Inequalities for entropies and dimensions

Alexander Shen

▶ To cite this version:

Alexander Shen. Inequalities for entropies and dimensions. Unity of Logic and Computation. CiE 2023, Computability in Europe association, Jul 2023, Batumi, Georgia. pp.120-131, 10.1007/978-3-031-36978-0_10 . lirmm-03843708

HAL Id: lirmm-03843708
https://hal-lirmm.ccsd.cnrs.fr/lirmm-03843708

Submitted on 8 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:2209.07243v1 [cs.IT] 15 Sep 2022

Inequalities for entropies and dimensions

Alexander Shen*

Abstract

We show that linear inequalities for entropies have a natural geometric interpretation in terms of Hausdorff and packing dimensions, using the point-to-set principle and known results about inequalities for complexities, entropies and the sizes of subgroups.

1 Introduction

Inequalities for entropies

Let $\xi_1, \ldots, \xi_m$ be jointly distributed random variables with finite ranges. Then, for every nonempty $I \subset \{1, \ldots, m\}$, we may consider the tuple of variables

$$\xi_I = \langle \xi_i \mid i \in I \rangle,$$

and its *Shannon entropy* $H(\xi_I)$. Recall that the Shannon entropy of a random variable that takes s values with probabilities $p_1, \ldots, p_s$ is defined as $\sum_i p_i \log(1/p_i)$. In this way we get $2^m - 1$ real numbers (for $2^m - 1$ non-empty subsets of $\{1, \ldots, m\}$). Shannon pointed out some inequalities that are always true for those quantities. For example, for every two variables ξ_1, ξ_2 we have

$$H(\xi_1) \leqslant H(\xi_1, \xi_2) \leqslant H(\xi_1) + H(\xi_1),$$

and for every triple of variables ξ_1, ξ_2, ξ_3 we have

$$H(\xi_1) + H(\xi_1, \xi_2, \xi_3) \leqslant H(\xi_1, \xi_2) + H(\xi_1, \xi_3).$$

The latter inequality corresponds to the inequality

$$H(\xi_2, \xi_3 | \xi_1) \leqslant H(\xi_2 | \xi_1) + H(\xi_3 | \xi_1)$$

for conditional entropies (defined as $H(\xi | \eta) = H(\xi, \eta) - H(\eta)$).

For a long time no other valid inequalities for entropies (except positive linear combinations of Shannon's inequalities) were known. Then Zhang and Yeung [7] found an inequality that was not a positive linear combination of Shannon's

*LIRMM, University of Montpellier, CNRS, Montpellier, France, alexander.shen@lirmm.fr, sasha.shen@gmail.com.

inequalities, and since then a lot of other inequalities were found. It became clear that the set of all valid linear inequalities for entropies has a complex structure (see, e.g., [5]). On the other hand, it became also clear that this set is very fundamental since it can be equivalently defined in combinatorial terms, in terms of subgroups size, and in terms of Kolmogorov complexity. In this paper we use these characterizations to get one more equivalent characterization of this set, now in terms of Hausdorff and packing dimensions. Let us recall briefly the characterizations in terms of Kolmogorov complexity and subgroup sizes; more details and proofs can be found in [6, Chapter 10].

Inequalities for complexities and subgroup sizes

For a binary string x, its *Kolmogorov complexity* is defined as the minimal length of a program that produces this string. The definition depends on the choice of programming language, but (as Kolmogorov and Solomonoff noted) there exist *optimal* programming languages that make the complexity function minimal, and we fix one of them. In this way we define the function $\mathrm{C}(x)$ up to a bounded additive term, so to get a meaningful statement we should consider the asymptotic behavior of complexity when the length of the strings goes to infinity. (For more details and proofs about Kolmogorov complexity see, e.g., [6].)

