Research Statement

Daniel Turetsky

Introduction

The two main areas of my research in the general area of computability theory are algorithmic randomness, and computable algebra and model theory. Randomness is a notion which is captured by in several different ways. An infinite sequence being random can be defined to mean that it contains large amounts of information, or that it is hard to predict the behavior of, or that it exhibits no atypical properties, and all are equivalent definitions. Effective dimension is a measurement of the extent to which a real is partially random. This can be defined by the real containing a smaller amount of information, or the real being only somewhat predictable, and these again yield equivalent definitions.

Computability theory seeks to understand the effective content of mathematics. Although many mathematical objects exist, computability theory asks the question of whether they can be algorithmically found. For example, although every vector space has a basis, computability theory tells us that finding a basis is not always algorithmically possible. Computable algebra is the analysis of classical mathematical structures such as rings, graphs or linear orders using the tools of computability theory, while computable model theory generalizes this to arbitrary mathematical structures.

1 Algorithmic Randomness

Attempts to formalize randomness began with von Mises in 1919, when he argued in [18] that an infinite sequence should be considered random if all reasonable subsequences obey the law of large numbers. Unfortunately, at the time, what constituted a reasonable subsequence was unclear. Later, with the advent of computability theory, it was suggested that the reasonable subsequences should be the computable ones. This, however, proved to be too weak a notion, and so stronger ones were sought. In 1966, Martin-Löf in [12] provided the first workable characterization of a random sequence as one which avoided all reasonable sets of measure 0. Von Mises's notion can still be seen as the precursor of the modern martingale definition.

Intuitively, an infinite sequence is random if the density of information in it is maximal. For instance, consider the following sequences:

$$0000000000000000000000000000000000000000000000000000000000000\ldots$$
$$1010101010101010101010101010101010101010101010101010101010\ldots$$
$$0101100000111000000101101000001110011011001001011010000010001 0010\ldots$$

The first sequence could be fully described as "the infinite sequence of all zeros", and the second as "the infinite sequence of alternating zeros and ones beginning with one". These are both finite descriptions which give all the information of the sequence, so these sequences are computable. The third sequence, meanwhile, is taken from neutrino bombardment data, and so is presumably random. Computable and random are thus in some way at opposite ends of the spectrum.

This intuition is made more precise by analyzing the complexity of finite initial segments of the sequence. For a finite string σ, the *Kolmogorov complexity* of that string is the smallest size to which it can be compressed. We denote this by $K(\sigma)$. For example, consider the string of one trillion zeros. It can be compressed as the much smaller string "one trillion zeros". Thus its Kolmogorov complexity is low, even though it is quite long.

If x is an infinite sequence, we use $x \upharpoonright k$ to denote the initial segment of length k. Note that one description of a string is itself, and so the Kolmogorov complexity of a string will never be much more than its own length. The very complicated strings will be those with complexity not much less than their own length. A sequence x is *random* if all its initial segments are very complicated, i.e. there is some b such that $K(x \upharpoonright k) \geq k - b$ for all k.

1.1 Effective Dimension and Topology

Very broadly, Lebesgue measure separates the world into the sets of positive measure and those of measure 0. This classification is rather coarse, however. In $\mathbb{R}^2$, for example, points and lines are indistinguishable by Lebesgue measure, as they all have measure 0. Notions of dimension, such as Hausdorff dimension or packing dimension, can strengthen this classification by separating certain sets of measure 0. Points have Hausdorff dimension 0, while lines have Hausdorff dimension 1.

Similarly, randomness separates sequences into those which are random and those which are not. Effective dimension refines this classification by separating certain nonrandom sequences. While the classical dimension of a singleton is always 0, the effective dimension of a singleton can be nonzero, so effective dimension often studies singletons (i.e. points).

The study of effective dimension began when Lutz in [11] proved an alternate characterization of the classical notion of Hausdorff dimension. Athreya, Hitchcock, Lutz and Mayordomo extended this to packing dimension in [1]. These alternate characterizations were effectivized, giving rise to the concept of effective Hausdorff dimension and effective packing dimension. Mayordomo in [13] and Athreya, Hitchcock, Lutz and Mayordomo in [1] then showed the equivalence of Definition 1.1.

