Partitioning multi-dimensional sets in a small number of "uniform" parts

Noga Alon* Ilan Newman† Alexander Shen‡ Gábor Tardos§
Nikolai Vereshchagin¶

*Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and Institute for Advanced Study, Princeton, NJ 08540, USA. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by the Hermann Minkowski Minerva Center for Geometry, and by the Von Neumann Fund.

†University of Haifa, Email: ilan@cs.haifa.ac.il.

‡Institute for Information Transmission Problems, Russian Academy of Science; Poncelet laboratory, CNRS. E-mail:shen@landau.ac.ru, shen@mccme.ru. Research supported in part by CNRS and RFBR.

§Rényi Institute, Budapest, Hungary. Partially supported by the Hungarian National Research Grants OTKA-T-037846 and OTKA-T-048826.

¶The corresponding author. Work done in part while visiting LIF, University of Provence. Moscow State University, Leninskie Gory, Moscow 119992. E-mail: ver@mccme.ru. Research supported in part by RFBR grants 03-01-00475, 358.2003.1.

Running head: Partitioning sets

The corresponding author is Nikolai Vereshchagin. Moscow State University, Leninskie Gory, Moscow, Russia 119992. E-mail: ver@mccme.ru

Abstract

Our main result implies the following easily formulated statement. The set of edges E of every finite bipartite graph can be split into poly$(\log |E|)$ subsets so that all the resulting bipartite graphs are almost regular. The latter means that the ratio between the maximal and minimal non-zero degree of the left nodes is bounded by a constant and the same condition holds for the right nodes. Stated differently, every finite 2-dimensional set $S \subset \mathbb{N}^2$ can be partitioned into poly$(\log |S|)$ parts so that in every part the ratio between the maximal size and the minimal size of non-empty horizontal section is bounded by a constant and the same condition holds for vertical sections.

We prove a similar statement for n-dimensional sets for any n and show how it can be used to relate information inequalities for Shannon entropy of random variables to inequalities between sizes of sections and their projections of multi-dimensional finite sets.

Let S be a finite n-dimensional set, that is, a subset of $X_1 \times X_2 \times \cdots \times X_n$ for some $X_1, X_2, \ldots, X_n$. For every set of indices $A \subset \{1, 2, \ldots, n\} = [n]$ consider the projection of S on coordinates in A. Let $m_S(A)$ denote the cardinality of this projection. We will consider also sections of projections of S. Let A and B be disjoint sets of indices. Choose any point s in S and consider the set of A-coordinates of all the points in S having the same B-coordinates as s. Sets of this form are called *$A|B$-sections of S*. Let $\max_S(A|B)$ stand for the largest cardinality of an $A|B$-section and $\min_S(A|B)$ for the smallest of them.

It is natural to define $\max(A|\varnothing) = \min(A|\varnothing) = m(A)$ and $m(\varnothing) = \max(\varnothing|B) = \min(\varnothing|B) = 1$.

For example let $S \subset \mathbb{N}^2$ (Fig. 1). Then $m_S(\{1\})$ is the number of elements in the

Figure 1: A 2-dimensional set and its characteristics

projection of S on the horizontal axis, $m_S(\{2\})$ is the number of elements in the projection on the vertical axis, $\max_S(\{2\}|\{1\})$ is the maximal number of elements in vertical sections, and $\max_S(\{1\}|\{2\})$ is the maximal number of elements in horizontal sections. The total number of elements in S is $m_S(\{1, 2\})$.

We have the following trivial inequality:

$$m(1, 2) \leq m(1) \cdot \max(2|1)$$

(we drop the subscript S and the brackets). Indeed, there are $m(1)$ vertical sections and each of them has at most $\max(2|1)$ elements. For n-dimensional sets and disjoint sets A and B of indices we have a similar inequality:

$$m(A \cup B) \leq m(B) \cdot \max(A|B).$$

Call a set S *uniform* if for all disjoint A, B this inequality specializes to equality, i.e., if all $A|B$-sections have the same cardinality, that is, $\max(A|B) = \min(A|B)$ (for all A, B). Note that it is enough to require the equality

$$m(A \cup B) = m(B) \cdot \max(A|B)$$

to be true only for $B = \bar{A}$ (the complement of A). The simplest example of a uniform set is a "parallelepiped"—a product of n sets S_i. There are other uniform sets, for instance, the 6-element set shown on Fig 2 is uniform: all its vertical and horizontal sections have 2 elements.