Let $x_1, \ldots, x_m$ be binary strings. A tuple $\langle x_1, \ldots, x_m \rangle$ of strings can be computably encoded as one string whose complexity is (by definition) the complexity of the tuple $\langle x_1, \ldots, x_m \rangle$ and is denoted by $\mathrm{C}(x_1, \ldots, x_m)$. The change of the encoding changes the complexity of a tuple at most by $O(1)$, so this notion is well defined. As it was noted by Kolmogorov [3], some inequalities for Shannon entropies have Kolmogorov complexity counterparts. For example,

$$\mathrm{C}(x_1) + \mathrm{C}(x_1, x_2, x_3) \leqslant \mathrm{C}(x_1, x_2) + \mathrm{C}(x_1, x_3) + O(\log n)$$

for all strings x_1, x_2, x_3 of length at most n; this statement corresponds to the inequality for entropies mentioned earlier. This statement is asymptotic; since $\mathrm{C}(x)$ is defined up to $O(1)$ precision, some error term is unavoidable. We use $O(\log n)$ term instead of $O(1)$, so the difference between different popular versions of Kolmogorov complexity, like plain and prefix complexity, does not matter for us.

It is easy to see that every linear inequality for complexities that is true for complexities with logarithmic precision, is also true for Shannon entropies: we replace strings by random variables, Kolmogorov complexity by entropy and omit the error term. It was proven by Romashchenko [2] that the reverse is also true and the same linear inequalities are valid for entropies and complexities (the exact statement and the proof can be found in [6, Chapter 10]). This result remains valid if we allow the programs in the definition of complexity access some oracle (set of strings) X (this is called *relativization* in computability theory); in particular, the class of valid linear inequalities for complexities with oracle X does not depend on X (since it coincides with the valid inequalities for entropies).

Another characterization of the same class of inequalities (in terms of sizes of subgroups of some group and their intersections) was given by Zhang and Yeung [7]. They have shown that it is enough to consider some special type of random variables that correspond to a group and its subgroups: if a linear inequality is valid for random variables of this type, it is valid for all random variables. (See below Section 4 for more details about this class.)

Dimensions and point-to-set principle

The notions of a Hausdorff dimension and packing dimension for sets in $\mathbb{R}^m$ are well known in the geometric measure theory, but for our purposes it is convenient to use their equivalent definitions provided by the point-to-set principle formulated by Jack Lutz and Neil Lutz [4]; the references to the classical definitions (and to previous results about effective dimensions) can be found in this paper.

Let $\alpha = 0.a_1a_2a_3\ldots$ be a real number represented as a binary fraction (the integer part does not matter, and we assume that $\alpha \in [0,1]$). Consider the Kolmogorov complexity of first n bits of α and then consider the limits

$$\mathit{dim}_H(\alpha) = \liminf_{n\to\infty} \frac{\mathrm{C}(a_1\ldots a_n)}{n} \quad \text{and} \quad \mathit{dim}_p(\alpha) = \limsup_{n\to\infty} \frac{\mathrm{C}(a_1\ldots a_n)}{n}$$

called *effective Hausdorff dimension* and *effective packing dimension* of α. Note that the classical dimension of every point is zero, so these notions do not have classical counterparts. Both dimensions of every real α are between 0 and 1.

Now we switch from points to sets and for every set $A \subset [0,1]$ we define the *effective Hausdorff* and *effective packing* dimension of A as the supremum of corresponding dimensions of the points in A:

$$\mathit{dim}_H(A) = \sup_{\alpha\in A} \mathit{dim}_H(\alpha); \qquad \text{and} \qquad \mathit{dim}_p(A) = \sup_{\alpha\in A} \mathit{dim}_p(\alpha).$$

This definition can also be relativized by some oracle X; for that we replace the Kolmogorov complexity C by its relativized version C^X. Adding oracle can make Kolmogorov complexity and effective dimension smaller; we denote the relativized effective dimensions by $\mathit{dim}_H^X(A)$ and $\mathit{dim}_p^X(A)$. The *point-to-set principle* says that for every set A there exists an oracle X that makes the effective dimensions minimal and these minimal dimensions are classical Hausdorff and packing dimensions:

$$\dim_H(A) = \min_X \mathit{dim}_H^X(A) \qquad \text{and} \qquad \dim_p(A) = \min_X \mathit{dim}_p^X(A)$$

Note that we use "*dim*" (italic) for effective dimensions and "dim" for classical dimensions to distinguish between them (for sets; for individual points only effective dimensions make sense). We take this characterization as (equivalent) definitions of Hausdorff and packing dimensions.