Definition 1.1. For an infinite sequence x, define the *effective Hausdorff dimension* of x as

$$\dim_H(x) := \liminf_{s\to\infty} \frac{K(x \upharpoonright s)}{s}.$$

Notice that this value is necessarily at most 1. Effective dimension thus measures when a sequence is partially random, with random sequences having dimension 1 (although the converse fails). I am interested in the question of whether, given a sequence of positive dimension, one can extract in an effective manner a sequence of greater dimension? This is important to any sort of computable application which depends on a source of randomness, such as weather modeling or cryptography. The question also arises in ergodic theory [4].

It is known that this is possible in some cases ([2]), while there are also examples where it is impossible ([15]).

Analogous to defining the effective dimension of an infinite sequence, one can define the effective Hausdorff dimension of a point in $\mathbb{R}^n$. There are several equivalent ways of doing this, one being to simply consider the binary expansions of the coordinates.

One can then ask how frequent points of any given dimension are. Several easy facts follow:

- For $z \in \mathbb{R}^n$, $\dim_H(z) \in [0, n]$.

- For every $\alpha \in [0, n]$, there are densely many $z \in \mathbb{R}^n$ with $\dim_H(z) = \alpha$.
- The set of $z \in \mathbb{R}^n$ with $\dim_H(z) < n$ has Lebesgue measure 0.
- The set of $z \in \mathbb{R}^n$ with $\dim_H(z) > 0$ is meager.

My interest is in the question of how various topological restraints on a set affect the effective dimension of the points within it.

One approach would be to consider sets of a particular classical Hausdorff dimension; indeed, the two notions are tightly connected. A consequence of Lutz's work in [11] is that if X has classical Hausdorff dimension β, it contains infinitely many points of effective Hausdorff dimension greater than $\beta - \epsilon$ for any positive ϵ.

Alternatively, one might consider sets which are manifolds of a given dimension. For any k-plane orthogonal to the axes, there is some α such that the effective Hausdorff dimension of every point on the plane lies in the interval $[\alpha, \alpha+k]$. In fact, for any such interval $[\alpha, \alpha+k]$ with $0 \leq \alpha < \alpha+k \leq n$, there exists an appropriate k-plane in $\mathbb{R}^n$ with every $\beta \in [\alpha, \alpha + k]$ occurring as the dimension of densely many points on the plane.

It is natural to wonder if this is best possible, or if one could make a k-manifold with fewer dimensions occurring. As partial answer to this question, I have shown the following:

Theorem 1.2 (Turetsky [17])**.** *If $X \subseteq \mathbb{R}^2$ is a 1-manifold (not necessarily smooth), X must contain densely many points x with* $\dim_H(x) = 1$.

The techniques used in the proof of this result have allowed me to show the following:

Theorem 1.3 (Turetsky [17])**.** *The set* $\mathrm{DIM}^{\{1\}} = \{x \in \mathbb{R}^2 : \dim_H(x) = 1\}$ *is connected.*

Higher dimension analogs of these results hold as well.

1.2 Future Work

I intend to investigate the question of dimension extraction, as discussed above. An important tool towards this is understanding when sequences can compute sequences of smaller dimension. These are both captured in the following general question:

Question 1.4. If a sequence has positive dimension α, for what other real numbers β can this sequence compute a sequence of dimension β?

Although I have proven some higher dimensional analogs of Theorems 1.2 and 1.3, there is still work to be done on them.

Question 1.5. If $X \subseteq \mathbb{R}^n$ is a k-manifold, precisely what are the possible sets $B \subseteq [0, n]$ such that $B = \{\dim_H(x) : x \in X\}$?

I would also like to strengthen Theorem 1.3:

Question 1.6. Is the set $\mathrm{DIM}^{\{1\}} = \{x \in \mathbb{R}^2 : \dim_H(x) = 1\}$ path-connected?

I am also interested in those sequence which are low for dimension. A sequence x is called *low for dimension* if it cannot decrease the dimension of any sequence. That is, if every sequence has the same dimension when Kolmogorov complexity is computed with full knowledge of x.

Question 1.7. What are the low for dimension sequences?

Leaving effective dimension, there is also a notion of low for random. Like low for dimension, a sequence is called *low for random* if every random sequence is still random when Kolmogorov complexity is computed with full knowledge of x.

Question 1.8. Is there a characterization of the low for randoms involving only notions from classical computability theory (not mentioning randomness, Kolmogorov complexity or measure)?

