On sumfree subsets of hypercubes

by

Daniel J. Katz (Providence, RI)

1. Introduction. Given an additive group Z, we refer to $A \subset Z$ as a *sumfree set* if $x + y \neq z$ for all $x, y, z \in A$. (Equivalently, using the notation of sumsets, A is sumfree if $(A+A) \cap A = \emptyset$.) These sets have been of interest since at least 1916, when Schur [10] proved that the positive integers could not be partitioned into finitely many such sets.

A common problem in this topic is as follows: given a particular additive group Z (or perhaps a subset Z of an additive group), how large can a sumfree subset of Z be, and further, what sort of structure do large sumfree subsets have? This problem has been considered for $Z = \mathbb{Z}_{>0}$ (see [2]), $\mathbb{Z}/p\mathbb{Z}$ (see [4, 9]), general finite groups (abelian [6] and non-abelian [7]), and $\{1, \ldots, n\} \subset \mathbb{Z}$ for arbitrary (usually large) n (see [1, 3, 11]).

The last of these cases suggests a study of the "discrete hypercube" $Z = \{1, \ldots, n\}^k \subset \mathbb{Z}^k$ for $k > 1$. In particular, we would like to know how proportionately large a sumfree subset of $\{1, \ldots, n\}^k$ can be when n is large. For this purpose, we define

$$(1) \qquad c_k := \limsup_{n \to \infty} \frac{1}{n^k} \max\{\#S : S \subset \{1, \ldots, n\}^k \text{ is sumfree}\}.$$

Previous work on sumfree subsets of $\{1, \ldots, n\}$ has shown that $c_1 = 1/2$. (For example, the set $\{\lfloor n/2 \rfloor + 1, \ldots, n\}$ is optimally large for all n, as is the set of odd elements.) Let S be a sumfree subset of $\{1, \ldots, n\}^k$ of size αn^k, and let $k' > k$. The inverse image S' of S under a natural projection from $\{1, \ldots, n\}^{k'}$ to $\{1, \ldots, n\}^k$ is also sumfree, and has size $\alpha n^{k'}$. In view of this fact, it is clear that $c_{k'} \geq c_k$ for $k' > k$, and thus $1/2 \leq c_k \leq 1$ for all k.

The largest sumfree subsets we have observed in the square $\{1, \ldots, n\}^2$ take the form of thick diagonal "stripes"; generalizing this construction, we can construct large sumfree subsets in $\{1, \ldots, n\}^k$ and thus prove a general lower bound for c_k.

THEOREM 1.1. *Define c_k as in* (1). *Then*

$$c_k \geq 1 - \frac{2}{k!} \sum_{i=0}^{\lfloor k/3 \rfloor} (-1)^i \binom{k}{i} \left(\frac{k}{3} - i \right)^k.$$

Analysis of this lower bound yields:

COROLLARY 1.2. *Define c_k as in* (1). *Then*

$$\lim_{k \to \infty} c_k = 1.$$

These statements are proved in Section 3.

We also prove a general upper bound for c_k using a combinatorial method, although it is difficult to write this bound as an explicit function of k.

THEOREM 1.3. *Define c_k as in* (1), *and let α^* be the unique root in* $[1/2, 1]$ *of the equation*

$$\alpha = (2 - 2\alpha) \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i.$$

Then

$$c_k \leq \alpha^*.$$

Our approach to the upper bound depends on the idea that if an element of a sumfree subset $S \in \{1, \ldots, n\}^k$ is the sum of many pairs of elements, none of these pairs can be in S. This means that if S contains a certain proportion of the full set, a certain number of elements cannot belong to S, which causes a contradiction if the proportion is large. This argument is given in Section 4.

To give an idea of the distance between our lower and upper bounds, here are the approximate bounds given by these theorems for $2 \leq k \leq 6$:

$$0.5555 \leq c_2 \leq 0.7880, \quad 0.7966 \leq c_5 \leq 0.9385,$$
$$0.6666 \leq c_3 \leq 0.8613, \quad 0.8388 \leq c_6 \leq 0.9586.$$
$$0.7407 \leq c_4 \leq 0.9080,$$

Our calculations for both the lower and upper bounds involve approximating numbers of lattice points in $\{1, \ldots, n\}^k$ by integrating over subsets of $[0, n]^k$. This approximation is less than exact, but the error becomes small compared to n^k when n is large, and thus it ultimately does not affect the value of c_k. These integrals become more complicated as k grows, but they can be calculated explicitly by induction, whereas counting the lattice points directly becomes cumbersome in higher dimensions.

This integral method suggests a non-discrete version of the problem: maximizing the volume of Lebesgue-measurable sumfree subsets of the "continuous hypercube" $[0, 1]^k \subset \mathbb{R}^k$. In Section 5, we will see that the bounds we calculated in Theorems 1.1 and 1.3 hold in this setting as well.

By applying an iterative process to the upper-bound argument, we can improve the result of Theorem 1.3 as follows:

THEOREM 1.4. *Define c_k as in* (1), *and let α^{**} be the unique root in $(1/2, 1)$ of the equation*

$$\alpha = \frac{1}{2} - \alpha + \sum_{i=k}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i.$$

Then

$$c_k \leq \alpha^{**}.$$

This theorem is proved in Section 6, and yields the following improved bounds for c_k:

$$0.5555 \leq c_2 \leq 0.7274, \quad 0.7966 \leq c_5 \leq 0.9351,$$
$$0.6666 \leq c_3 \leq 0.8407, \quad 0.8388 \leq c_6 \leq 0.9572.$$
$$0.7407 \leq c_4 \leq 0.9000,$$

This result, like Theorem 1.3, is also applicable in the continuous setting.