Figure 2: A uniform set

Uniform sets were used in [3] to provide a combinatorial interpretation to inequalities for Shannon entropies of random variables, called also *information inequalities*. Let $\xi_1, \ldots, \xi_n$ be random variables with finite domains having a joint distribution. Consider linear inequalities of the form

$$\sum_A \lambda_A H(\xi_A) \leq 0. \tag{1}$$

Here A ranges over non-empty subsets of the set of indices $\{1, \ldots, n\}$ and ξ_A stands for the random variable consisting of all ξ_i for $i \in A$, that is, ξ_A is the A-projection of the vector $\langle \xi_1, \ldots, \xi_n \rangle$. Here are two examples of such inequalities

$$\begin{gathered} H(\xi_1) + H(\xi_2) \geq H(\langle \xi_1, \xi_2 \rangle), \\ H(\xi_1) + H(\langle \xi_1, \xi_2, \xi_3 \rangle) \leq H(\langle \xi_1, \xi_2 \rangle) + H(\langle \xi_1, \xi_3 \rangle). \end{gathered} \tag{2}$$

Both inequalities are true for all ξ_1, ξ_2, ξ_3. These two inequalities correspond to the following two combinatorial inequalities

$$m(1)m(2) \geq m(1,2), \qquad m(1)m(1,2,3) \leq m(1,2)m(1,3)$$

that are true for all uniform sets. Actually the first one is obviously true for all sets. However the second one is false for some sets: consider, for instance, the disjoint union of a parallelepiped $U \times V \times W$ with large U, V, W with another parallelepiped $P \times \{1\} \times \{1\}$ where $|P|$

is much greater than $|U|$ and much less than $|U\times V|$ and $|U\times W|$. The first parallelepiped (as well as the second one) satisfies the equality $m(1)m(1,2,3) = m(1,2)m(1,3)$. However when we join it with the second parallelepiped, all the terms $m(1,2,3), m(1,2), m(1,3)$ increase only a little but the term $m(1)$ increases much; the inequality becomes false.

For uniform sets the second inequality can be proved as follows. Replace the term $m(1,2,3)$ in the left hand side by $m(1)\max(\{2,3\}|1)$ and make similar replacements in the right hand side: $m(1,2) = m(1)\max(2|1)$ and $m(1,3) = m(1)\max(3|1)$. Then the inequality becomes trivial:

$$\max(\{2,3\}|1) \le \max(2|1)\max(3|1)$$

(the size of every 2-dimensional section does not exceed the product of its linear projections).

Most of the known information inequalities are consequences of inequalities of type (2), such inequalities are called Shannon type inequalities. However, there are some exceptions found recently, see [4, 1].

For every inequality for Shannon entropies of random variables $\xi_1,\ldots,\xi_n$ of the form (1) we can consider the corresponding inequality for the size of n-dimensional finite sets and its projections. It is obtained by formal substitution of $\log m(A)$ for $H(\xi_A)$ in the formula (1):

$$\sum_A \lambda_A \log m(A) \le 0 \tag{3}$$

or, in equivalent form,

$$\prod_A m(A)^{\lambda_A} \le 1. \tag{4}$$

In [3] it is shown that if the inequality (1) for Shannon entropy is true for all random variables then the corresponding combinatorial inequality (4) is true for all uniform sets and vice versa.

The goal of our paper is to go further: for every linear inequality for Shannon entropies we provide a combinatorial interpretation that is true for *every* finite set (another interpretation of this kind was presented in [2]). Namely we show that if the inequality (1) is true for all random variables then every finite set can be partitioned into a small number of parts so that every part "almost" satisfies the corresponding inequality (4). The number of parts is bounded by a polynomial of the logarithm of the cardinality of the set and "almost" means that the constant 1 in the right hand side of (4) is replaced by some constant depending only on n, the number of variables:

Theorem 1. *For every n there is a constant d and a polynomial p such that the following holds. Every finite set $S\subset\mathbb{N}^n$ can be partitioned into $p(\log|S|)$ parts so that for every part we have*

$$\prod_A m(A)^{\lambda_A} \le d \tag{5}$$

whenever the parameters λ_A satisfy $\sum_A|\lambda_A|\le 1$ and the inequality (1) is true for all random variables.

