

COMPLEXITY OF DEEP COMPUTATIONS VIA TOPOLOGY OF FUNCTION SPACES

EDUARDO DUEÑEZ¹ JOSÉ IOVINO¹ TONATIUH MATOS-WIEDERHOLD²
LUCIANO SALVETTI² FRANKLIN D. TALL²

¹Department of Mathematics, University of Texas at San Antonio

²Department of Mathematics, University of Toronto

ABSTRACT. We bring two theories to bear on the complexity of deep computations, that is, of limits of computations along ultrafilters. The first is the topology of spaces of functions of the first Baire class, which culminates in Todorćević’s classification of Rosenthal compacta; the second is Shelah’s model-theoretic notion of independence. Such limits arise whenever an iterative procedure is carried to its conclusion. Their instances range from the classical methods of numerical analysis, such as Newton’s method and the power method, to the recurrent, equilibrium and continuous-depth architectures of deep learning. A deep computation, however, is in general a nonconstructive object: it is an ultrafilter limit of possibly uncountably many computations, and it need not be the limit of any sequence of them. Shelah’s *independence property* is a dividing line of model theory akin to the Vapnik–Chervonenkis dimension of statistical learning. Suppose that states are observed through countably many real-valued features, and consider the regions of the state space on which these features are bounded and which the computations of a given set preserve. We show that the set lacks the independence property exactly when its deep computations, extended to the limit states of each such region, are of the first Baire class there. In that case, every deep computation is, on each such region, the limit of a sequence of computations. Todorćević’s classification then yields a strict hierarchy of complexity classes of deep computations. From the top down, its levels are attained by a rotation of the circle, the power method, Bresenham’s line algorithm in the plane and in space, and the Euclidean algorithm. Newton’s method for the square roots of unity on the Riemann sphere, by contrast, lies beyond the hierarchy. For algorithms that iterate continuous steps on a compact metrizable space of states, the hierarchy coincides with the Glasner–Megrelishvili hierarchy of tame dynamical systems. Finally, for any number of features, the absence of the independence property makes every deep computation, extended to the limit states of each such region, universally measurable there. Consequently, whatever Radon probability measure governs the inputs, each output feature of such a deep computation is, outside a set of arbitrarily small probability, a uniform limit of polynomials in the input features.

Date: October 26, 2026.

2020 Mathematics Subject Classification. Primary 68Q15, 03C45, 54H30; Secondary 03C98, 03D15, 05D10, 26A21, 37B05, 54C35, 68T07.

Key words and phrases. Deep computations, limit computations, complexity classes, deep learning, recurrent neural networks, deep equilibrium models, compositional computation structures, independence property, NIP, model theory, Rosenthal compacta, Baire class 1 functions, Todorćević’s trichotomy, Argyros–Dodos–Kanellopoulos heptachotomy, Bourgain–Fremlin–Talagrand theorem, tame dynamical systems, infinite Ramsey theory.

INTRODUCTION

This paper brings Shelah’s theory of independence and Todorćević’s classification of Rosenthal compacta to bear on the complexity of deep computations, the limits of computations along ultrafilters; with these tools we determine when a deep computation is the limit of a sequence of computations, and we stratify deep computations into a strict hierarchy of complexity classes. The interest of the undertaking lies in the fact that a deep computation need not inherit the properties of the computations that converge to it. A pointwise limit of continuous maps, for instance, need not be continuous. Likewise, a limit taken along an ultrafilter over uncountably many computations need not be the limit of any sequence of them, in which case it lies beyond the reach of sequential approximation. Deep computations arise whenever an iterative procedure is carried to its conclusion. The classical methods of numerical analysis, such as Newton’s method for the roots of a polynomial and the power method for the eigenvectors of a matrix, furnish instances of such procedures. In Sections 4 and 5 we analyze the deep computations of these two methods. In deep learning, several architectures are defined as limits of computations. A recurrent network applies one layer over and over, with the same weights in every layer. The deep equilibrium models of Bai, Kolter and Koltun [BKK] compute directly, by root finding, the fixed point that an infinitely deep network of this kind would reach. A residual network, in turn, may be regarded as a discretization of a differential equation. The neural ordinary differential equations of Chen, Rubanova, Bettencourt and Duvenaud [CRBD] replace the discrete sequence of layers of a residual network by a differential equation for the hidden state. The equation for the hidden state is obtained in the limit in which the time step between consecutive layers tends to zero. Neural ordinary differential equations belong to a broader effort to bring differential equations into deep learning. This effort also includes the physics-informed neural networks of Raissi, Perdikaris and Karniadakis [RPK19], in which the differential equations of a physical system are built into the training of a network. In these models, as in the classical methods, the object of interest is a limit of computations rather than a single computation.