We will also discuss in Section 7 some results that generalize our processes to l-fold-sumfree sets, that is, sets A such that $x_1 + \cdots + x_l \neq z$ for all $x_1, \ldots, x_l, z \in A$. The lower bound for sumfree sets extends easily to $l > 2$; the upper bound is difficult to apply when $l > 4$, but interestingly in the $l = 3$ case it gives a bound which is explicit rather than the root of an equation.

Finally, in Section 8, we will present some concluding remarks, suggesting two divergent paths for future investigation in the subject.

2. Introductory lemmas. In order to bound the constants under consideration, we will need the following volume formula.

LEMMA 2.1. *Given $a \in [0, k]$, the volume of the region*

$$\{(x_1, \ldots, x_k) \in [0, 1]^k : x_1 + \cdots + x_k \leq a\} \subset \mathbb{R}^k$$

is equal to

$$\frac{1}{k!} \sum_{i=0}^{\lfloor a \rfloor} (-1)^i \binom{k}{i} (a - i)^k.$$

Proof. This is a special case of Theorem 1 in Section I.9 of [5]. ∎

REMARK 2.1. The proof in [5] uses probability theory, but the formula can also be obtained directly using an inclusion-exclusion argument. The latter proof is useful in that it can easily be adapted to count the number of lattice points in the region; however, we will not use this application, so we omit the alternative proof.

We will also need the following integral formula, easily proven by induction.

LEMMA 2.2. *Given $c \in (0, 1]$, we have the following equality*:

$$\int_c^1 dx_1 \int_{c/x_1}^1 dx_2 \int_{c/x_1 x_2}^1 dx_3 \cdots \int_{c/x_1 \cdots x_{k-1}}^1 dx_k = 1 - c \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{c} \right)^i.$$

REMARK 2.2. In a sense, the domain of integration is a multiplicative analogue of the k-simplex. Also note that the right side of the equation approaches 0 as $k \to \infty$, since the sum is a truncated Maclaurin series for e^x evaluated at $x = -\ln c$.

Proof of Lemma 2.2. Let $J(k, c)$ represent the left side of the equation. The assertion is clearly true when $k = 1$, so we proceed by induction on k. Assume the statement is true for k; then

$$J(k+1, c) = \int_c^1 J\left(k, \frac{c}{x_1}\right) dx_1 = \int_c^1 \left(1 - \frac{c}{x_1} \sum_{i=0}^{k-1} \frac{1}{i!} (\ln x_1 - \ln c)^i\right) dx_1$$

$$= (1 - c) - c \sum_{i=0}^{k-1} \frac{1}{i!} \int_0^{-\ln c} (\ln x_1 - \ln c)^i \, d(\ln x_1 - \ln c)$$

$$= (1 - c) - c \sum_{i=0}^{k-1} \frac{1}{(i+1)!} (-\ln c)^{i+1} = 1 - c \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{c}\right)^i,$$

and thus the assertion holds for all k. ∎

Finally, we quote a theorem from Lang, adapted for our purposes, which will allow us to use integrals to approximate subsets of lattices.

THEOREM 2.3 (Lang). *Let D be a subset of $[0, 1]^k$ such that the boundary of D has a Lipschitz-continuous parametrization in $k - 1$ variables, and let $nD = \{nx : x \in D\}$. Then*

$$\#(\{1, \ldots, n\}^k \cap nD) = n^k \operatorname{Vol}(D) + O(n^{k-1}).$$

Proof. Apply [8, Theorem 2, p. 128] with $L = \mathbb{Z}^k$ and $F = (0, 1]^k$. There are fewer than $k(n + 1)^{k-1}$ lattice points in the intersection of nD and $(\{0, \ldots, n\}^k \backslash \{1, \ldots, n\}^k)$, and these can be absorbed into the error term. ∎

3. Bounding c_k from below. One method of generating sumfree sets in $\{1, \ldots, n\}$ is to consider "cross-section" sets

(2) $$K_a := \{(x_1, \ldots, x_k) \in \{1, \ldots, n\}^k : x_1 + \cdots + x_k = a\}.$$

If A is a sumfree set in $\{k, k+1, \ldots, kn\}$, the set $S = \bigcup_{a \in A} K_a$ is sumfree, because if $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_k)$ are both contained in S, then

$$(x_1 + y_1) + \cdots + (x_k + y_k) = (x_1 + \cdots + x_k) + (y_1 + \cdots + y_k) \notin A,$$

so the sum of these two elements is not in S.