The proof of the theorem consists of two parts. First we prove that every set $S \subset \mathbb{N}^n$ can be partitioned into poly($\log |S|$) almost uniform parts and then we prove that every almost uniform set satisfies the inequality (5).

Let us give the definition of an almost uniform set. Fix a constant c and call a set S *c-uniform* if

$$c \cdot m(A \cup B) \geq m(B) \cdot \max(A|B)$$

for all disjoint sets of indices A, B. In other words, the cardinality of the largest $A|B$-section exceeds the average cardinality of $A|B$-sections, that is equal to $m(A \cup B)/m(B)$, by at most a factor of c. Uniform sets are 1-uniform sets. Call a set S *weakly c-uniform* if

$$c \cdot m(\bar{A} \cup A) \geq m(A) \cdot \max(\bar{A}|A)$$

for every set of indices A. One can show that weak 2-uniformity does not imply c-uniformity: consider the 3-dimensional set $\{(0, i, i) : 1 \leq i \leq n\} \cup \{(i, i, j) : 1 \leq i, j \leq n\}$. This set is weakly 2-uniform, but it is not $n/2$-uniform.

Almost uniform sets have the following simple property.

Lemma 1. *If the inequality* (1) *is true for all random variables and* $\sum_A |\lambda_A| \leq 1$ *then for every weakly c-uniform set* $S \subset \mathbb{N}^n$ *we have*

$$\sum_A \lambda_A \log m(A) \leq \log c.$$

Proof. Let $\xi = \langle \xi_1, \ldots, \xi_n \rangle$ be the random variable that is uniformly distributed in S. Then the Shannon entropy of its projection ξ_A on any set of coordinates A is at most $\log m(A)$ and at least $\log m(A) - \log c$.

Indeed, as ξ_A has at most $m(A)$ different outcomes, its entropy does not exceed $\log m(A)$. Every outcome of ξ_A has probability at most $\max(\bar{A}|A)/|S|$. As S is c-uniform, this is less than $c/m(A)$. Hence the entropy of ξ_A is greater than the minus logarithm of this ratio.

By assumption the inequality (1) is true for $\xi_1, \ldots, \xi_n$. Replace each term $H(\xi_A)$ in it by the term $\log m(A)$, which differs from it by at most $\log c$. $\square$

Thus to prove Theorem 1 it is enough to prove the following combinatorial statement.

Theorem 2. *For any n there exist a constant c and a polynomial p such that every finite set* $S \subset \mathbb{N}^n$ *can by partitioned into* $p(\log |S|)$ *c-uniform parts.*

Proof. We associate a *weight* with every partition of S into subsets. We will show that the minimal weight partition (that exists because the set of partitions is finite) satisfies the statement of the theorem.

We first define a weight for every element $s \in S$. Let X be a part from the partition. The weight of every element $s \in X$ is defined by the formula

$$w(s) = -d \log |X| + \sum_{A,B} \log \max_X(B|A), \tag{6}$$

where the sum is taken over all disjoint pairs of indices $A, B \subseteq \{1, 2, \dots, n\}$ and d is a constant (depending on n, to be chosen later). Note that the sum includes the terms $\log m(B)$ for all $B \subseteq \{1, 2, \dots, n\}$ as we can let $A = \varnothing$. Let us stress that the weights of all elements in the same part coincide. We then define the weight of a partition as the sum of the weights of all the elements.