2 Computable Algebra and Computable Model Theory

Computable algebra is the investigation of the complexity of a class of algebraic structures and the relations on those structures. A typical question is which members of a class of countable structures are computable. For example, one might consider which abelian groups are computable. This complexity is usually analyzed in terms of Kleene's arithmetic hierarchy:

Definition 2.1. A logical formula φ, using the standard symbols from arithmetic, is called both Σ_0 and Π_0 if it contains no quantifiers (i.e. it has neither a $\forall$ nor an $\exists$).

A logical formula φ is called Σ_{n+1} if $\varphi = \exists x_0 \dots \exists x_n \psi$, where ψ is Π_n.

A logical formula φ is called Π_{n+1} if $\varphi = \forall x_0 \dots \forall x_n \psi$, where ψ is Σ_n.

A set $X \subseteq \mathbb{N}$ is called Σ_n (Π_n) if there is a Σ_n (Π_n) formula φ such that $X = \{x \in \mathbb{N} : \varphi \text{ holds of } x\}$.

A set which is both Σ_n and Π_n is called Δ_n.

The arithmetic hierarchy is closely connected to computability. For example, the computable sets are precisely the Δ_1 sets. Notice that if a set X is either Σ_n or Π_n, then it is both Σ_{n+1} and Π_{n+1}. These form a proper hierarchy, since there are Σ_{n+1} sets which are neither Σ_n nor Π_n. Similarly for Π_{n+1}. Many natural relations on a structure are Σ_n or Π_n for some n, and computable algebra is often interested in determining the least such n.

Rather than restricting attention to particular classes of classical algebraic structures, computable model theory generalizes the notion of what an algebraic structure is and asks these same questions in this general context.

2.1 Boolean Algebras

Boolean algebras are a difficult structure to work with, because from the point of view of computability theory, they are incredibly homogenous. Thus it can be very hard to gain any information from a boolean algebra in an algorithmically effective way.

On the elements of a boolean algebra, a natural property to consider is that of being an atom (i.e., bounding no non-zero elements). For a boolean algebra A, we will denote the set of atoms of A by atom(A). Atom(A) is easily seen to be Π_1, and one can construct boolean algebras B for which atom(B) cannot be Π_0. In fact, it turns out that practically any boolean algebra has a copy that can be used as such an example:

Theorem 2.2 (Remmel [16])**.** *A computable boolean algebra A is isomorphic to a computable boolean algebra B such that atom(B) computes all Δ_2 sets iff A has infinitely many atoms.*

On the other hand, every boolean algebra has a copy for which the set of atoms is not computationally powerful:

Theorem 2.3 (Downey [3])**.** *Every computable boolean algebra A is isomorphic to a computable boolean algebra B such that atom(B) does not compute all Δ_2 sets.*

I have considered the properties of being atomless (bounding no atoms) or infinite (bounding infinitely many elements). These relations are both Π_2, and I have proven results analogous to Theorems 2.2 and 2.3 for these relations.

A longstanding open question is the Low$_n$ conjecture [10]. In an effort to better understand this question, Harris and Montalbán have made an effort to understand the possible relations on boolean algebras [7]. I intend to look further at these.

2.2 Linear Orders

Linear orders are an interesting algebraic structure, because they naively seem so simple. They are also closely tied to boolean algebras. The properties of atom, atomless and infinite all have analogs for linear orders: adjacent, dense and infinitely-far-apart (e.g., two elements x and y are adjacent if $x < y$ and the interval (x, y) is empty).

For adjacent, Downey, Lempp and Wu have shown that the natural analog of Theorem 2.2 holds. I have shown the corresponding result for dense, but for infinite the situation is more complex. I have a number of partial results, but with Downey, Kach and Lempp, I have constructed a linear order demonstrating that the natural analog fails. The full theorem is still unknown.

Question 2.4. For which computable linear orders A is there a computable copy B such that the infinite relation on B computes all Δ_3 sets?

Interestingly, Theorem 2.3 fails for linear orders. Downey and Moses showed that there is a computable linear order such that in any isomorphic copy, the set of adjacencies computes all Δ_2 sets. I have shown the analogous result for dense and infinite, and for every n, have generalized it to hold for a particular relation which is Π_n.

A different sort of question one might consider is how one can code sets into linear orders (this sort of coding is impossible for boolean algebras because of their homogeneity). For example, let ζ denote the integers (positive and negative) as a linear order. Then for a set $X \subseteq \mathbb{N}$, $X = \{x_0 < x_1 < x_2 < \dots\}$, consider the linear order:

$$L_X = \zeta + x_0 + \zeta + x_1 + \zeta + x_2 + \zeta + \dots$$

That is, we have a copy of ζ, then x_0 many points, then a copy of ζ, then x_1 many points, then a copy of ζ, etc. One might ask, for which sets X is L_X computable? Lerman showed that L_X is computable iff X is Σ_3.