We will determine a lower bound for c_k using sets of the form

$$S(n, k, a) := \{(x_1, \ldots, x_k) \in \{1, \ldots, n\}^k : a \le x_1 + \cdots + x_k < 2a\}.$$

Since $\{a, \ldots, 2a - 1\}$ is clearly sumfree in $\{k, k + 1, \ldots, kn\}$, $S(n, k, a)$ is sumfree. To obtain an optimal lower bound for this method, we need to choose a value of a that maximizes the size of $S(n, k, a)$. We approximate this size using the region

$$\widetilde{S}(n, k, a) := \{(x_1, \ldots, x_k) \in [0, n]^k \subset \mathbb{R}^k : a \le x_1 + \cdots + x_k < 2a\}.$$

Note that since $\widetilde{S}(1, k, a)$ is just a scaled-down copy of $\widetilde{S}(n, k, an)$, we have

$$\widetilde{S}(n, k, an) = n^k \widetilde{S}(1, k, a).$$

Proof of Theorem 1.1. By Lemma 2.1, the volume of

$$\{(x_1, \ldots, x_k) \in [0, 1]^k \subset \mathbb{R}^k : x_1 + \cdots + x_k < a\}$$

is equal to

$$V_1(k, a) := \frac{1}{k!} \sum_{i=0}^{\lfloor a \rfloor} (-1)^i \binom{k}{i} (a - i)^k.$$

Changing variables, we see that the volume of

$$\{(x_1, \ldots, x_k) \in [0, 1]^k \subset \mathbb{R}^k : x_1 + \cdots + x_k > 2a\}$$

is equal to

$$V_2(k, a) := \frac{1}{k!} \sum_{i=0}^{\lfloor k - 2a \rfloor} (-1)^i \binom{k}{i} (k - 2a - i)^k.$$

We wish to choose a value of a (for each k) which maximizes

$$\mathrm{Vol}(\widetilde{S}(1, k, a)) = 1 - V_1(k, a) - V_2(k, a).$$

A computer search (for $k < 60$) suggests the optimal choice satisfies $a = k/3 + O(1)$, although it is difficult to determine an exact formula. For our lower bound, we choose $a = k/3$; this value appears to be close to optimal, and it gives a concise expression for $\mathrm{Vol}(\widetilde{S}(1, k, a))$ (since $V_1(k, k/3) = V_2(k, k/3)$). The regions $\widetilde{S}(1, k, k/3)$ for $k = 2, 3$ are shown in Figures 1 and 2.

By Theorem 2.3, $\#S(n, k, kn/3) = \mathrm{Vol}(\widetilde{S}(n, k, kn/3)) + O(n^{k-1})$, since all of the boundaries of the region are hyperplanes and are thus Lipschitz
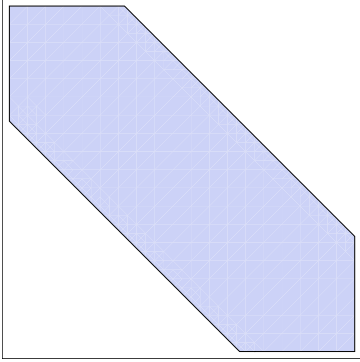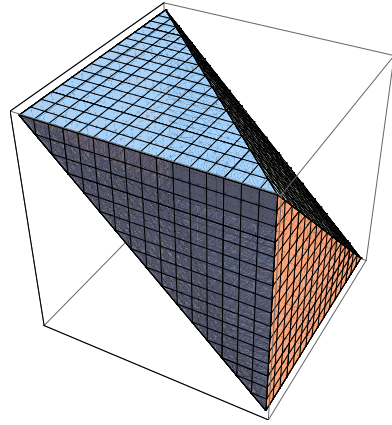
Fig. 1. Sumfree region for $k = 2$

Fig. 2. Sumfree region for $k = 3$

parametrizable. Then we have

$$c_k \geq \limsup_{n \to \infty} \frac{1}{n^k} \#S(n, k, kn/3) = \limsup_{n \to \infty} \frac{1}{n^k} \left(\mathrm{Vol}(\widetilde{S}(n, k, kn/3)) + O(n^{k-1}) \right)$$

$$= \limsup_{n \to \infty} \frac{1}{n^k} \left(n^k \, \mathrm{Vol}(\widetilde{S}(1, k, k/3)) + O(n^{k-1}) \right)$$

$$= \mathrm{Vol}(\widetilde{S}(1, k, k/3)) = 1 - \frac{2}{k!} \sum_{i=0}^{\lfloor k/3 \rfloor} (-1)^i \binom{k}{i} \left(\frac{k}{3} - i \right)^k.$$

This last expression is the lower bound we wished to prove. ■

To determine the behavior of this bound, we need the following lemma.

LEMMA 3.1. *Let a, b satisfy $0 < a < b \leq 1/3$ and*

$$\frac{1}{3} - a < \frac{b^b (1 - b)^{1-b}}{e}.$$

Then

$$\lim_{k \to \infty} \frac{1}{k!} \sum_{i=\lceil ak \rceil}^{\lfloor bk \rfloor} \binom{k}{i} \left(\frac{k}{3} - i \right)^k = 0.$$

Proof. We compute

$$\limsup_{k \to \infty} \frac{1}{k!} \sum_{i=\lceil ak \rceil}^{\lfloor bk \rfloor} \binom{k}{i} \left(\frac{k}{3} - i \right)^k \leq \limsup_{k \to \infty} \frac{1}{k!} \sum_{i=\lceil ak \rceil}^{\lfloor bk \rfloor} \binom{k}{bk} \left(\frac{k}{3} - ak \right)^k$$

$$\leq \limsup_{k \to \infty} \frac{(b - a + 1)k}{(bk)!(k - bk)!} \, k^k \left(\frac{1}{3} - a \right)^k.$$

