Bienvenu, Laurent; **Csima, Barbara F.**; **Harrison-Trainor, Matthew**
Optimal bounds for single-source Kolmogorov extractors. (English) Zbl 1443.03022
Trans. Am. Math. Soc. 373, No. 3, 1983-2006 (2020).

The Kolmogorov complexity $C(\sigma)$ of a finite binary string σ measures its information content. Formally, it is defined as the length of the shortest program which, when run on a previously fixed universal interpreter, produces the string in question. As the choice of the universal interpreter is arbitrary within a certain class of admissible interpreters, this number is only meaningful up to a positive constant. This is why this notion is most interesting when applied to *in*finite binary sequences X by studying how fast Kolmogorov complexity approaches infinity when looking at longer and longer finite initial segments of such a sequence.

For infinite sequences with strong regularities (for instance, computable sequences, that is, sequences output by some algorithm) the growth rate will be quite low, meaning that the information content of the sequence becomes negligible in the limit. For other infinite sequences without regularities, such as the typical result of infinitely often tossing a fair coin, the Kolmogorov complexity of each initial segment might be roughly equal to its length. Thus, intuitively, we can think of these two kinds of sequences as possessing 0% and 100%, respectively, of what we might informally call information density. That density can easily be diluted, for example if we intersperse a sequence of high density with another one of low density according to some systematic pattern; for example, after each three bits from the one sequence we might insert one bit from the other.

The inverse problem is more interesting: If we have a sequence of information density strictly between 0% and 100%, can we algorithmically boil it down to a sequence of strictly higher density? This is known as the problem of randomness extraction and is the subject of the present article.

Of course, if we write $X\upharpoonright n$ for the first n bits of an infinite binary sequence X, $\lim_{n\to\infty} C(X\upharpoonright n)/n$ need not exist. Thus, it is necessary to look at the *effective Hausdorff dimension* $\dim(X) = \liminf_{n\to\infty} C(X\upharpoonright n)/n$ or the *effective packing dimension* $\mathrm{Dim}(X) = \limsup_{n\to\infty} C(X\upharpoonright n)/n$ instead. Then the question above becomes: Are there algorithms which can be run on infinite sequences X and which are guaranteed to produce as output infinite sequences Y such that $\dim(Y) > \dim(X)$ or such that $\mathrm{Dim}(Y) > \mathrm{Dim}(X)$?

In the case of dim, the answer is negative: *J. S. Miller* [Adv. Math. 226, No. 1, 373–384 (2011; Zbl 1214.03030)] showed that for every possible value of $\dim(X)$ there is some X for which no Y as above can be produced from X algorithmically.

But the case of Dim is very different, as *L. Fortnow* et al. [Lect. Notes Comput. Sci. 4051, 335–345 (2006; Zbl 1223.68060)] used results from complexity theory to show that from every infinite sequence X with $\mathrm{Dim}(X) > 0$ a sequence Y with $\mathrm{Dim}(Y)$ arbitrarily close to 1 can be extracted. Their extraction procedure is based on "almost uniform" extraction for *finite* strings; namely, they show that for every $0 < \alpha < \beta < 1$ there is an extraction function E which works on finite strings of information density at least α to extract another finite string of information density at least β if provided with some additional finite advice. *M. Zimand* ["Symmetry of information and bounds on nonuniform randomness extraction via Kolmogorov extractors", in: 26th Annual IEEE conference on computational complexity. Los Alamitos, CA: IEEE Computer Society. 148–156 (2011; doi:10.1109/CCC.2011.21)] made more exact how much advice is needed for which values of α and β.

The present paper improves Zimand's results. Its contents can be summarized as follows:

For $k \geq 1$, the authors define the set $\mathsf{EXT}(k)$ as those pairs (α, β) for which the extraction works by k different total computable functions (which corresponds to requiring roughly $\log_2 k$ bits of advice). In other words, $(\alpha, \beta) \in \mathsf{EXT}(k)$ if there is a function $f\colon \mathbb{N} \to \mathbb{N}$ and an algorithm which, for all n, on a string σ of length $f(n)$ as input, produces k strings of length n as output, and which has the property that if σ had information density at least α, then *at least one* of the k output strings has information density at least β.

The first main result of the article is an exact characterization of $\mathsf{EXT}(k)$ as follows: If $k = 1$, then $(\alpha, \beta) \in \mathsf{EXT}(k)$ if and only if $\beta \leq \alpha$, and therefore no meaningful randomness extraction is possible in

this case. For $k \geq 2$, $(\alpha, \beta) \in \mathsf{EXT}(k)$ if and only if either $\alpha = \beta \in \{0, 1\}$ or $\beta < \frac{k\alpha}{1+(k-1)\alpha}$. The proof goes by first proving the equivalence of the possibility of extraction for some (α, β) to the existence of a sequence of hypergraphs whose hyperedges are distributed somewhat evenly, where the exact quality of this evenness is parametrized by α and β.

Then, to get the positive part (when extraction is possible) of their characterization, the authors use the probabilistic method to show the existence of the needed hypergraphs for the right choices of α and β. The authors point out that they do not know whether their extraction technique can be carried out in polynomial time, because it is unknown whether hypergraphs with the properties they require can be constructed in polynomial time.