Formally speaking, we should first prove that the minimal values are achieved for some oracle X; this is easy to see, because a countable sequence of oracles that give better and better approximations can be combined in one oracle. (Or we can just use "inf" instead of "min" in the definition.)

In the same way the dimensions of a set $A \subset \mathbb{R}^m$ are defined. Now a point $\alpha \in A$ has m coordinates and is represented by a tuple $\langle \alpha_1, \ldots, \alpha_m \rangle$ of binary fractions. To define the effective dimension of α we now consider the Kolmogorov complexity of an m-tuple that consists of n-bit prefixes of $\alpha_1, \ldots, \alpha_m$, and divide this complexity by n. Taking the limits, we get the effective Hausdorff and packing dimension of the point α; they are between 0 and m. Now, taking supremum over all α in A, we define the effective dimensions of $A \subset \mathbb{R}^m$, and their relativized versions $dim_H^X(A)$ and $dim_p^X(A)$. Taking the minimum over X, we get classical Hausdorff and packing dimensions for the set A.

In the next section we give two examples where the inequalities for Kolmogorov complexities are used to prove some results about (classical) Hausdorff and packing dimensions. Then (Section 3) we generalize this approach to arbitrary inequalities for Kolmogorov complexity. Finally (Section 4) we prove the reverse statement that shows that this translation can be used to characterize exactly the class of all linear inequalities valid for entropies or complexities.

2 Inequalities for dimensions: examples

Example 1

Consider the inequality for Kolmogorov complexities

$$2\,\mathrm{C}(x,y,z) \leqslant \mathrm{C}(x,y) + \mathrm{C}(x,z) + \mathrm{C}(y,z) + O(\log n) \tag{1}$$

that is true for all strings x, y, z of length n (see, e.g., [6, Section 2.3]). We apply it to the first n bits of three real numbers α, β, γ (considered as binary sequences; as we have said, we ignore the integer part and assume that all the real numbers are between 0 and 1).

$$2\,\mathrm{C}((\alpha)_n, (\beta)_n, (\gamma)_n) \leqslant \mathrm{C}((\alpha)_n, (\beta)_n) + \mathrm{C}((\alpha)_n, (\gamma)_n) + \mathrm{C}((\beta)_n, (\gamma)_n) + O(\log n).$$

Here we denote by $(\rho)_n$ the first n bits of a real number ρ considered as a binary sequence. We can divide this inequality by n and take lim sup of both parts: recall that $\limsup(x_n + y_n) \leqslant \limsup x_n + \limsup y_n$. In this way we get the inequality

$$2\,dim_p(\alpha,\beta,\gamma) \leqslant dim_p(\alpha,\beta) + dim_p(\alpha,\gamma) + dim_p(\beta,\gamma). \tag{2}$$

This inequality remains can be relativized with arbitrary X used as an oracle in the definition of Kolmogorov complexity and effective packing dimension.

