An alternate proof of Szemerédi's cube lemma using extremal hypergraphs

David S. Gunderson*and Vojtěch Rödl
Emory University, Atlanta GA 30322

Abstract

A collection H of integers is called an *affine d-cube* if there exist $d+1$ positive integers $x_0, x_1, \ldots, x_d$ so that

$$H = \left\{ x_0 + \sum_{i \in I} x_i : I \subseteq \{1, 2, \ldots, d\} \right\}.$$

In 1969, Szemerédi found a density result for affine cubes, namely, that for any positive integer d, there exists a constant c so that if $A \subseteq \{1, 2, \ldots, n\}$ and $|A| \geq cn^{1-\frac{1}{2^d}}$, then A contains an affine d-cube. Using extremal hypergraphs, we offer an entirely different proof of this fact (though with worse constant) which also yields a slightly stronger statement.

1 Introduction

For any positive integer m we use the notation $[m] = [1, m] = \{1, 2, \ldots, m\}$.

Definition 1.1 *A collection H of integers is called a d-*dimensional affine cube*, or more simply, an* affine d-cube *if there exist $d+1$ positive integers $x_0, x_1, \ldots, x_d$ so that*

$$H = \left\{ x_0 + \sum_{i \in I} x_i : I \subseteq [d] \right\}. \tag{1}$$

If all sums in (1) are distinct, then $|H| = 2^d$, and thus H is saturated, or replete.

For a set X we use the standard notations $\mathcal{P}(X) = \{Y : Y \subseteq X\}$ and $[X]^s = \{S \subset X : |S| = s\}$. It will often be convenient to use $X = [n]$. An arithmetic progression of length k will be referred to as an AP_k.

If an affine d-cube H is generated by $x_0, x_1, \ldots, x_d$, then we write $H = H(x_0, x_1, \ldots, x_d)$. For example, $H(1,1,1) = \{1,2,3\}$, while a replete affine 2-cube is $H(1,3,9) = \{1,4,10,13\}$. Note that $H(x_0, x_1, \ldots, x_d)$ may differ from, say, $H(x_1, x_0, \ldots, x_d)$.

In 1892, Hilbert [11] proved a non-trivial Ramsey-type theorem for affine cubes. The following is a finite version thereof, following from Hilbert's original result by compactness.

*Partially supported by DFB-SFB 343, Diskrete Structuren in der Mathematik, Bielefeld

Theorem 1.2 (Hilbert [11]) *For every r, d, there exists a least number $h(d,r)$ so that for every coloring*

$$\chi : [h(d,r)] \to [r],$$

there exists an affine d-cube monochromatic under χ.

The original statement of the theorem asserted that if one colored the positive integers with finitely many colors, then one color class contained a (monochromatic) d-cube. Theorem 1.2 follows from van der Waerden's Theorem [17] since an AP_{d+1} given by $\{a, a+k, a+2k, \ldots, a+dk\}$ is the affine d-cube $H(a, k, k, \ldots, k)$. See [1] for further references.

In 1969, Szemerédi [15] proved that if a set A of positive integers has positive upper density, (i.e., $\overline{\lim_{n\to\infty}} |A \cap [1,n]| > 0$) then A contains an AP_4. In that same paper[15] (Lemma $p(\delta, l)$, p. 93), Szemerédi gave a density version of Theorem 1.2; Szemerédi's so-called "cube lemma" can also be found in [8]. Explicit bounds which follow directly from Szemeré'di's argument are also mentioned in [6].

Theorem 1.3 (Szemerédi [15]) *For each d, there exists a constant c so that for sufficiently large n, if $A \subset [n]$ and*

$$|A| \geq cn^{1-\frac{1}{2^d}}, \tag{2}$$

then A contains an affine d-cube.

Szemerédi's 1975 paper [16] proving a density result for arbitrarily long arithmetic progressions has occasionally been errantly attributed as the source of the cube lemma. Since the 1975 paper is so complicated, such claims were hard to verify. Perhaps as a result, typographical misprints such as the "2" in the exponent of (2) printed as "e" have survived—but with positive effect. One such misprint led the present authors to find the following following slight improvement, but with a more complicated proof:

Theorem 1.4 *For each integer $d \geq 2$ there exists an n_0 so that for every $n \geq n_0$, if $A \subset [1,n]$ satisfies*

$$|A| \geq (1+\delta)2^{d-3/2}n^{1-1/2^d} \tag{3}$$

then A contains $(1-o(1))n^2/2^{d-1}$ replete affine d-cubes.

We make a few comments on some subtle differences between our density result and Szemerédi's. The results given in this paper guarantee not only one d-cube as does Szemerèdi's, but on the order of n^2 of them. It is not clear what the precise order of magnitude should be, since by a result in [4] (Theorem 2, with $k = 2^{d+1}$ and $l = 2^{d+1} - (d+1)$), for a fixed $\epsilon \in (0,1]$ one can conclude that if $A \subset [1,n]$ satisfies $|A| \geq \epsilon n$, then A contains on the order of n^{d+1} affine d-cubes, that is, a positive fraction of all such cubes. We also observe that our results guarantee *replete* affine cubes.

With Theorem 1.3 in hand, we observe that the upper bound for $h(d,r)$ appearing in [1] is now immediately improved by applying the pigeonhole principle. The details have been noted in [9] and [10].

2 The density result; Proof of Theorem 1.4

The basic idea used here for proving Theorem 1.4 is to first give an extremal result for a very special class of partite hypergraphs, then to interpret hyperedges thereof as numbers.