By Stirling's approximation, this last expression is equal to

$$\limsup_{k \to \infty} \frac{b - a + 1}{2\pi\sqrt{b(1-b)}} \left(\frac{(1/3 - a)e}{b^b(1-b)^{1-b}} \right)^k = 0.$$

Since the upper limit is equal to zero, the limit is as well. ∎

Proof of Corollary 1.2. Define a sequence $\{a_i\}$ as follows:

$$a_0 = \frac{1}{3}, \quad a_{i+1} = \frac{1}{3} - \frac{a_i^{a_i}(1 - a_i)^{1-a_i}}{e}.$$

Calculating the initial terms of this sequence, we find that $a_1, \dots, a_6$ are positive and a_7 is negative. Thus we can split the sum as follows:

$$\sum_{i=0}^{\lfloor k/3 \rfloor} (-1)^i \binom{k}{i} \left(\frac{k}{3} - i \right)^k \le \sum_{i=\lceil a_1 k \rceil}^{\lfloor a_0 k \rfloor} + \sum_{i=\lceil a_2 k \rceil}^{\lfloor a_1 k \rfloor} + \dots + \sum_{i=0}^{\lfloor a_6 k \rfloor}.$$

By Lemma 3.1, each of these partial sums approaches zero as k approaches infinity, and so the entire sum does as well. This means that the lower bound determined in Theorem 1.1 approaches 1 as k grows, and therefore so does c_k. ∎

REMARK 3.1. We can alternatively consider the constants

$$\underline{c_k} := \liminf_{n \to \infty} \frac{1}{n^k} \max\{\#S : S \subset \{1, \dots, n\}^k \text{ is sumfree}\},$$

using the limit inferior rather than the limit superior. It is worth noting that Theorem 1.1 and Corollary 1.2 remain true if c_k is replaced by $\underline{c_k}$, which results in stronger statements.

4. Bounding c_k from above. The process of finding an upper bound for c_k is a bit more complicated, since we cannot do so simply by exhibiting a sumfree set. Here our procedure is to assume that our sumfree set has a certain size, and from this we force a contradiction if the set is too large.

Proof of Theorem 1.3. Let S be a sumfree subset of $\{1, \dots, n\}^k$ with $\#S \ge \alpha n^k$. Suppose that $b = (b_1, \dots, b_k)$ is an element of S. There are $\frac{1}{2} \prod_{i=1}^k (b_i - 1)$ disjoint pairs of elements in $\{1, \dots, n\}^k$ which sum to b, unless all of the b_i's are even, in which case there are $\frac{1}{2}(\prod_{i=1}^k (b_i - 1) + 1)$ of them, to account for the point $(b_1/2, \dots, b_k/2)$. Either way, the number of pairs is equal to $\frac{1}{2} b_1 \cdots b_k + O(n^{k-1})$.

At least one element from each of these pairs must be absent from S, so

$$\alpha n^k \le \#S \le n^k - \frac{1}{2} b_1 \cdots b_k + O(n^{k-1})$$

and thus

$$b_1 \cdots b_k \le (2 - 2\alpha)n^k + O(n^{k-1}) = \beta n^k,$$

where $\beta = (2 - 2\alpha) + O(1/n)$.

This "disqualifies" a number of lattice points from being contained in S, namely

$$T(n, k, \alpha) := \{(b_1, \ldots, b_k) \in \{1, \ldots, n\}^k : b_1 \cdots b_k > \beta n^k\}.$$

As in the last section, we will approximate this collection of lattice points by the region

$$\widetilde{T}(n, k, \alpha) := \{(b_1, \ldots, b_k) \in [0, n]^k \subset \mathbb{R}^k : b_1 \cdots b_k > \beta n^k\}.$$

We can calculate the volume of $\widetilde{T}(n, k, \alpha)$ using an integral:

$$\mathrm{Vol}(\widetilde{T}(n, k, \alpha)) = \int_{\beta n}^{n} dx_1 \int_{\beta n/x_1}^{n} dx_2 \int_{\beta n/x_1 x_2}^{n} dx_3 \cdots \int_{\beta n/x_1 \cdots x_{k-1}}^{n} dx_k$$

$$= n^k \int_{\beta}^{d} x_1 \int_{\beta/x_1}^{d} x_2 \int_{\beta/x_1 x_2}^{d} x_3 \cdots \int_{\beta/x_1 \cdots x_{k-1}}^{d} x_k$$

$$= n^k \left(1 - \beta \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{\beta} \right)^i \right),$$

using Lemma 2.2 with $c = \beta$ in the final step. Since, by Theorem 2.3, $\mathrm{Vol}(T(n, k, \alpha)) = \mathrm{Vol}(\widetilde{T}(n, k, \alpha)) + O(n^{k-1})$, this indicates that for any α such that

$$(3) \qquad \alpha = 1 - \frac{1}{2}\beta + O(1/n) \geq f(\beta) := \beta \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{\beta} \right)^i,$$

any set larger than αn^k is simultaneously smaller than αn^k, yielding a contradiction.