Now instead of one point $\langle \alpha, \beta, \gamma \rangle$ consider a set $S \in \mathbb{R}^3$. Consider also three its two-dimensional projections onto each of three coordinate planes; we denote them by S_{12}, S_{13} and S_{23}. As we have seen in the previous section, the point-to-set principle says that the (classical) packing dimension of a set can be characterized as the minimal (over all oracles) *effective* packing dimension of the set relativized to the oracle, and the latter is the supremum (over all points in the set) of effective packing dimensions of its points. Fix an oracle X that makes the effective packing dimension of all three projections minimal (we may combine the oracles for all three sets) and use it everywhere. Then for every point $s = \langle s_1, s_2, s_3\rangle$ in S the effective packing dimension of $\langle s_1, s_2\rangle$ does not exceed the packing dimension of S_{12}, etc. Applying the inequality (2), we conclude that effective packing dimension of every point $(s_1, s_2, s_3) \in S$ satisfies the inequality

$$2\, \mathit{dim}_p^X(s_1, s_2, s_3) \leqslant \dim_p S_{12} + \dim_p S_{13} + \dim_p S_{23}.$$

where the effective dimension in the left hand side is taken with the fixed oracle (and classical dimensions in the right hand side do not depend on any oracles). Now we take the maximum of all points $\langle s_1, s_2, s_3\rangle \in S$ and get the bound for effective packing dimension of S (with oracle X), and, therefore, for the classical packing dimension of S. In this way we get the following result that deals exclusively with classical dimensions:

Proposition 1. *For every set $S \subset \mathbb{R}^3$ and three its two-dimensional projections S_{12}, S_{13} and S_{23} we have*

$$2 \dim_p S \leqslant \dim_p S_{12} + \dim_p S_{13} + \dim_p S_{23}.$$

Example 2

Now consider another inequality for Kolmogorov complexities of three strings mentioned earlier:

$$\mathrm{C}(x) + \mathrm{C}(x, y, z) \leqslant \mathrm{C}(x, y) + \mathrm{C}(x, z) + O(\log n). \tag{3}$$

We can try the same reasoning, but some changes are necessary. First, we use that

$$\liminf x_n + \liminf y_n \leqslant \liminf(x_n + y_n)$$

(and that $\liminf \leqslant \limsup$) to get an inequality that combines the effective Hausdorff and effective packing dimensions:

$$\mathit{dim}_H^X(s_1) + \mathit{dim}_H^X(s_1, s_2, s_3) \leqslant \mathit{dim}_p^X(s_1, s_2) + \mathit{dim}_p^X(s_1, s_3), \tag{4}$$

for every oracle X. Then, for some strong enough oracle X, we get

$$\mathit{dim}_H^X(s_1) + \mathit{dim}_H^X(s_1, s_2, s_3) \leqslant \dim_p S_{12} + \dim_p S_{13}.$$

Still we cannot make any conclusions about the classical dimensions of the projection S_1 and the entire set S, since the point s_1 where the first term is maximal is unrelated to the point $s = \langle s_1, s_2, s_3\rangle$ when the second term is maximal.

To see what we can do, let us recall that a similar problem appears for the combinatorial interpretation of inequalities for Kolmogorov complexity (see [6, Chapter 10] for details). The inequality (1) from our previous example has a direct combinatorial translation (a special case of the Loomis–Whitney inequality): if a three-dimensional body has volume[1] V and its three projections have areas V_{12}, V_{13} and V_{23}, then

$$2\log V \leqslant \log V_{12} + \log V_{13} + \log V_{23} \qquad (\text{or } V^2 \leqslant V_{12}V_{13}V_{23}).$$

However, the inequality

$$\log V_1 + \log V \leqslant \log V_{12} + \log V_{13},$$

written in a similar way for the inequality (3) , does not hold for every three-dimensional body. Consider, for example, the union of a cube $N \times N \times N$ with a parallelepiped $N^{1.5} \times 1 \times 1$: the left hand side is about $1.5\log N + 3\log N = 4.5\log N$, while the right hand side is about $2\log N + 2\log N = 4\log N$ (for large N).