The intuition behind the weight function is as follows. The term $\sum_X (-d|X| \log |X|)$ (the sum of $-d \log |X|$ over all $s \in S$) in the formula for the weight of S handles the number of parts: it increases when a part is split in 2 parts and decreases if parts are glued together. Moreover, if the cardinalities of the glued parts are similar, this decrease is large. For instance, gluing together 2 parts of the same cardinality k decreases the sum by $2dk \log 2k - (dk \log k + dk \log k) = 2dk$. The term $\sum_{A,B} \log \max_X(B|A)$ in the formula for $w(s)$ ensures almost uniformity: every part X that is highly non-uniform can be split in parts X_0, X_1 so that this term decreases a lot for all $s \in X$. Indeed, assume that the maximal $B|A$-section of X is much larger than the average one. Let then X_0 consist of all large sections of X and X_1 consist of all the remaining sections. Then $m_{X_0}(A)$ is much smaller than $m_X(A)$ (there are few sections whose cardinality is much larger than the average one). On the other hand, $\max_{X_1}(B|A)$ is much smaller than $\max_X(B|A)$ (all large sections are in X_0). Later we will make this arguments precise.

Let us prove first that if d is sufficiently large then the number of parts in every partition of the smallest weight is small. Namely, we will prove that if we glue together any two parts for which the terms $\log \max(B|A)$ are close enough (differ by at most 1) for every B, A, then the weight of the partition decreases. Indeed, let X, Y be two distinct parts for which $\log \max_X(B|A)$ differs from $\log \max_Y(B|A)$ by at most 1, for every disjoint B, A. This assumption on X, Y implies that $|X \cup Y| \geq 1.5 \max(|X|, |Y|)$ by choosing $B = [n], A = \varnothing$. Similarly, $\max_{X\cup Y}(B|A) \leq 3 \max_X(B|A)$ and $\max_{X\cup Y}(B|A) \leq 3 \max_Y(B|A)$.

Thus fixing B, A and summing up the contribution of all elements in $X \cup Y$ to the term $\log \max_{X\cup Y}(B|A)$ gives:

$$|X \cup Y| \cdot \log \max_{X\cup Y}(B|A) \;\leq\; |X| \cdot \log \max_X(B|A) + |Y| \cdot \log \max_Y(B|A) + |X \cup Y| \cdot \log 3.$$

(Recall that all the logarithms are binary.) Hence, there is an increase of at most $|X \cup Y| \cdot \log 3$ compared to the contribution of the corresponding term before gluing. On the other hand the term $d \log |X \cup Y|$ contributes (summing up for all elements in $X \cup Y$) at least $d(|X \cup Y|) \log(|X \cup Y|)$. Plugging in that $|X \cup Y| \geq 1.5 \max(|X|, |Y|)$ we get that this is at least $d|X| \log |X| + d|Y| \log |Y| + d|X \cup Y| \log(1.5)$. Thus if we choose $d \geq 3^n \log 3 / \log(1.5)$ (to compensate the increase for all $A, B \subseteq [n]$), the value of the partition will certainly decrease.

Let d be chosen as described. Let us classify the parts in the partition according to the integer parts of $\log \max(B|A)$ for all A and B. As shown above, no two parts fall into the same class. Thus the number of parts is bounded by a polynomial of the logarithm of the cardinality of the partitioned set (recall that n is fixed). This implies the upper bound on the number of parts for a minimal weight partition.

It remains to show that in every partition of the smallest weight all the parts are almost uniform. We will show that every part that is considerably non-uniform can be split into two parts so that the weight of the partition decreases. Splitting a part does not affect weights of points in other parts so we may consider only the change of the weight in the split part. As the result of such a splitting, all the logarithms in the equation (6) decrease. We need to split the part in such a way that the total decrease of the sum $\sum_{A,B} \log m_X(B|A)$ is greater than the total decrease of the term $d \log |X|$. The decrease of the term $d \log |X|$ can be expressed by a simple formula: if X is split in two parts of cardinalities $p|X|$ and $q|X|$, respectively (thus $p + q = 1$), then the total decrease of $d \log |X|$ is equal to $d \cdot |X| \cdot H(p, q)$, where

$$H(p, q) = p(-\log p) + q(-\log q) \leq 1$$

is the binary entropy function. Therefore the average decrease per element of the term $d \log |X|$ is at most d. Hence it suffices to find a splitting such that the last term in the weight of every point in X decreases by more than d.

