A_I. To spell it out, write v_i^T for ith row of A and suppose that $v_i^T = \sum_{j \in I \setminus \{i\}} \lambda_j v_j^T$. Then, if $\mu \in \ker A_{I \setminus \{i\}}$, we have $\mu \cdot v_i^T = \sum_{j \in I \setminus \{i\}} \lambda_j \mu \cdot v_j^T = 0$. So $\ker A_{I \setminus \{x\}} = \ker A_I$.

Thus, the vectors $v_1^T, \ldots, v_t^T$ are linearly independent, and $v_i \notin \operatorname{span}\{v_j : j \in I \setminus \{i\}\}$ for $i \in I \setminus [t]$. But, then, $|I| > r$ and the rows of I are linearly independent vectors of the dimension r, which is a contradiction. Hence, we can proceed until we reach $|I| = r$, as desired. □

The matrix

$$I(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_t, T_1, T_2, \ldots, T_r)$$

satisfies the conditions of the lemma since $n(X_1, X_2, \ldots, X_t; S_1, S_2, \ldots, S_t) = 0$. By applying the lemma, we can pick $S_{t+1}, S_{t+2}, \ldots, S_r$ among the sets in $T_1, T_2, \ldots, T_r$ so that

$$\begin{aligned} &n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r) \\ &\quad = n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_t, T_1, T_2, \ldots, T_r) \\ &\quad \leqslant n(X_1, X_2, \ldots, X_r; T_1, T_2, \ldots, T_r) \leqslant 1. \end{aligned}$$ □

Similarly to Corollary 5.2, the next corollary is consistent with the vague idea that the combinatorial subspaces are the source of the non-trivial $\mathcal{F}_{N,d,m}$-incident sets. In particular, we show that the $\mathcal{F}_{N,d,m}$-incident sets behave like the combinatorial subspaces when it comes to taking unions: the size of a union of two $\mathcal{F}_{N,d,m}$-incident sets of size 2^{m+1} is at least $3 \cdot 2^m$.

Corollary 6.7. *Let $d, m \in \mathbb{N}$ be given.*

(i) *If $T \subseteq \{0,1\}^N$ is $\mathcal{F}_{N,d,m}$-incident, then $|T| \geqslant \min\{d+2, 2^{m+1}\}$.*

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

(ii) *If* $T_1, T_2 \subseteq \{0,1\}^N$ *are distinct, of size at most* $d+1$ *and minimal (with respect to inclusion)* $\mathcal{F}_{N,d,m}$*-incident, then* $|T_1 \cup T_2| \geqslant 3 \cdot 2^m$.

Proof. *Part (i).* Suppose that $T = \{x_0, x_1, x_2, \dots, x_r\} \subseteq \{0,1\}^N$ is $\mathcal{F}_{N,d,m}$-incident and that $r < 2^{m+1} - 1, d+1$. Note that, for a fixed set A, the map $X \mapsto X \Delta A$, induces a reflection $\alpha : \mathbb{R}^N \to \mathbb{R}^N$ (with the natural correspondence between sets and points in $\{0,1\}^N$). In particular, taking A to be the set of non-zero coordinates of x_0, we have an affine isomorphism α that preserves the cube $\{0,1\}^N$ and sends x_0 to zero. Let $X_i \subseteq [N]$ be the set corresponding to $\alpha(x_i)$, that is, the set of indices j such that $\alpha(x_i)_j = 1$. As $r < 2^{m+1} - 1$, Corollary 6.4 yields non-empty sets $S_1, S_2, \dots, S_r \subseteq [N]$ of size at most m, such that

$$n(X_1, X_2, \dots, X_r; S_1, S_2, \dots, S_r) = 0.$$

Choosing ($\leqslant m$)-functions $\varphi_1, \varphi_2, \dots, \varphi_r$ with images $S_1, S_2, \dots, S_r$ respectively, we obtain that the vectors