If we let η denote the rational numbers as a linear order, and replace ζ with η, we get a linear order S_X which is called a *strong η-representation for X*. If we remove the requirement that the x_i be in increasing order, the resulting linear order is called a *weak η-representation for X*. Harris [6] classified which sets have computable weak η-representations, but the question for strong η-representations remains open. I have shown, however, that the tools used for the classification of the weak η-representations will not suffice.

Question 2.5. For which sets X is the strong η-representation of X computable?

2.3 Abelian Groups

The group operation imposes a great deal of structure that is lacking in linear orders and even boolean algebras. One can freely add a single element to a linear order and still have a linear order, but adding an element to a group requires adding a product for that element with every other element in the group, and the relationships of these products must be chosen intelligently if the result is to still be a group. Restricting attention to abelian groups makes this simpler, but there are still interesting questions.

We can start by considering torsion free abelian groups. For a prime p, let $\mathbb{Q}_p$ denote the subgroup of the rational numbers consisting of the fractions with denominator p^n for some n. Then for P a set of primes, consider the groups

$$G_P = \bigoplus_{p \in P} \mathbb{Q}_p,$$

$$H_P = \bigoplus_{p \in P} \mathbb{Q}_p \oplus \bigoplus_{n \in \mathbb{N}} \mathbb{Z}.$$

Again, the natural question to ask is for what sets P are these computable. Melnikov [14] has shown that H_P is computable iff P is Σ_3. However, a classification of those P for which G_P is computable is unknown. Downey, Kach, Melnikov and I are close to showing that there is a Σ_3 set P for which G_P is not computable.

Question 2.6. For which P is G_P computable?

Turning instead to reduced abelian p-groups, Khisamiev [8, 9] has used Ulm invariants to partially classify which are computable. This classification has been proven to work on groups of height strictly less than ω^2. The classification does not appear to extend to groups of height ω^2 or more, but no counterexample is known.

Question 2.7. Which reduced abelian p-groups are computable?

2.4 Computably Model Theory

Computable categoricity is an effective version of categoricity from model theory. For a cardinal κ, a system of axioms is said to be κ-categorical if every structure of cardinality κ which satisfies those axioms is isomorphic. In making this notion effective, we restrict our attention to computable isomorphisms. Rather than considering all structures which satisfy a given axiom system, we consider only those which are isomorphic (but not necessarily computably so).

Definition 2.8. A computable structure $\mathfrak{A}$ is *computably categorical* if for any other computable structure $\mathfrak{B}$ with $\mathfrak{A} \cong \mathfrak{B}$, there exists a total computable function f with $f : \mathfrak{A} \cong \mathfrak{B}$.

For example, any linear ordering which contains no adjacencies is computably categorical; given two computable copies, one can perform a back-and-forth construction to create an isomorphism. A linear order with only finitely many adjacencies is also computably categorical, because one could begin by correctly mapping the finitely many points, then run the back-and-forth construction. In fact, this completely characterizes the computably categorical linear orders; any computable linear order with infinitely many adjacencies (e.g., the integers as a linear order) has two computable copies between which there is no computable isomorphism.

One might expect that every computably categorical structure is such because one can run a back-and-forth construction to create the isomorphism, but this turns out to correspond to a stronger notion.

Definition 2.9. A computable structure $\mathfrak{A}$ is *relatively computably categorical* if for any other structure $\mathfrak{B}$ (not necessarily computable) with $\mathfrak{A} \cong \mathfrak{B}$, there exists a total function f computable from $\mathfrak{B}$ with $f : \mathfrak{A} \cong \mathfrak{B}$.

There are analogs of both computable categoricity and relative computable categoricity for Δ_n.

Very surprisingly, relative computable categoricity is a strictly stronger property than computable categoricity. With Day, Downey, Kach, Lempp and Ng, I have shown that the index set of all relatively computably categorical structures is Σ_3-complete. White showed that the the index set of all computably categorical structures is at least Π_4 [19]. It is not known if the set of computably categorical structures is arithmetical, or even hyperarithmetical.

Although these properties are different, under certain additional assumptions they are equivalent.