Assume that X is not c-uniform, that is for some disjoint sets of indices A and B we have $\max_X(B|A) \geq c \cdot d_X(B|A)$ where $d_X(B|A)$ is the average size of the $(B|A)$ sections. We split X into two parts. The first part contains all small $(B|A)$ sections and the second contains all the remaining ones. As the threshold take the geometric mean of the size of the maximal section and the size of the average section. In the first part, the size of the maximal section (compared to X) decreases by a factor of at least $\sqrt{c}$. In the other part, all the sections exceed $\sqrt{c}$ times the average section of X, hence the number of sections in the second part is $\sqrt{c}$ times smaller than that in X. That is, the size of the A-projection of the second part is $\sqrt{c}$ times smaller than that of X.

As the result of the splitting, in both parts at least one term of the sum in equation (6) decreases by $\log \sqrt{c}$ (and all the other do not increase). Therefore if c is large, that is, $\log \sqrt{c} > d$, the decrease in the contribution of the last term in (6) dominates the contribution of the $d \log |X|$ term and hence the total weight of elements of X decreases. This means that in every partition of the minimal weight all parts are c-uniform for a constant c depending only on n. □

Theorem 2 can be strengthened by requiring that in all the parts the ratio between the largest section and the smallest section is bounded by a constant. We do not need this for Theorem 1. However we think this is interesting in its own right.

Call a set *strongly d-uniform* if for every disjoint sets A, B of indices

$$\max(A|B)/\min(A|B) \leq d.$$

Every uniform set is strongly 1-uniform.

Theorem 3. *For any n there exist a constant d and a polynomial $p(\cdot)$ such that every finite set S can by partitioned into $p(\log |S|)$ strongly d-uniform parts.*

Proof. First let us note that it is enough to prove that for some polynomial q and a constant d every set S has a strongly d-uniform subset T of size at least $|S|/q(\log |S|)$. Indeed, remove a large strongly d-uniform part T from S. We obtain a set $S' \subset S$ of cardinality at most

$|S|(1 - 1/q(\log|S|))$. Then remove from S' another d-uniform subset T' getting a set S'' of cardinality at most $|S|(1 - 1/q(\log|S|))^2$. Repeating this $O(q(\log|S|)\log|S|)$ times we get the empty set and obtain the partition satisfying the theorem.

To find a large strongly d-uniform subset T of a given set S apply Theorem 2 to S and take the largest part T in the partition provided by the theorem (i.e., the part with the largest cardinality). That part T is c-uniform and has at least $1/p(\log|S|)$ fraction of the elements of S. However for some A, B it may have $B|A$-sections that are much smaller (say, d times smaller) than the largest $B|A$-section. If this is the case, pick any small section and remove it from T, that is, remove all the elements of T whose projection on coordinates in $A \cup B$ belongs to that section. Repeat such removals in any order until either T becomes empty or strongly d-uniform. We claim that if the constant d is chosen appropriately then T cannot become empty and moreover it loses at most half of its elements. Indeed, fix a pair of set indices A, B and count the total number of elements removed due to $B|A$-sections. After removing any small $B|A$-section the set T loses at most

$$\frac{\max_T(B|A)\max_T(C|A \cup B)}{d}$$

elements, where C stands for the complement of $A \cup B$. This bound does not increase as T is shrinking, and the total number of $B|A$-sections in T is equal to $m_T(A)$, so the total number of removed elements is bounded by

$$\frac{m_T(A)\max_T(B|A)\max_T(C|A \cup B)}{d},$$

As T is c-uniform, the product of the first two terms in the numerator is at most $c \cdot m_T(A \cup B)$, and again by c-uniformity, the product of all three terms does not exceeds $c^2 \cdot |T|$. Hence we can let d be equal to the number of pairs (A, B) times $2c^2$. □

We conclude by a simple observation that the converse to Theorem 1 is true, even in a stronger form.

Theorem 4. *Assume that an integer n and coefficients λ_A (for all $A \subseteq \{1, \dots, n\}$) are fixed. Assume that every finite set $S \subseteq \mathbb{N}^n$ can be partitioned into $O(|S|^{o(1)})$ parts so that*

$$\sum_A \lambda_A \log m_T(A) = o(\log|S|).$$

for every part T. Then the inequality (1) holds for all random variables $\xi_1, \dots, \xi_n$.