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(\alpha(x_1)) \\ \varphi_2(\alpha(x_1)) \\ \vdots \\ \varphi_r(\alpha(x_1)) \end{pmatrix}, \quad \begin{pmatrix} \varphi_1(\alpha(x_2)) \\ \varphi_2(\alpha(x_2)) \\ \vdots \\ \varphi_r(\alpha(x_2)) \end{pmatrix}, \dots, \begin{pmatrix} \varphi_1(\alpha(x_r)) \\ \varphi_2(\alpha(x_r)) \\ \vdots \\ \varphi_r(\alpha(x_r)) \end{pmatrix}$$

are affinely independent. However, as $r \leqslant d$, Proposition 4.4 applies to T, affine map α and functions $\varphi_1, \varphi_2, \dots, \varphi_r$, which tells us that these vectors are affinely dependent, which is a contradiction. Thus $|T| = r + 1 \geqslant \min\{2^{m+1}, d+2\}$, as desired.

Part (ii). If T_1, T_2 are disjoint, then by part (i), $|T_1 \cup T_2| \geqslant 2^{m+2}$, so we are done. Thus, assume that some t_0 belongs to both sets. Pick an affine isomorphism $\alpha : \mathbb{R}^N \to \mathbb{R}^N$ which sends t_0 to zero and preserves the cube $\{0,1\}^N$ (given by a suitable reflection). Let $X_1, X_2, \dots, X_t$ be the sets that correspond to the non-zero points of $\alpha(T_1 \cap T_2)$. Next, let $X_{t+1}, \dots, X_{t+r_1}$ be the sets that correspond to points in $\alpha(T_1 \setminus T_2)$ and let $X_{t+r_1+1}, \dots, X_{t+r_1+r_2}$ be the sets corresponding to points of $\alpha(T_2 \setminus T_1)$. If $|T_1 \cup T_2| \geqslant 3 \cdot 2^m$, we are done. Otherwise $1 + t + r_1 + r_2 = |T_1 \cup T_2| < 3 \cdot 2^m$.

Since they are minimal and distinct, T_1, T_2 cannot contain one another. So $T_1 \cap T_2$ is a proper subset of T_1 and hence it is not $\mathcal{F}_{N,d,m}$-incident. Therefore, by Observation 6.2, we can find non-empty sets $S_1, S_2, \dots, S_t \in \mathbb{N}^{(\leqslant m)}$ such that

$$n(X_1, X_2, \dots, X_t; S_1, S_2, \dots, S_t) = 0.$$

Applying Corollary 6.5 (as $r + t_1 + t_2 < 3 \cdot 2^m - 1$), we obtain non-empty sets

$$S_{t+1}, \dots, S_{t+r_1+r_2} \in \mathbb{N}^{(\leqslant m)}$$

such that

$$n(X_1, X_2, \dots, X_{r+t_1+t_2}, S_1, S_2, \dots, S_{r+t_1+t_2}) \leqslant 1.$$

Now, take any ($\leqslant m$)-functions $\varphi_1, \dots, \varphi_{t+r_1+r_2}$ to $[N]$ with images $S_1, S_2, \dots, S_{t+r_1+r_2}$, and let $x_i \in T_1 \cup T_2$ be the point such that the set X_i corresponds to $\alpha(x_i)$. Write y_i for the vector $(y_i)_j = \varphi_j(x_i)$, $j = 1, 2, \dots, t + r_1 + r_2$. Thus, $y_1, y_2, \dots, y_t$ are linearly independent

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

and the rank of $y_1, y_2, \ldots, y_{t+r_1+r_2}$ is at least $t + r_1 + r_2 - 1$. Since $|T_1| \leqslant d + 1$, we can apply Proposition 4.4 to T_1, map α and functions $f_1, \ldots, f_{t+r_1}$. Note that since the sets S_i are non-empty, we have $f_i(0) = 0$ for all i. Thus, vectors $y_1, y_2, \ldots, y_{t+r_1}$ have rank at most $t + r_1 - 1$. Similarly, the rank of the vectors $y_1, y_2, \ldots, y_t, y_{t+r_1+1}, \ldots, y_{t+r_1+r_2}$ is at most $t + r_2 - 1$.