The framework of this paper is that of compositional computation structures. Alva, Dueñez, Iovino and Walton introduced these structures in the first paper of this series [ADIW26] to accommodate all the limits just described in a single setting. The framework concerns classical models of computation, in which real-valued quantities are known and manipulated to a finite degree of precision. In this framework, computations act on states, and each state is observed through the values of real-valued *features*. In that paper the authors introduced two closely related notions of limit of computations. A *deep computation*, so called after the usage in machine learning [BKK], is a pointwise limit of computations along an ultrafilter, evaluated on the states themselves. An *ultracomputation*, so called because it arises as an ultrafilter limit of computations, is a pointwise limit of the extensions of computations to the ideal limit states, the types described below. Every deep computation is the restriction of an ultracomputation to the states, and for the iterative algorithms of Section 5 the two notions coincide. The analysis takes place on *shards*, the sets of states on which each feature obeys a prescribed bound, and on *type shards*. The type shards, the closures of the shards in the space of types, adjoin to each shard its limit states. Strictly speaking, the classification below concerns

ultracomputations restricted to type shards, but we speak of deep computations wherever the distinction plays no role. A set of computations is *stable* when all its deep computations, extended to the limit states, are continuous on every type shard. The analysis of stable sets rests on a theorem proved by Grothendieck in 1952 [Gro52], his double limit criterion. By this criterion, a uniformly bounded set of continuous functions on a compact space has only continuous functions in its pointwise closure exactly when the set allows the exchange of iterated limits. In other words, for every sequence of functions from the set and every sequence of points of the space, the two iterated limits agree whenever both exist. Using this criterion, Alva, Dueñez, Iovino and Walton characterized the stable sets by an exchange of iterated limits and showed that their deep computations are *polynomially approximable*. That is, on every shard, each feature of the output of such a deep computation is a uniform limit of polynomials in the features of the input. Polynomial approximability corresponds to continuous extendibility in topology, and to *stability* and *type definability* in model theory.

Beyond stability, Shelah’s classification theory [She90] singles out the *independence property*, which Shelah isolated in 1971 [She71, Definition 4.1]. A formula has the independence property when, for every natural number n , there are n parameters on which every pattern of truth and falsity of the formula is realized by some element of a model. A theory none of whose formulas has the independence property is called *NIP*, or *dependent*. Within a few years, the same dividing line appeared independently in two other fields, as a dichotomy between polynomial and exponential growth. A family of sets shatters a finite set when it cuts out all subsets of that set as traces, that is, as intersections of the set with members of the family. In these terms, either a family shatters arbitrarily large finite sets, or the number of its traces on every finite set is bounded by a polynomial in the size of the set. In statistical learning, Vapnik and Chervonenkis announced this dichotomy in 1968 [VC68] and published its proof in 1971 [VC71], in their study of the uniform convergence of relative frequencies to probabilities. In combinatorics, Sauer [Sau72] and Shelah [She72] proved the dichotomy in 1972, and it is now known as the Sauer–Shelah lemma. The largest size of a finite set shattered by a family is called the *Vapnik–Chervonenkis dimension* of the family.

In analysis, the same combinatorial pattern emerged from the study of Banach spaces. Rosenthal’s ℓ_1 -theorem of 1974 [Ros74] states that every bounded sequence in a Banach space has a subsequence of one of two kinds. Such a subsequence is either weakly Cauchy or equivalent to the unit vector basis of ℓ_1 , the space of absolutely summable sequences. At the core of the proof of the ℓ_1 -theorem is a dichotomy for bounded sequences of real functions on a set. Such a sequence either has a pointwise convergent subsequence or has an *independent* subsequence, one that admits a lower and an upper threshold in the following sense. For any two disjoint finite sets of indices, at some point the functions indexed by the first fall below the lower threshold and those indexed by the second exceed the upper one. Bourgain, Fremlin and Talagrand [BFT78] passed in 1978 from sequences to sets, and from convergence to the first Baire class, which consists of the pointwise limits of sequences of continuous functions. By a condition of the same kind, they characterized the pointwise bounded sets of continuous functions on a Polish space whose pointwise closures lie in the first Baire class (see Theorem 1.4). Recall that a Polish space is a separable and completely metrizable space. On a compact

metrizable space, this condition is exactly the absence of the independence property for the set of functions (see Theorem 1.6). Compact sets of functions of the first Baire class on Polish spaces, with the topology of pointwise convergence, were investigated by Rosenthal in 1977 [Ros77]. Such compacta were named *Rosenthal compacta* after him by Godefroy in 1980 [God80], and have since been studied intensively in functional analysis, especially in the theory of Banach spaces.