Proof. We will use a result of [3]: for every tuple $\xi_1, \dots, \xi_n$ of jointly distributed random variables and for every natural N there is an uniform set $S \subseteq \mathbb{N}^n$ such that $\log m_S(A) = N \cdot H(\xi_A) + o(N)$ for every set A of indices (recall that n is fixed and N tends to infinity). In particular, $\log|S| = N \cdot H(\xi) + o(N) = O(N)$. Choose a large N and let S be the uniform set as above. By the assumption the set S can be split into $c = |S|^{o(1)} = 2^{o(N)}$ parts so that

$$\sum_A \lambda_A \log m_T(A) = o(\log|S|) = o(N). \tag{7}$$

for every part T.

Pick the largest part T in the partition. We claim that $\log m_T(A)$ is close to $\log m_S(A)$ and hence close to $N \cdot H(\xi_A)$ for all A. More specifically,

$$m_S(A)/c \le m_T(A) \le m_S(A).$$

The second inequality is obvious, as T is a subset of S. To prove the first one let B be the complement of A. Compare the inequality (which is always true), $m_T(A)\max_T(B|A) \ge |T|$, with the equality $m_S(A)\max_S(B|A) = |S|$, which is true since S is uniform. Using $|T| \ge |S|/c$ and $\max_T(B|A) \le \max_S(B|A)$ the bound follows.

Thus if we replace $\log m_T(A)$ in the inequality (7) by $\log m_S(A)$ the left hand side can increase by at most $O(\log c) = o(N)$. Replacing $\log m_S(A)$ by $N \cdot H(\xi_A)$ changes it also by at most $o(N)$. Thus we obtain the inequality

$$\sum_A \lambda_A \cdot N \cdot H(\xi_A) \le o(N).$$

Divide it by N and take the limit. □

It is interesting to estimate the minimal degree of a polynomial p in Theorem 2. We can find good estimates for its degree in the case of *weakly* c-uniform sets (note that these sets are enough for the proof of Theorem 1).

Theorem 5. *Let us fix n and let $k = 2^n - 2$. There exists a $c > 0$ such that every finite n-dimensional set S has a weakly c-uniform subset of cardinality at least $\frac{|S|}{(\log|S|)^k}$. On the other hand for all m and c there exists a n-dimensional set S of cardinality $\Omega(m^k 2^m)$ such that all its weakly c-uniform subsets have cardinality at most $O(2^m(\log m + \log c)^k)$. The constants in the O- and Ω-notations depend on n.*

This implies that the minimal degree of a polynomial p such that for some c every n-dimensional set S can be partitioned into $p(\log|S|)$ weakly c-uniform subsets is in the range $[2^n - 2; 2^n - 1]$.

Proof. We start with the upper bound. To this end we prove the following

Lemma 2. *For $k = 2^n - 2$ every weakly α-uniform n-dimensional set S has a weakly $2^k\sqrt{\alpha}$-uniform subset of cardinality at least $|S|/2^k$.*

Proof. Consider a subset A of $\{1,\dots,n\}$ and let d denote the average cardinality of $\bar{A}|A$-sections of S. Let $S_0 \subseteq S$ contain all $\bar{A}|A$-sections of cardinality $d\sqrt{\alpha/2}$ or less. Let S_1 contain all the remaining points. Let T denote the largest set among S_0, S_1. We claim that T is $\sqrt{2\alpha}$-uniform with respect to $\bar{A}|A$-sections. Indeed, if $T = S_0$ then the cardinality of all $\bar{A}|A$-sections in T is at most $d\sqrt{\alpha/2}$, and the average size of a $\bar{A}|A$-section is at least $d/2$ (removing points does not increase the A-projection of S, and the cardinality of T is at least $|S|/2$). If $T = S_1$ then the cardinality of $\bar{A}|A$-sections in T is at least $d\sqrt{\alpha/2}$ and at most

$d\alpha$ hence T is $\sqrt{2\alpha}$-uniform with respect to $\bar{A}|A$-sections. In both cases ($T = T_1$ or $T = T_0$) T is 2α-uniform with respect to $\bar{B}|B$-sections for $B \neq A$, as section size cannot increase.