Observe that

$$f'(\beta) = \beta \left(\sum_{i=0}^{k-2} \frac{1}{i!} \left(\ln \frac{1}{\beta} \right)^i \right) \left(-\frac{1}{\beta} \right) + \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{\beta} \right)^i = \frac{1}{(k-1)!} \left(\ln \frac{1}{\beta} \right)^{k-1} > 0.$$

Thus, as β increases from 0 to 1, $f(\beta)$ increases monotonically from 0 to 1, while $1 - \beta/2$ decreases monotonically from 1 to 1/2. Therefore, the equation $1 - \beta/2 = f(\beta)$ has a unique root $\beta^* \in [0, 1]$, and letting $\alpha^* = 1 - \beta^*/2$, we must have $\alpha < \alpha^* + O(1/n)$ to avoid a contradiction. Letting n approach infinity, we conclude that $c_k \leq \alpha^*$. ∎

5. A continuous analogue. In the previous two sections, we used the volume of continuous regions to estimate the size of discrete sets. Alternatively, we could have asked our question about the continuous regions in the first place. Let us consider

$$(4) \qquad \widetilde{c}_k := \max\{\mathrm{Vol}(S) : S \subset [0, 1]^k \text{ is measurable and sumfree}\}.$$

THEOREM 5.1. *Define $\widetilde{c}_k$ as in* (4). *Then*

$$\widetilde{c}_k \geq 1 - \frac{2}{k!} \sum_{i=0}^{\lfloor k/3 \rfloor} (-1)^i \binom{k}{i} \left(\frac{k}{3} - i\right)^k.$$

COROLLARY 5.2. *Define $\widetilde{c}_k$ as in* (4). *Then*

$$\lim_{k \to \infty} \widetilde{c}_k = 1.$$

THEOREM 5.3. *Define $\widetilde{c}_k$ as in* (4), *and let α^* be the unique root in* $[1/2, 1]$ *of the equation*

$$\alpha = (2 - 2\alpha) \sum_{i=0}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha}\right)^i.$$

Then

$$\widetilde{c}_k \leq \alpha^*.$$

Proof. The proofs of these statements are virtually identical to the proofs of Theorem 1.1, Corollary 1.2, and Theorem 1.3 respectively. The only difference is that $S(n, k, a) = \widetilde{S}(n, k, a)$ and $T(n, k, a) = \widetilde{T}(n, k, a)$, so there are no error terms to incorporate.

The proof of Theorem 5.3 warrants one additional comment. If S is a Lebesgue-measurable sumfree set, and $(b_1, \ldots, b_k) \in S$, then

$$S_{b_1, \ldots, b_k} := S \cap ([0, b_1] \times \cdots \times [0, b_k]), \quad S'_{b_1, \ldots, b_k} := \{(b_1, \ldots, b_k) - x : x \in S_{b_1, \ldots, b_k}\}$$

are disjoint sets of equal volume and contained in $[0, b_1] \times \cdots \times [0, b_k]$. Therefore we have

$$\mathrm{Vol}(S_{b_1, \ldots, b_k}) \leq \frac{1}{2} b_1 \cdots b_k.$$

This substitutes for the combinatorial argument that begins the proof of Theorem 1.3. ∎

It should be noted that while the constants c_k and $\widetilde{c}_k$ are similar in nature (and indeed we apply similar methods when bounding them), there is no obvious relation between them; it is not even clear which of these values is larger for a given k.

6. Improving the upper bound. The upper bound for c_k and $\widetilde{c}_k$ may be improved by a slightly different approach. Recall the definition of $f(\beta)$ from (3) in the proof of Theorem 1.3, and suppose $\mathrm{Vol}(S) = \alpha$, where $\alpha = f(\alpha)$. This would require S to consist of all of $[0, 1]^k$ except the "integral wedge" $\widetilde{T}(n, k, a)$ that we removed from the upper right corner. But this would mean that S contains *all* of a smaller set $[0, m]^k$. Scaling by a factor of $1/m$ then violates the upper bound we have just determined. We can

improve our upper bound by exploiting this condition and iterating the process. We deal with the continuous setting first.

THEOREM 6.1. *Define* $\widetilde{c}_k$ *as in* (4), *and let* α^{**} *be the unique root in* $(1/2, 1)$ *of the equation*

$$\alpha = \frac{1}{2} - \alpha + \sum_{i=k}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i.$$

Then

$$\widetilde{c}_k \leq \alpha^{**}.$$

Proof. If $\mathrm{Vol}(S) = \alpha$, consider the set $R = [0, (2 - 2\alpha)^{1/k}]^k \in [0, n]^k$, which is disjoint from $\widetilde{T}(n, k, \alpha)$ except for a single point. Since S cannot intersect $\widetilde{T}(n, k, \alpha)$, the *smallest* density $\mathrm{Vol}(S \cap R) / \mathrm{Vol}(R)$ we can achieve is

$$\widetilde{\varphi}_k(\alpha) := \frac{\alpha - (1 - \mathrm{Vol}(\widetilde{T}(n, k, \alpha))) - (2 - 2\alpha)}{2 - 2\alpha} = \frac{\mathrm{Vol}(\widetilde{T}(n, k, \alpha))}{2 - 2\alpha} + \frac{1}{2}$$

$$= \frac{\alpha}{2 - 2\alpha} - \sum_{i=1}^{k-1} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i,$$

where we apply Lemma 2.2 in the final step.