The solution for the combinatorial case is to allow splitting of the set V into two parts: one has (relatively) small projection length, the other has (relatively) small volume. Namely, the following statement is true (see [6, Section 10.7]):

> if for some three-dimensional set S the areas V_{12} and V_{13} of its two-dimensional projections onto coordinates $(1,2)$ and $(1,3)$ satisfy the inequality
> $$\log V_{12} + \log V_{13} \leqslant a + b,$$
> then the set can be split in two parts
> $$S = S' \cup S''$$
> in such a way that
> $$\log V_1' \leqslant a \quad \text{and} \quad \log V'' \leqslant b.$$
> Here V_1' is the measure of the (one-dimensional) projection of S' onto the first coordinate, and V'' is the volume of S''.

(Why do we introduce a and b? In a sense, we replace the inequality $u + v \leqslant w$ by an equivalent statement "for every a, b, if $w \leqslant a + b$, then either $u \leqslant a$ or $w \leqslant b$.)

We use a similar approach for dimensions, and get the following statement.

Proposition 2. *Let $S \subset \mathbb{R}^3$, and let $a, b \geqslant 0$ be two numbers such that*

$$\dim_p S_{12} + \dim_p S_{13} \leqslant a + b.$$

Then there exist a splitting $S = S' \cup S''$ such that

$$\dim_H(S_1') \leqslant a \quad \textit{and} \quad \dim_H(S'') \leqslant b.$$

[1]To be more combinatorial, one could consider finite sets and the cardinalities of those sets and their projections.

Here S'_1 is the projection of S' onto the first coordinate.

Proof. As before, fix an oracle X that minimizes the effective packing dimensions of S_{12} and S_{13}, and use it everywhere when speaking about complexities and effective dimensions. Then for every point $(s_1, s_2, s_3) \in S$ we have

$$dim_p^X(s_1, s_2) + dim_p^X(s_1, s_3) \leqslant a + b.$$

The inequality (4) then guarantees that

$$dim_H^X(s_1) + dim_H^X(s_1, s_2, s_3) \leqslant a + b$$

for every $(s_1, s_2, s_3) \in S$, and this inequality implies that *either* $dim_H^X(s_1) \leqslant a$ *or* $dim_H^X(s_1, s_2, s_3) \leqslant b$. Therefore we may split S into two sets S' and S'' and guarantee that for all elements $\langle s_1, s_2, s_3 \rangle \in S'$ we have $dim_H^X(s_1) \leqslant a$ and for all elements $\langle s_1, s_2, s_3 \rangle \in S''$ we have $dim_H^X(s_1, s_2, s_3) \leqslant b$. This implies that $\dim_H(S'_1) \leqslant a$ and $\dim_H(S'') \leqslant b$, as required. $\square$

3 Corollaries for dimensions: general statement

Proposition 2 can be generalized (with essentially the same proof) to arbitrary linear inequalities for Kolmogorov complexities. Fix some m, and consider an m-tuple of strings $\langle x_1, \ldots, x_m \rangle$. For every non-empty $I \subset \{1, \ldots, m\}$ we consider a sub-tuple x_I that consists of all x_i with $i \in I$. Consider some linear inequality for Kolmogorov complexities $\mathrm{C}(x_I)$ for all I; we assume that it is split between two parts to make the coefficients positive:

$$\sum_{I \in \mathcal{I}} \lambda_I \, \mathrm{C}(x_I) \leqslant \sum_{J \in \mathcal{J}} \mu_J \, \mathrm{C}(x_J) + O(\log n) \tag{5}$$

Here $\mathcal{I}$ and $\mathcal{J}$ are two disjoint families of subsets of $\{1, \ldots, m\}$, and λ_I and μ_J are positive reals defined for $I \in \mathcal{I}$ and $J \in \mathcal{J}$. Assume that this inequality is true for all n and for all tuples $\langle x_1, \ldots, x_m \rangle$ of n-bit strings (with a constant in $O(\log n)$-notation that does not depend on n and $x_1, \ldots, x_m$).

Then we have the corresponding result about dimensions.