The classification of Rosenthal compacta took a decisive form in 1999, when Todorćević [Tod99] proved his trichotomy. By this trichotomy, every separable Rosenthal compactum that is not metrizable contains a copy of one of a few canonical compacta. Which of the canonical compacta occur depends on whether the compactum is hereditarily separable, first countable but not hereditarily separable, or not first countable (see Theorem 3.3). Todorćević’s proof relies on set-theoretic forcing and infinite Ramsey theory, and his methods continue to be analyzed and exploited [Tod23, HT19]. Glasner and Megrelishvili [GM22] used the trichotomy to arrange into a hierarchy the tame dynamical systems on compact metrizable spaces [Köh95, Gla06]. The tame systems on such spaces are those whose enveloping semigroup, the pointwise closure of the set of maps of the system, consists of maps of the first Baire class. The levels of the Glasner–Megrelishvili hierarchy are defined by first countability, hereditary separability and metrizability of the enveloping semigroup.

The theories just recalled are the instruments with which we classify the unstable sets of computations. An unstable set of computations is *wild* when it has Shelah’s *independence property* (IP), and *tame*, or *NIP*, when it does not. The independence property means, roughly, that for some feature the computations of the set realize every finite pattern of high and low values on suitable inputs. Between stability and NIP we identify three new levels of tameness, NIP_3 , NIP_2 and NIP_1 , which together with NIP form a hierarchy of four tame classes beyond the stable ones. In this classification, the theory of Rosenthal compacta assumes on the topological side the role that Grothendieck’s criterion plays for stable sets. On the model-theoretic side of this classification, the role of stability passes correspondingly to NIP, a central notion of *neostability*, the part of classification theory beyond stability.

Model theory is the medium through which topology and computation are translated into each other. With every intermediate state of a computation we associate its *type*, the record of the real values that all the features of the computation take at that state. The types and their limits form a space of real-valued functions on the set of features, the *space of types*, on which computations act. The embedding of a structure into its space of types is a central device of model theory. This embedding allows us to study topological limits of computations through the theory of function spaces with the topology of pointwise convergence, known as C_p -theory. The space of types thus becomes a “Rosetta stone” connecting three classification programs. In topology, the first of them is the classification of Rosenthal compacta pioneered by Todorćević [Tod99]; in logic, the second is the classification of theories developed by Shelah [She71, She90]. In statistical learning, the third is the theory of Vapnik–Chervonenkis (VC) dimension and of PAC (probably approximately correct) learnability [VC71, VC74, Val84]. In Theorem 3.1 we show that, under this translation, NIP sets of computations correspond to spaces of deep computations that are Rosenthal compacta on each type shard. On the side of learning, finiteness

of VC dimension implies NIP, and the two notions coincide for first-order formulas (see Section 1.5).

Argyros, Dodos and Kanellopoulos [ADK08] refined Todorćević's trichotomy into a heptachotomy, which we likewise bring to bear on deep computations. The objects of the heptachotomy are families of functions indexed by the binary tree, compared along dyadic subtrees, the copies of the binary tree inside itself. Two such families are *equivalent* when the bijection between them that respects the indexing extends to a homeomorphism of their closures. Argyros, Dodos and Kanellopoulos exhibited seven pairwise nonequivalent families of Baire class 1 functions indexed by the binary tree. Each of these seven families is *minimal*, in the sense that it is equivalent to each of its subfamilies indexed by a dyadic subtree. The same three authors proved the heptachotomy for relatively compact families of Baire class 1 functions on a Polish space, indexed by the binary tree. This theorem asserts that every such family contains, along a dyadic subtree, a family equivalent to one of the seven.

Approximation by sequences rests on a property shared by two settings, that of Grothendieck [Gro52] and that of Bourgain, Fremlin and Talagrand [BFT78]. In the first, the functions are continuous on a compact space, whereas in the second they are of Baire class 1 on a Polish space. In both settings, *compactness and sequential compactness coincide* for sets of such functions with the topology of pointwise convergence. In either setting, moreover, whenever the closure of such a set is compact and consists of such functions, every point of the closure is the limit of a sequence from the set. Consequently, when the features are countably many and the set of computations is NIP, every deep computation is, on each shard, the limit of a *sequence* of computations.