If any $\widetilde{\varphi}_k^m(\alpha) > 1$ (that is, the mth iterate of $\widetilde{\varphi}_k$, not the mth power), we have a contradiction. We wish to show that the function

$$\widetilde{\psi}_k(\alpha) = \widetilde{\varphi}_k(\alpha) - \alpha$$

has a unique root α^{**} in the interval $(0.5, 1)$, and that any $\alpha > \alpha^{**}$ will grow larger than 1 through repeated application of $\widetilde{\varphi}_k$. First we observe that $\widetilde{\psi}_k(0.5) = 0$. On the interval $(1/2, 1)$,

$$\widetilde{\psi}_k(\alpha) = \frac{\alpha}{2 - 2\alpha} - \left(\frac{1}{2 - 2\alpha} - 1 - \sum_{i=k}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i \right) - \alpha$$

$$= \frac{1}{2} - \alpha + \sum_{i=k}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i.$$

Next we compute the first and second derivatives:

$$\widetilde{\psi}_k'(\alpha) = \frac{1}{1 - \alpha} \sum_{i=k-1}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i - 1,$$

$$\widetilde{\psi}_k''(\alpha) = \frac{1}{1 - \alpha} \left(\sum_{i=k-2}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i \right) + \frac{1}{(1 - \alpha)^2} \left(\sum_{i=k-1}^{\infty} \frac{1}{i!} \left(\ln \frac{1}{2 - 2\alpha} \right)^i \right).$$

Inspecting these derivatives, we see that $\widetilde{\psi}_k'(1/2) = -1 < 0$, and $\widetilde{\psi}_k''$ is positive on the interval $(1/2, 1)$. Thus, $\widetilde{\psi}_k$ has at most one root in the interval.

Finally, since the quantity $\ln \frac{1}{2-2\alpha}$ approaches infinity as α approaches 1 from below, it is clear that

$$\lim_{\alpha \to 1^-} \widetilde{\psi}_k(\alpha) = \infty.$$

This implies that $\widetilde{\psi}_k(\alpha) = \widetilde{\varphi}_k(\alpha) - \alpha$ has a unique root $\alpha^{**} \in (1/2, 1)$, and furthermore, since $\widetilde{\psi}_k$ is increasing for $\alpha > \alpha^{**}$, iteration of $\widetilde{\varphi}_k$ on any $\alpha > \alpha^{**}$ will eventually give a result larger than 1. Thus we must have $\widetilde{c}_k \leq \alpha^{**}$. ∎

Figure 3 illustrates the method applied to prove Theorem 5.3, in which one region of the hypercube is ruled out, while Figure 4 illustrates the method of Theorem 6.1, in which successive regions are removed from hypercubes of decreasing size.
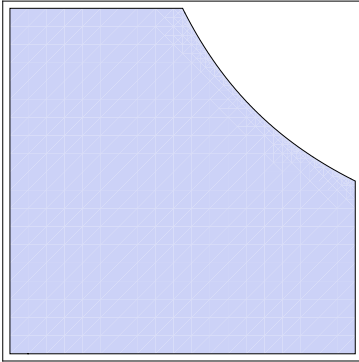

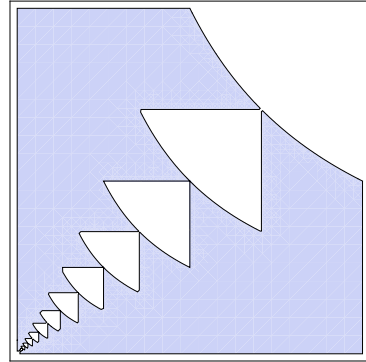Fig. 3. Method of Theorem 5.3

Fig. 4. Method of Theorem 6.1

When we attempt to adapt the proof of Theorem 6.1 to the discrete case (already stated as Theorem 1.4), we encounter several obstacles. First, the side lengths of the smaller cubes are not integers for most choices of α. Second, continuous approximation introduces an additional error term for each iteration. And third, after some number of iterations, we will run out of lattice points. However, all of these problems can be addressed by making n sufficiently large.

Proof of Theorem 1.4. We follow the proof of Theorem 6.1, but for each iteration we must replace $\mathrm{Vol}(\widetilde{T}(n,k,\alpha))$ with $\mathrm{Vol}(\widetilde{T}(n,k,\alpha)) + O(n^{k-1})$, and $(2-2\alpha)^{1/k}$ with $(2-2\alpha)^{1/k} + O(1)$, to adjust for the continuous approximation step.

The consequence is that if a sumfree set $S \subset \{1, \ldots, n\}^k$ has volume αn^k, there exists a smaller cube R of side length $(2-2\alpha)^{1/k} n + O(1)$ such that the density $\mathrm{Vol}(S \cap R)/\mathrm{Vol}(R)$ is at least

$$\varphi_k(\alpha) := \widetilde{\varphi}_k(\alpha) + O(1/n).$$

We similarly define

$$\psi_k(\alpha) := \varphi_k(\alpha) - \alpha = \widetilde{\psi}_k(\alpha) + O(1/n).$$

Now, let $\overline{\alpha} > \alpha^{**}$. Since $\widetilde{\psi}_k(\overline{\alpha})$ is positive, the same is true of $\psi_k(\overline{\alpha})$ if we choose n large enough. It follows from the proof of Theorem 6.1 that some iterate $\varphi_k^m(\overline{\alpha})$ is greater than 1, yielding a contradiction.