Theorem 1. *Under these assumptions, for every set $S \subset \mathbb{R}^m$ and every non-negative reals a_I (defined for all $I \in \mathcal{I}$) such that*

$$\sum_{J \in \mathcal{J}} \mu_J \dim_p S_J \leqslant \sum_{I \in \mathcal{I}} \lambda_I a_I,$$

where S_J is the projection of S onto J-coordinates, there exist a splitting $S = \bigcup_{I \in \mathcal{I}} S^I$ such that

$$dim_H(S_I^I) \leqslant a_I.$$

The number of parts in the splitting is the same as the number of terms in the left-hand side of the inequality; they are indexed by $I \in \mathcal{I}$. By S_I^I we denote the I-projection of the part S^I; it is a set in $\mathbb{R}^k$ for $k = \#I$. The special case of Proposition 2 has two terms on both sides of the inequality ($\#\mathcal{I} = \#\mathcal{J} = 2$), and the coefficients λ_I and μ_J are all equal to 1.

Proof. As before, we consider some oracle X that makes the packing dimension of all S_J for all $J \in \mathcal{J}$ minimal. Then we have

$$\sum_{J\in\mathcal{J}} \mu_J \, \mathit{dim}_p^X \, s_J \leqslant \sum_{I\in\mathcal{I}} \lambda_I a_I,$$

for every point $s = \langle s_1, \ldots, s_m \rangle \in S$; here s_J stands for the projection of s onto J-coordinates. The inequality for Kolmogorov complexities (that we assumed to be true) gives (after dividing by n and taking the limit)

$$\sum_{I\in\mathcal{I}} \lambda_i \, \mathit{dim}_H^X \, s_I \leqslant \sum_{J\in\mathcal{J}} \mu_J \, \mathit{dim}_p^X \, s_J \left[\leqslant \sum_{I\in\mathcal{I}} \lambda_I a_I \right],$$

as before. (Note that the left hand side uses effective Hausdorff dimensions while the middle part uses effective packing dimensions, because of the limits.) This inequality is true for every point $s \in S$. Therefore, for every point $s \in S$ there exists some coordinate set $I \in \mathcal{I}$ such that

$$\mathit{dim}_H^X \, s_I \leqslant a_I,$$

and we can split the set S according to these indices and get sets S^I such that

$$\mathit{dim}_H^X \, s_I \leqslant a_I$$

for all points $s \in S^I$, and therefore

$$\dim_H S_I^I \leqslant a_I,$$

as required. □

4 Equivalence

We have shown that every (valid) linear inequality for entropies can be translated to a statement in dimension theory. In this section we show that this is an equivalent characterization:

Theorem 2. *If a linear inequality is not true for entropies, then the corresponding statement about dimensions, constructed as in Theorem 1, is false.*

Proof sketch. We combine several well-known tools to achieve this result.

1. The first one is the characterization of inequalities in terms of the size of subgroups mentioned above. Let G be some group, and let $H_1, \ldots, H_m$ be its subgroups. (We do not require them to be normal.) For every element $g \in G$ consider the cosets $g_1 = gH_1, \dots, g_m = gH_m$. If $g \in G$ is taken at random, the cosets $g_1, \dots, g_m$ become (jointly distributed) random variables (with common probability space G). We may consider tuples of them (g_I is a tuple of all g_i with $i \in I$), and check whether the inequality for entropies is valid for g_I. The result of Chan and Yeung [1] guarantees that if the inequality

$$\sum_{I \in \mathcal{I}} \lambda_I H(\xi_I) \leqslant \sum_{J \in \mathcal{J}} \mu_J H(\xi_J).$$

is not universally true (for all tuples of random variables $\langle \xi_1, \dots, \xi_n \rangle$), then there exists a counterexample with groups, i.e., a group G and its subgroups $H_1, \dots, H_m$ that make the inequality false:

$$\sum_{I \in \mathcal{I}} \lambda_I H(g_I) > \sum_{J \in \mathcal{J}} \mu_J H(g_J).$$

It is easy to see that every g_I has the same distribution as the coset gH_I for $H_I = \cap_{i \in I} H_i$, and that the distribution of g_I is uniform on its range. So the entropy $H(g_I)$ is the logarithm of the range size, i.e., $\log \#G - \log \#H_I$, and our inequality is essentially about the size of the subgroups and their intersections.