To obtain a contradiction, look at

$$\begin{aligned} U &= \operatorname{span}\{y_1, y_2, \ldots, y_{t+r_1}\}, \\ V &= \operatorname{span}\{y_1, y_2, \ldots, y_t, y_{t+r_1+1}, y_{t+r_1+2}, \ldots, y_{t+r_1+r_2}\}, \\ W &= \operatorname{span}\{y_1, y_2, \ldots, y_{t+r_1+r_2}\}, \\ Z &= \operatorname{span}\{y_1, y_2, \ldots, y_t\}. \end{aligned}$$

Thus, $\dim Z = t, \dim U \leqslant t + r_1 - 1, \dim V \leqslant t + r_2 - 1, \dim W \geqslant t + r_1 + r_2 - 1$, $Z \subseteq U$, $V \subseteq W$ and $W = U + V$. Therefore $W/Z = U/Z + V/Z$. Finally,

$$\begin{aligned} r_1 + r_2 - 1 &\leqslant \dim W - \dim Z = \dim W/Z \\ &\leqslant \dim U/Z + \dim V/Z \\ &\leqslant r_1 - 1 + r_2 - 1 = r_1 + r_2 - 2, \end{aligned}$$

which is a contradiction. □

Theorem 6.8. *Suppose that $d, m \in \mathbb{N}$ satisfy $2^{m+1} - 1 \leqslant d \leqslant 3 \cdot 2^m - 3$. Let $N \geqslant 1$. Then*

$$\alpha(2^N, d) \leqslant \left(\frac{25}{N}\right)^{1/2^{m+1}} \cdot 2^N.$$

Proof. Let $X = \{0, 1\}^N \subseteq \mathbb{R}^N$. Applying Proposition 3.1, we obtain a function $f \in \operatorname{span} \mathcal{F}_{N,d,m}$, which is a bijection onto its image when restricted to X and such that if $S \subseteq f(X)$ is affinely dependent then $f^{-1}(S)$ is $\mathcal{F}_{N,d,m}$-incident. Let $Y = f(X)$. Note that $|Y| = 2^N$, since f is injective on X. We claim that the set Y has no more than $d + 1$ points on a same hyperplane, but all sufficiently large subsets of Y have $d + 1$ cohyperplanar points.

No more than $d + 1$ points on a hyperplane. Suppose that we have a cohyperplanar set $S = \{s_1, s_2, \ldots, s_{d+2}\} \subseteq Y$. Pick a maximal affinely independent subset $S' \subseteq S$. Without loss of generality $S' = \{s_1, \ldots, s_r\}$, for some r. As S' is a subset of a hyperplane, we have $r \leqslant d$. Look at $S'_1 = S' \cup \{s_{d+1}\}$. By the choice of S', the set S'_1 is not affinely independent. By the choice of f, the pre-image $f^{-1}(S'_1)$ is $\mathcal{F}_{N,d,m}$-incident. Find a subset T_1 of $f^{-1}(S'_1)$ which is minimal $\mathcal{F}_{N,d,m}$-incident, and arbitrary point p in T_1. We also have $S'_2 = S \setminus \{p\}$ affinely dependent, as it is a subset of a hyperplane of size at $d + 1$. By the choice of f, $f^{-1}(S'_2)$ is $\mathcal{F}_{N,d,m}$-incident, and has a minimal $\mathcal{F}_{N,d,m}$-incident subset T_2. Note that $p \in T_1 \setminus T_2$, so T_1, T_2 are distinct, and $|T_1|, |T_2| \leqslant d + 1$. Corollary 6.7(ii) applies to give $d + 2 = |S| \geqslant |T_1 \cup T_2| \geqslant 3 \cdot 2^m > d + 2$, which is a contradiction.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Dense subsets are not in general position. Let $T \subseteq Y$ have size at least

$$\left(\frac{25}{N}\right)^{1/2^{m+1}} 2^N,$$

Then, by Theorem 6.1, $f^{-1}(T)$ contains an $(m+1)$-dimensional combinatorial subspace. Applying Lemma 4.1, we have that the points of $T = f(f^{-1}(T))$ are affinely dependent. Adding any $d+1-2^{m+1}$ points to the set T proves the claim. □