We must still ensure that the decreasing side length does not reach zero before we reach our contradiction. However, the number of necessary iterations for a fixed $\overline{\alpha}$ does not change if we increase n, so we are free to choose n large enough to avoid this situation. Thus, any initial density $\overline{\alpha} > \alpha^{**}$ results in a contradiction for sufficiently large n; therefore, $c_k \leq \alpha^{**}$. ∎

7. Generalization to l-fold-sumfree sets. A sumfree set S is, by definition, a set such that $L(x, y, z) = x + y - z \neq 0$ for all $x, y, z \in S$. We can generalize this definition by replacing $x + y - z$ with any other linear form $L(x_1, \ldots, x_n)$ and considering sets such that this form is non-zero for all $x_1, \ldots, x_n \in S$.

As a natural generalization, we call S an *l-fold-sumfree set* if

$$\forall x_1, \ldots, x_l, z \in S, \quad x_1 + \cdots + x_l - z \neq 0,$$

or equivalently, using sumset notation, if

$$lA \cap A = \emptyset.$$

We define

$$(5) \quad c_{k,l} := \limsup_{n \to \infty} \frac{1}{n^k} \max\{\#S : S \subset \{1, \ldots, n\}^k \text{ is } l\text{-fold sumfree}\},$$

or in the continuous setting,

$$(6) \quad \widetilde{c}_{k,l} := \max\{\text{Vol}(S) : S \subset [0, 1]^k \text{ is measurable and } l\text{-fold-sumfree}\}.$$

REMARK 7.1. In some of the literature ([2], for instance), these sets are simply referred to as l-sumfree. However, this description is used with various meanings (see [11]), so we will use the term l-fold-sumfree for added clarity.

As in the sumfree ($l = 2$) case, we can construct large sumfree sets using "diagonal stripes", leading to a similar lower bound.

THEOREM 7.1. *Define $c_{k,l}$ and $\widetilde{c}_{k,l}$ as in* (5) *and* (6). *Then*

$$c_{k,l} \geq 1 - \frac{2}{k!} \sum_{i=0}^{\lfloor k/(l+1) \rfloor} (-1)^i \binom{k}{i} \left(\frac{k}{l+1} - i \right)^k,$$

$$\widetilde{c}_{k,l} \geq 1 - \frac{2}{k!} \sum_{i=0}^{\lfloor k/(l+1) \rfloor} (-1)^i \binom{k}{i} \left(\frac{k}{l+1} - i \right)^k.$$

Proof. We follow the proof of Theorem 1.1, except now we use the *l*-fold-sumfree set

$$\widetilde{S}(n, k, a) := \left\{ (x_1, \ldots, x_k) \in [0, n]^k : \frac{1}{l+1} \leq x_1 + \cdots + x_k < \frac{l}{l+1} \right\}.$$

The given bound is the volume of this set, by Lemma 2.1. ∎

COROLLARY 7.2. *Define $c_{k,l}$ and $\widetilde{c}_{k,l}$ as in* (5) *and* (6). *Then*

$$\lim_{k \to \infty} c_{k,l} = \lim_{k \to \infty} \widetilde{c}_{k,l} = 1.$$

Proof. The lower bound for $c_{k,l}$ and $\widetilde{c}_{k,l}$ given in Theorem 7.1 is larger than the lower bound for c_k given in Theorem 1.1 (as it is the volume of a larger region). Since the previous bound approaches 1 as k grows large, this one does as well. ∎

Our upper bound does not extend as easily. Adapting our methods, we can deal with the $l = 3$ case, and in fact find an upper bound which is both explicit and reasonably effective; however, it is not evident how to deal with any of the cases where $l \geq 4$.

THEOREM 7.3. *Define $c_{k,l}$ and $\widetilde{c}_{k,l}$ as in* (5) *and* (6). *Then*

$$c_{k,3} \leq 1 - \frac{1}{(1 + 2^{1/k})^k}, \quad \widetilde{c}_{k,3} \leq 1 - \frac{1}{(1 + 2^{1/k})^k}.$$

Proof. Let S be an *l*-fold-sumfree subset of $[0, 1]^k$ with $\mathrm{Vol}(S) = \alpha$. Suppose that

$$\alpha > 1 - \frac{1}{2^k}.$$

Let

$$\gamma = (1 - \alpha)^{1/k} + \varepsilon, \quad \text{where } 0 < \varepsilon < \frac{1}{2} - (1 - \alpha)^{1/k}.$$

Note that

$$\alpha > 1 - \gamma^k \quad \text{and} \quad \gamma < \frac{1}{2}.$$

We define two sets

$$A_1 = S \cap [0, \gamma]^k, \quad A_2 = S \cap [1 - \gamma, 1]^k$$

and consider the element

$$a := (1 - \gamma, 1 - \gamma, \ldots, 1 - \gamma).$$

Suppose that the sets A_2 and the translation $A_1 + a \subset [1 - \gamma, 1]^k$ are disjoint. Then

$$\alpha \leq 1 - 2\gamma^k + \mathrm{Vol}(A_1) + \mathrm{Vol}(A_2) = 1 - 2\gamma^k + \mathrm{Vol}(A_1 + a) + \mathrm{Vol}(A_2)$$
$$\leq 1 - 2\gamma^k + \gamma^k = 1 - \gamma^k,$$

which contradicts our assumption on γ.