2. We also use standard results about the dimension of Cantor-type sets. Consider N-ary positional system where every real from $[0,1]$ is represented by an infinite sequence of digits $0 \dots N-1$. (As usual, the double representations for finite N-ary fractions do not matter much, and we ignore this problem.) Let X be a subset of $\{0, \dots, N-1\}$. We may consider the set C_X of all N-ary fraction with digits only in X (for example, the classical Cantor set is $C_{\{0,2\}}$ for $N = 3$). It is well known that $\dim_H(C_X) = \dim_p(C_X) = \log \#X / \log N$; this directly follows, for example, from the point-to-set principle.

One can consider similarly defined sets in $\mathbb{R}^2, \mathbb{R}^3$ etc. For example, let Y be a subset of $\{0, \dots, N-1\} \times \{0, \dots, N-1\}$. Then one can construct a set $C_Y \subset [0,1] \times [0,1]$ that consists of the pairs of N-ary fractions $(u_1 u_2 \dots, v_1 v_2 \dots)$ such that $(u_i, v_i) \in Y$ for every Y. The Hausdorff and packing dimensions of the set C_Y are $\log \#Y / \log N$.

A similar result is true for subsets of $[0,1]^m$ constructed in a similar way. Consider some set $A \subset \{0, \dots, N-1\}^m$. (Later, we let m be the number of variables in the inequality we consider, and we will explain how the set A can be constructed starting from the group G and its subgroups $H_1, \dots, H_m$.) Then consider the set $C_A \subset [0,1]^m$ that is constructed as explained above: $\langle x_1^1 x_2^1 \dots, x_1^2 x_2^2 \dots, \dots, x_1^m x_2^m \dots \rangle \in C_A$ if $\langle x_i^1, \dots, x_i^m \rangle \in A$ for all i. As we have noted, the dimension (packing or Hausdorff) of C_A is $\log \#A / \log n$.

The projection of the set C_A on some subset I of coordinates is the set C_{A_I} of the same type that is constructed starting from the projection A_I of A onto the same coordinates. Therefore, to find the dimensions of all projections of C_A, we need to know only the size of the projections of A.

3. We can consider a finite set $A \subset U_1 \times \dots \times U_m$ for arbitrary finite sets $U_1, \dots, U_m$. Then we identify arbitrarily all U_i with some subsets of $\{1, \dots, N\}$

for large enough N, and construct the corresponding set $C_A \subset [0,1]^m$. For that we need that $\#U_i \leqslant N$ for all i; the exact choice of N is not important since the factor $1/\log N$ is the same for all the projections.

Using this remark, we let U_i be the family of all cosets with respect to the subgroup H_i, and let

$$A = \{(gH_1, \ldots, gH_m)\colon g \in G\}.$$

Then, as we have seen, the dimension of $(C_A)_I$ is proportional to the logsize of the corresponding projection A_I, which equals the entropy of g_I:

$$\dim(C_A)_I = \frac{H(g_I)}{\log N}.$$

Therefore, we have

$$\sum_{I\in\mathcal{I}} \lambda_I \dim(C_A)_I > \sum_{J\in\mathcal{J}} \mu_J \dim(C_A)_J.$$

Here we may use both packing and Hausdorff dimensions, since for our sets they are the same.