Apply this procedure to all non-empty proper subsets A of $\{1, \dots, n\}$. Each step decreases the cardinality of the set by a factor of at most 2, thus the cardinality of the resulting set is at least $|S|/2^k$. For each A the ratio between the cardinalities of the maximum and the average ($\bar{A}|A$) sections has been multiplied by a factor of 2 for at most $k-1$ times, while at least once it has been decreased from some r to $\sqrt{2r}$. Hence the resulting set is weakly $2^k\sqrt{\alpha}$-uniform. □

To end the proof of the upper bound in Theorem 5, apply the lemma $N = \lfloor \log\log |S| \rfloor$ times to the given set S. As S is certainly $|S|$-uniform, we obtain that it contains a subset of cardinality at least

$$\frac{|S|}{2^{kN}} \geq \frac{|S|}{(\log |S|)^k}$$

that is weakly c-uniform with

$$c = (2^k)^{1+1/2+\cdots+2^{1-N}} \cdot |S|^{2^{-N}} < 2^{2k+2}.$$

It remains to prove the lower bound. To this end we first establish the following

Lemma 3. *For $k = 2^n - 2$ and for all m there is a family of $\Omega(m^k)$ uniform n-dimensional sets of cardinality 2^m each with the following properties. For every set S in the family and for every set $A \subset [n]$, the cardinality of $\bar{A}|A$-sections of S is equal to 2^i for some natural i. In addition, for every different S_1, S_2 in the family there is $A \subseteq [n]$ such that the cardinality of $\bar{A}|A$-sections of S_1 differs from the cardinality of $\bar{A}|A$-sections of S_2.*

We first finish the proof of Theorem 5 using this lemma. Let $\mathcal{F}$ be the family provided by the Lemma. We may assume without loss of generality that $|\mathcal{F}| \leq m^k$ and the sets $S \in \mathcal{F}$ are pairwise disjoint. We claim that the union $\cup\mathcal{F}$ satisfies the statement of the theorem.

Indeed, let T be a weakly c-uniform subset of this union. We need to prove that T has at most $O(2^m(\log m + \log c)^k)$ points. Let $d_T(\bar{A}|A)$ be the average cardinality of an $\bar{A}|A$-section of T.

Every point in T belongs to some set $S \in \mathcal{F}$. Divide all points in T into three groups.

(1) Let T_1 consist of those points in T that belong to some $S \in \mathcal{F}$ such that

$$\max_S(\bar{A}_S|A_S) \leq \frac{d_T(\bar{A}_S|A_S)}{2 \cdot m^k}$$

for some set $A_S \subseteq [n]$. We claim that $|T_1| \leq |T|/2$. Indeed,

$$|T_1| \leq \sum_S m_{T_1}(A_S) \cdot \max_S(\bar{A}_S|A_S) \leq \sum_S m_T(A_S) \cdot \frac{d_T(\bar{A}_S|A_S)}{2 \cdot m^k} = |T|/2.$$

(2) Let T_2 consist of those points in T that are in some $S \in \mathcal{F}$ such that there exists a set A_S of indices such that

$$\max_S(\bar{A}_S|A_S) \geq c \cdot m^k \cdot d_T(\bar{A}_S|A_S).$$

For every such S

$$|T_2 \cap S| \le m_S(A_S) \cdot \max_{T\cap S}(\bar{A}_S|A_S) \le m_S(A_S) \cdot c \cdot d_T(\bar{A}_S|A_S);$$

the last inequality holds since T is weakly c-uniform.

Plugging the bound on $d_T(\bar{A}_S|A_S)$ implied by the definition of T_2, we get

$$|S \cap T_2| \le m_S(A_S) \cdot \max_S(\bar{A}_S|A_S)/m^k \le 2^m/m^k.$$

Summing over all S we get that $|T_2| \le 2^m$.