Thus, there exist elements $w, z \in S$ such that $w + a = z$. This means S cannot contain any pair of elements $x, y \in S$ such that $x + y = a$, or else we would have $w + x + y = z$, a contradiction since S is 3-fold-sumfree. Then, as in the proof of Theorem 5.3, we must have

$$\mathrm{Vol}(S \cap [1 - \gamma]^k) \leq \frac{1}{2}\,(1 - \gamma)^k.$$

Using this result we obtain

$$1 - \gamma^k < \alpha \leq 1 - \frac{1}{2}\,(1 - \gamma)^k,$$

and consequently

$$\frac{1}{2}\,(1-\gamma)^k < \gamma^k, \quad \frac{1}{1 + 2^{1/k}} < \gamma = (1-\alpha)^{1/k} + \varepsilon, \quad \alpha < 1 - \left(\frac{1}{1 + 2^{1/k}} - \varepsilon\right)^k.$$

Letting ε approach zero, we achieve the stated bound for $\widetilde{c}_{k,3}$.

We obtain the upper bound for $c_{k,3}$ using the same sort of integral approximation technique we applied to Theorems 1.1, 1.3, and 1.4. The process is virtually identical, so we omit the details here. ∎

Theorems 7.1 and 7.3 give us lower and upper bounds for $c_{k,3}$; looking at the cases where $2 \leq k \leq 6$, we get the following results:

$$0.7500 \leq c_{2,3} \leq 0.8285, \quad 0.9492 \leq c_{5,3} \leq 0.9782,$$
$$0.8593 \leq c_{3,3} \leq 0.9134, \quad 0.9686 \leq c_{6,3} \leq 0.9891.$$
$$0.9166 \leq c_{4,3} \leq 0.9565,$$

These bounds illustrate that for 3-fold-sumfree sets, the largest "diagonal stripe" sets have close to maximal size.

8. Concluding remarks. All of the large sumfree (and l-fold-sumfree) sets we have constructed are unions of sets of the form K_a as defined in (2) in Section 3. These are certainly the simplest sets to grasp, but there is no guarantee that the largest sumfree sets have this structure.

If we limit ourselves to these K_a-unions, the problem is simplified to choosing an optimal sumfree set $A \subset \{k, k+1, \ldots, kn\}$. (To conserve space in this section, we will use the discrete notation to discuss both the continuous and discrete problems.) Since the sets K_a are not of equal size, this is a different task than finding a large sumfree set A. This suggests a more general combinatorial problem.

QUESTION 8.1. *Given an additive set with a weight assigned to each element, what methods can we use to construct sumfree (resp. l-fold-sumfree) sets that maximize the sum of the weights of the elements?*

On the other hand, if we relax this structural constraint, we know virtually nothing about whether the upper bound on the size increases.

QUESTION 8.2. *Are there optimally large sumfree (resp. l-fold-sumfree) subsets of $\{1, \ldots, n\}^k$ which are not the union of "cross-section" sets?*

Addressing both of these questions would solve the problems we have been studying. Question 8.1 is unlikely to have an answer in full generality, although if the weight distribution is highly structured, as it is in this context, there may be methods of approach.

References

[1]   A. Baltz, P. Hegarty, J. Knape, U. Larsson and T. Schoen, *The structure of maximum subsets of $\{1, \ldots, n\}$ with no solutions to $a + b = kc$*, Electron. J. Combin. 12 (2005), research paper 19, 16 pp.

[2]   J. Bourgain, *Estimates related to sumfree subsets of sets of integers*, Israel J. Math. 97 (1997), 71–92.

[3]   J.-M. Deshouillers, G. A. Freiman, V. Sós and M. Temkin, *On the structure of sum-free sets. II*, Astérisque 258 (1999), 149–161.

[4]   J.-M. Deshouillers and V. F. Lev, *A refined bound for sum-free sets in groups of prime order*, Bull. London Math. Soc. 40 (2008), 863–875.

[5]   W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II*, Wiley, New York, 1966.

[6]   B. Green and I. Z. Ruzsa, *Sum-free sets in abelian groups*, Israel J. Math. 147 (2005), 157–188.

[7]   K. S. Kedlaya, *Product-free subsets of groups*, Amer. Math. Monthly 105 (1998), 900–906.

[8]   S. Lang, *Algebraic Number Theory*, 2nd ed., Grad. Texts in Math. 110, Springer, New York, 1994.

[9]   V. F. Lev, *Large sum-free sets in $\mathbb{Z}/p\mathbb{Z}$*, Israel J. Math. 154 (2006), 221–233.

[10] I. Schur, *Über die Kongruenz $x^m + y^m = z^m \pmod{p}$*, Jahresber. Deutsche Math. Ver. 25 (1916), 114–117.

[11] E. Szemerédi and V. H. Vu, *Long arithmetic progressions in sum-sets and the number of x-sum-free sets*, Proc. London Math. Soc. (3) 90 (2005), 273–296.

Department of Mathematics
Brown University
Box 1917
Providence, RI 02912-1917, U.S.A.
E-mail: thedan@math.brown.edu