But this is not what we need: we need to show that a splitting of C_A into sets with bounded dimensions of projections does not exist for some bounds a_I. Let us choose a_I slightly smaller than $\dim(C_A)_I$ so that still

$$\sum_{I\in\mathcal{I}} \lambda_I a_I > \sum_{J\in\mathcal{J}} \mu_J \dim(C_A)_J.$$

We want to show that the assumption about dimensions is false for those a_I, namely, that one cannot split C_A into a family of C^I (for $I \in \mathcal{I}$) in such a way that

$$\dim_H (C^I)_I \leqslant a_I$$

(here we have to specify the Hausdorff dimension since for the sets C^I the Hausdorff and packing dimensions may differ). For that we note that the last inequality implies

$$\dim_H (C^I)_I < \dim(C_A)_I.$$

It remains to show that the last inequality implies

$$\dim_H C^I < \dim C_A;$$

then we get a contradiction, since the set C_A cannot be represented as a finite union of sets of smaller Hausdorff dimensions.

To get the bound for $\dim_H C^I$ in terms of the dimension of its I-projection we use special properties of the set A that corresponds to the group G and its subgroups $H_1, \ldots, H_m$. Namely, for every set of indices I the projection of A onto A_I is uniform (every element that has preimages has the same number of preimages). We already mentioned a similar property when saying that the variable g_I is uniformly distributed on its image. The number of preimages is the ratio $\#\bigcap_{i\in I} H_i / \#\bigcap_{i=1,\ldots,m} H_i$. Using this property, we prove the following lemma.

Lemma. *Assume that the projection $\pi_I\colon A \to A_I$ (only I-coordinates remain) is uniform. Consider the set C_A and its projection $(C_A)_I$; let d be the difference in their dimensions: $d = \dim C_A - \dim(C_A)_J$. Then, for every $X \subset C_A$ we have*

$$\dim_H X \leqslant \dim_H X_I + d.$$

Note that this lemma deals with two different projections: mappings $\pi_I\colon A \to A_I$ (finite sets) and $\Pi_I : C_A \to (C_A)_I$ (coordinate spaces); the second one applies the first one simultaneously for every position in N-ary notation.

Proof of the lemma. The dimension of C_A is equal to $\log \#A/\log N$, and the dimension of $(C_A)_I = \Pi_I(C_A)$ is equal to $\log \#A_I/\log N$, so the difference is equal to $\log(\#A/\#A_I)/\log N$. Here the ratio $\#A/\#A_I$ is just the size of preimages for the uniform projection $\pi_i\colon A \to A_I$. To specify the first k digits in a point $x \in X$ we have to specify k digits of its projection $x_I \in X_I$, and also for every of k positions choose one of the preimages of some element of A_I. Now we may apply the point-to-set principle to get the desired result. □

The application of this lemma, as we have discussed, finishes the proof. □

Acknowledgments

The author is grateful to all his colleagues, especially to the members of the ESCAPE team in LIRMM and the Kolmogorov seminar, and the participants of the meeting organized by the American Institute of Mathematics in 2022 where some of the work presented here was discussed. Part of the work was supported by ANR FLITTLA grant.

References

[1] Chan T.H., Yeung R.W., On a relation between information inequalities and group theory, *IEEE Transactions on Information Theory*, **48**(7), 69–95 (2002)

[2] Hammer D., Romashchenko A., Shen A., Vereshchagin N., Inequalities for Shannon entropies and Kolmogorov complexities, *Journal of Computer and System Sciences*, **60**(2), 442–464 (2000).

[3] Kolmogorov A.N., Logical basis for information theory and probability theory, *IEEE Transactions of Information Theory*, **14**(5), 662–664 (1968)

[4] Lutz J., Lutz N., Algorithmic information, plane Kakeya sets, and conditional dimension, *ASM Transaction of Computation Theory*, **10**(2), 1–22.

[5] Matúš F., Csirmaz L., Entropy Region and Convolution, *IEEE Transactions on Information Theory*, **62**(11), 6007–6018 (2016)

[6] Shen A., Uspensky V.A., Vereshchagin N., *Kolmogorov complexity and algorithmic randomness*, American Mathematical Society, 2017. 511 pp.

[7] Zhang Z., Yeung R.W., On characterization of entropy function via informational inequalities, *IEEE Transactions on Information Theory*, **48**(7), 1440–1450.