Theorem 2.10 (Goncharov [5])**.** *Every 2-decidable, computably categorical structure is relatively computably categorical.*

A structure $\mathfrak{A}$ is 2-decidable if there is a computable process to decide if, given a Σ_2 formula φ and a finite tuple of elements $\overline{a} \in \mathfrak{A}$, does φ hold of $\overline{a}$.

2.5 Further Work in Computable Model Theory

It is known that there are structures which are computably categorical but not relatively computably categorical. In a sense, the computable categoricity of these structures is not due to any algebraic properties, but rather a fluke of the computable structure. However, I conjecture (and hope to prove) the following:

Conjecture 2.11. Every computably categorical structure is relatively Δ^0_3-categorical.

Day, Downey, Kach, Lempp, Ng and I are close to showing:

Conjecture 2.12. Every 1-decidable, computably categorical structure is relatively Δ^0_2-categorical.

There are also some natural index set questions. Every computably categorical structure has a Σ_3 index set. Day, Downey, Kach, Lempp, Ng and I have shown that there is a computably categorical structure with index set Σ_3-complete. However, every known example of a relatively computably categorical structure has only a Π_2 index set.

Question 2.13. Is there a relatively computably categorical structure with index set Σ_3-complete?

For computably categorical, there is a different index set question:

Question 2.14. What is the complexity of the index set of all computably categorical structures?

If Conjecture 2.11 is true, then the answer to the above is at most Σ_6.

References

[1] Krishna B. Athreya, John M. Hitchcock, Jack H. Lutz, and Elvira Mayordomo. Effective strong dimension in algorithmic information and computational complexity. In *STACS 2004*, volume 2996 of *Lecture Notes in Comput. Sci.*, pages 632–643. Springer, Berlin, 2004.

[2] Laurent Bienvenu, David Doty, and Frank Stephan. Constructive dimension and weak truth-table degrees. In *Computation and Logic in the Real World - Third Conference of Computability in Europe (Sienna, 2007)*, pages 63-72.

[3] Rod Downey. Every recursive Boolean algebra is isomorphic to one with incomplete atoms. *Ann. Pure Appl. Logic*, 60(3):193–206, 1993.

[4] Péter Gács. The Boltzmann Entropy and Randomness Tests. *Proceedings of the Workshop on Physics and Computation*, pages 209–216. IEEE Computer Soc., 1994.

[5] Sergei S. Goncharov. Autostability and computable families of constructivizations. *Algebra i Logika*, 14:392–408, 1975 (English translation).

[6] Kenneth Harris. η-representation of sets and degrees. *J. Symbolic Logic*, 73(4):1097–1121, 2008.

[7] Kenneth Harris and Antonio Montalbán. On the n-Back-and-Forth types of Boolean Algebras.

[8] N. G. Khisamiev. The arithmetic hierarchy of abelian groups. *Sibirsk. Mat. Zh.*, 29(6):144–159, 1988.

[9] N. G. Khisamiev. Constructive abelian groups. In *Handbook of recursive mathematics, Vol. 2*, volume 139 of *Stud. Logic Found. Math.*, pages 1177–1231. North-Holland, Amsterdam, 1998.

[10] Julia F. Knight and Michael Stob. Computable Boolean algebras. *J. Symbolic Logic*, 65(4):1605–1623, 2000.

[11] Jack H. Lutz. Dimension in complexity classes. In *15th Annual IEEE Conference on Computational Complexity (Florence, 2000)*, pages 158–169. IEEE Computer Soc., Los Alamitos, CA, 2000.

[12] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[13] Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Inform. Process. Lett.*, 84(1):1–3, 2002.

[14] Alexander G. Melnikov. Enumerations and completely decomposable torsion-free abelian groups. *Theory of Comp. Sys.*, 45(4):897–916, 2009.

[15] Joseph S. Miller. Extracting information is hard: a Turing degree of non-integral effective Hausdorff dimension. To appear in *Advances in Mathematics.*

[16] J. B. Remmel. Recursive isomorphism types of recursive Boolean algebras. *J. Symbolic Logic*, 46(3):572–594, 1981.

[17] Daniel Turetsky. Connectedness properties of dimension level sets. In preparation.

[18] Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.*, vol. 5 (1919), 52-99.

[19] Walker M. White. On the complexity of categoricity in computable structures. *MLQ Math. Log. Q.*, 49(6):603–614, 2003.