7. Conclusion

Even though there are now some non-trivial lower bounds on $\alpha(n,d)$ [1], the gap between the lower and the upper bounds remains very large. Of course, the first question is still to determine the $\alpha(n,d)$. Regarding the current lower bounds on $\alpha(n,2)$, both in [3] and in [1], we note that their proofs are based on some relatively general probabilistic estimates of the independence number of hypergraphs. However, these approaches use very little of the structure of the given sets of points. In fact, possible algebraic properties of such sets have not been exploited. For example, if X is a set of points with no more than three on a line, but with no dense set in general position, we can expect that plenty of pairs of points in X have a unique third point in X on their line. This gives rise to an algebraic operation: given two points x_1, x_2 of X, set $x_1 * x_2$ to be the third point of X on their line, if such a point exists. Of course, there is an issue of how to define $x_1 * x_2$ for all pairs, but at least for plenty of pairs it can be defined. Possibly, if X is a set for which the $\alpha(|X|,2)$ is attained, we could deduce some properties of the operation $*$.

Leaving determination of $\alpha(n,d)$ aside, note that it would be surprising if the $\alpha(n,d)/n$ did not decrease in d. In particular, the current situation with the upper bounds is that we have infinitely many d for which $\alpha(n,d)/n = O(1/\log^{\beta_d} n)$ for some $\beta_d > 0$, while for infinitely many other d the bounds for $\alpha(n,d)/n$ currently come from the general density Hales–Jewett theorem, and are roughly comparable to the inverse of the Ackermann function. It is very unlikely that $\alpha(n,d)/n$ is actually close to these estimates for all d. An obvious question is the following.

Question 7.1. What is the relationship between $\alpha(n,d_1)$ and $\alpha(n,d_2)$ for $d_1 < d_2$? Do we always have $\alpha(n,d)/n \geqslant \alpha(n,d+1)/n$?

Another question is the relationship between the bounds in the density Hales–Jewett theorem and the $\alpha(n,d)$. Do these have to be related?

Finally, some key tools in this paper were the algebraic Lemmas 5.1 and 6.3. It could be of interest to study $n(X_1, X_2, \ldots, X_r; S_1, S_2, \ldots, S_r)$ further.

Acknowledgements

I would like to thank Trinity College and the Department of Pure Mathematics and Mathematical Statistics of Cambridge University for their generous support, and Imre

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press

Leader for helpful discussions concerning this paper. I am also grateful to the anonymous referee for a careful reading and for comments that led to a clarification of the paper.

References

[1] Cardinal, J., Tóth, C. and Wood, D. R. (2017) General position subsets and independent hyperplanes in d-space. *Journal of Geometry* **108** 33–43.
[2] Erdős, P. (1986) On some metric and combinatorial geometric problems. *Discrete Math.* **60** 147–153.
[3] Füredi, Z. (1991) Maximal independent subsets in Steiner systems and in planar sets. *SIAM J. Discrete Math.* **4** 196–199.
[4] Furstenberg, H. and Katznelson, Y. (1991) A density version of the Hales–Jewett theorem. *J. d'Analyse Mathématique* **57** 64–119.
[5] Polymath, D. H. J. (2012) A new proof of the density Hales–Jewett theorem. *Ann. of Math.* **175** 1283–1327.

https://doi.org/10.1017/S0963548317000098 Published online by Cambridge University Press