(3) The remaining points $T_3 = T \setminus (T_1 \cup T_2)$. These points belong to sets $S \in \mathcal{F}$ such that for all A the cardinality of the A-section of S is in the range

$$d_T(\bar{A}|A)/(2m^k) \le \max_S(\bar{A}|A) \le c \cdot m^k \cdot d_T(\bar{A}|A),$$

and hence may take at most $\log(4c\cdot m^{2k})$ different values. Thus the vector $(\max_S(\bar{A}|A)\colon A \subset [n])$ may take at most $\log^k(4c \cdot m^{2k})$ different values for any S as above. However, by assumption on the family S, no two S's can share the same vector. We conclude that the number of such S is at most $\log^k(4c \cdot m^{2k})$. As every such S has 2^m points we obtain $|T_3| \le 2^m \log^k(4c \cdot m^{2k})$.

Therefore we have

$$|T| = |T_1| + |T_2| + |T_3| \le |T|/2 + 2^m + 2^m \log^k(4c \cdot m^{2k}),$$

that is,

$$|T| \le 2^{m+1} + 2^{m+1} \log^k(4c \cdot m^{2k}).$$

This proves Theorem 5. □

It remains to prove the lemma.

Proof of Lemma 3. Consider a function J that maps every i in $\{1,\dots,n\}$ to a subset of $[m] = \{1,\dots,m\}$ such that the union of all $J(i)$ covers $[m]$. Associate with J the following n-dimensional set S: For every binary string x of length m include the point $\langle x_1,\dots,x_n\rangle$ in S, where x_i is a substring of x formed by bits of x whose indices are in $J(i)$. The condition on the union of all $J(i)$ guarantees that different strings x become different points in S, thus $|S| = 2^m$.

Every $(\bar{A}|A)$-section of S consists of points associated with strings x having the same projection onto the coordinates in the set $J(A) = \bigcup_{i\in A} J(i)$. The number of such strings is $2^{m-|J(A)|}$. As it depends only on A, the set S is uniform.

We need also that for different sets S_1, S_2 in the family there exists A such that the cardinality of $\bar{A}|A$-sections of S_1 differs from the cardinality of $\bar{A}|A$-sections of S_2.

This means that the functions J used in the construction should be "essentially different", i.e., the mappings $A \mapsto |J(A)|$ should be different. Let us see how many sets S we can obtain in this way, i.e., how many essentially different mappings $J : [n] \to \mathcal{P}([m])$ exist.

To this end, given a function J, consider the 'atoms' of the set family $\{J(i) : i \in [n]\}$, namely all sets of type

$$K(I) = \bigcap_{i \in I} J(i) \cap \bigcap_{i \notin I} \overline{J(i)},$$

where I is a non-empty subset of $\{1, \ldots, n\}$.

It is clear that the sizes of the atoms determine uniquely the sizes of all sets $J(A)$ for all $A \subseteq [n]$. Also, the converse is true due to the inclusion-exclusion formula.

Therefore the number of essentially different mappings J equals the number of decomposition of m into a sum of $2^n - 1$ non-negative integers, that is, $\binom{m+2^n-2}{2^n-2} = \binom{m+k}{k}$, a polynomial in m of degree k, and therefore $\Omega(m^k)$ as claimed. □

Questions.

1. Is it true that every 2-dimensional finite set S can be partitioned into poly$(\log |S|)$ *uniform* (=1-uniform) parts? Is this true for higher dimensions?
2. Theorem 5 asserts that for any n there is some $c > 1$ for which there is a big weakly c-uniform subset in every n-dimensional set. How big such set can be found for (weakly) $(1+\varepsilon)$-uniform subsets, for ε tending to 0 (or even below 1)? We can obtain some good estimates for the 2-dimensional case, but the general case seems to be more difficult.

References

[1] K. Makarychev, Yu. Makarychev, A. Romashchenko, N. Vereshchagin. A new class of non Shannon type inequalities for entropies, *Communications in Information and Systems*, 2:2 (2002), 147–166.

[2] A. Romashchenko, A. Shen, N. Vereshchagin. Combinatorial interpretation of Kolmogorov complexity, *Theoretical Computer Science*, 271 (2002), 111–123.

[3] T.H. Chan. A combinatorial approach to information inequalities, *Communications in Information and Systems*, 1:3 (2001), 241–254.

[4] R. W. Yeung, Z. Zhang, On characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory*, 44 (July 1998), 1440–1450.