On the other hand, to get the negative part of their characterization, the authors give a lengthy combinatorial argument that shows that the wrong values of α and β lead to contradictions in connection with the length $f(n)$ of the extraction source string σ, the amount of information that can be contained within this string, and information conservation.

Next, the authors modify the setting of their first main result, by modifying the definition of $\mathsf{EXT}(k)$ in such a way as to allow the extraction algorithm to be partial on some bad advices. They call $\mathsf{EXT}^p(k)$ the set of pairs (α, β) resulting from this modification. Kolmogorov complexity is a function that is only algorithmically approximable from above and not computable; meaning that there is never certainty about whether a short description that was found for some string is already the shortest. Thus, one might think that allowing partial functions here might enable more risky extraction strategies that are allowed to fail, for instance, when they get stuck in an infinite loop searching for non-existent shorter descriptions for some strings.

However, the authors show that allowing such strategies only makes a marginal difference when compared with the case where the extraction algorithm is required to always be total; namely, they show that $\mathsf{EXT}^p(k)$ equals $\mathsf{EXT}(k)$ plus some computable pairs on the border of $\mathsf{EXT}(k)$.

In the last section of the article, the authors use a variant of their hypergraph argument to slightly improve another of Zimand's results [loc. cit.] in the setting where the advice is not of constant size anymore, but is allowed to grow computably in the input length.

Reviewer: Rupert Hölzl (Neubiberg)

MSC:

03D32 Algorithmic randomness and dimension
05C80 Random graphs (graph-theoretic aspects)
68Q30 Algorithmic information theory (Kolmogorov complexity, etc.)

Keywords:

randomness extraction; Kolmogorov complexity; effective packing dimension; hypergraphs

Full Text: DOI

References:

[1] Bienvenu, Laurent; Doty, David; Stephan, Frank, Constructive dimension and Turing degrees, Theory Comput. Syst., 45, 4, 740-755 (2009) · Zbl 1183.68281

[2] Barak, Boaz; Impagliazzo, Russell; Wigderson, Avi, Extracting randomness using few independent sources, SIAM J. Comput., 36, 4, 1095-1118 (2006) · Zbl 1127.68030

[3] Chung, Fan R. K., Constructing random-like graphs. Probabilistic combinatorics and its applications, San Francisco, CA, 1991, Proc. Sympos. Appl. Math. 44, 21-55 (1991), Amer. Math. Soc., Providence, RI

[4] Conidis, Chris J., A real of strictly positive effective packing dimension that does not compute a real of effective packing dimension one, J. Symbolic Logic, 77, 2, 447-474 (2012) · Zbl 1251.03047

[5] Downey, Rodney G.; Hirschfeldt, Denis R., Algorithmic randomness and complexity, Theory and Applications of Computability, xxviii+855 pp. (2010), Springer, New York · Zbl 1221.68005

[6] Fortnow, Lance; Hitchcock, John M.; Pavan, A.; Vinodchandran, N. V.; Wang, Fengming, Extracting Kolmogorov complexity with applications to dimension zero-one laws. Automata, languages and programming. Part I, Lecture Notes in Comput. Sci. 4051, 335-345 (2006), Springer, Berlin · Zbl 1223.68060

[7] Li, Ming; Vitányi, Paul, An introduction to Kolmogorov complexity and its applications, Texts in Computer Science, xxiv+790 pp. (2008), Springer, New York · Zbl 1185.68369

[8] Miller, Joseph S., Extracting information is hard: a Turing degree of non-integral effective Hausdorff dimension, Adv. Math., 226, 1, 373-384 (2011) · Zbl 1214.03030

[9] Mitzenmacher, Michael; Upfal, Eli, Probability and computing, xx+467 pp. (2017), Cambridge University Press, Cambridge · Zbl 1368.60002

[10] Nies, André, Computability and randomness, Oxford Logic Guides 51, xvi+433 pp. (2009), Oxford University Press, Oxford · Zbl 1237.03027

Rei04Reimann2004 Jan Reimann. \newblock \em Computability and fractal dimension. \newblock PhD thesis, Universität Heidelberg, 2004.

[11] Thomason, Andrew, Pseudorandom graphs. Random graphs '85, Poznań, 1985, North-Holland Math. Stud. 144, 307-331 (1987), North-Holland, Amsterdam

[12] Thomason, Andrew, Random graphs, strongly regular graphs and pseudorandom graphs. Surveys in combinatorics 1987, New Cross, 1987, London Math. Soc. Lecture Note Ser. 123, 173-195 (1987), Cambridge Univ. Press, Cambridge

[13] Vereshchagin, Nikolai K.; Vyugin, Michael V., Independent minimum length programs to translate between given strings, Theoret. Comput. Sci., 271, 1-2, 131-143 (2002) · Zbl 0992.68083

Zim10Zimand2010 Marius Zimand. \emph Possibilities and impossibilities in Kolmogorov complexity extraction, SIGACT News, Dec 2010.

[14] Zimand, Marius, Symmetry of information and bounds on nonuniform randomness extraction via Kolmogorov extractors. 26th Annual IEEE Conference on Computational Complexity, 148-156 (2011), IEEE Computer Soc., Los Alamitos, CA