Our results show how the classification invariants used in computation, topology, descriptive set theory and model theory translate into one another. The paper presupposes no acquaintance with advanced topology, model theory or the theory of computation; undergraduate topology and measure theory suffice, and the further background is assembled in Section 1.

Our first main result, Theorem 3.1, concerns compositional computation structures with countably many features. In it we characterize the sets of computations that have the NIP on every shard that they preserve, one feature at a time. These are exactly the sets whose deep computations, extended to the limit states, form a Rosenthal compactum on each of the corresponding type shards. Equivalently, on each of these type shards the extended deep computations are of the first Baire class. Guided by Todorćević's trichotomy, we then define three levels of tameness by the topology of these compacta (see Definition 3.2). A countable NIP set of computations is NIP_3 , NIP_2 or NIP_1 when its deep computations form respectively a metrizable, a hereditarily separable or a first countable space on each such type shard. For countable sets of computations in compositional computation structures with countably many features, the levels form the hierarchy

$$\text{stable} \Rightarrow \text{NIP}_3 \Rightarrow \text{NIP}_2 \Rightarrow \text{NIP}_1 \Rightarrow \text{NIP},$$

and in Section 5 we show that every implication in it is strict.

In that section we realize each level by a classical algorithm and exhibit algorithms beyond the hierarchy:

- A rotation of the circle is stable.
- The power method is NIP_3 but not stable.

- Bresenham’s line algorithm occupies a level that depends on the dimension of the space in which it draws lines:
 - For lines in the plane, it is NIP_2 but not NIP_3 .
 - For lines in three-dimensional space, it is NIP_1 but not NIP_2 . In that case, the deep computations of the algorithm contain a copy of the extended duplicate of the split Cantor set but none of the Alexandroff duplicate of the Cantor set.
- The Euclidean algorithm, acting on the projective line by the group that its steps generate, is NIP but not NIP_1 , and its deep computations include exact equality tests. Moreover, a countable family of compositions of Euclidean steps and their inverses is NIP_1 but not NIP_2 , and its deep computations contain a copy of the Alexandroff duplicate of the Cantor set.
- The logistic map $x \mapsto rx(1-x)$ on $[0, 1]$ is stable for $0 < r \leq 1$, and it is NIP_3 but not stable for $1 < r < r_\infty$, where $r_\infty \approx 3.5699$ is the Feigenbaum parameter. The map is not NIP for $r_\infty < r \leq 4$, whereas at r_∞ itself, as we sketch, it is again NIP_3 but not stable.
- Newton’s method on the Riemann sphere is not NIP , for the roots of $z^2 - 1$ as well as for those of the cubics $z^3 - 2z + 2$ and $z^3 - 1$.

For iterative algorithms, the hierarchy coincides with the Glasner–Megrelishvili hierarchy of tame dynamical systems [GM22]. Our two examples at the level NIP but not NIP_1 are the Euclidean algorithm and free reduction in the free group on two generators. Both are actions of groups with two generators rather than iterations of a single map. Whether a single iterated map can reach this level is an open question, closely related to a question of Glasner and Megrelishvili (see Question 7.2). For separable Rosenthal compacta that are first countable but not hereditarily separable, Todorćević’s trichotomy allows two alternatives. At the level NIP_1 but not NIP_2 , the three-dimensional Bresenham algorithm and the countable family of compositions of Euclidean steps and their inverses realize these two alternatives. Whether a single iterated map can realize this level with a copy of the Alexandroff duplicate among its deep computations is likewise open (see Question 7.3).

The Argyros–Dodos–Kanellopoulos heptachotomy refines this picture for families of computations indexed by the binary tree. On each type shard and for each feature, every tame family of this kind contains, along a dyadic subtree, a family equivalent to one of the seven minimal families. We interpret the seven minimal families as primitives of binary search (see Section 3.3). Finally, in Section 6 we replace topology by measure, which frees the theory from the countability of the set of features. In Theorem 6.7 we show that a set of computations is NIP exactly when the restrictions of all its ultracomputations to the type shards are *universally measurable*. Universal measurability here means measurability with respect to every Radon probability measure. In that case, the deep computations of the set are *universally Monte Carlo computable*. Whatever Radon probability measure governs the inputs, each output feature of such a deep computation is, outside a set of arbitrarily small probability, a uniform limit of polynomials in the input features. We also relate NIP to *Talagrand stability*, a measure-theoretic condition introduced by Talagrand in his study of Pettis integration. Talagrand stability likewise yields universal Monte Carlo computability (see Corollary 6.15), and for countable sets of computations we show that the two conditions coincide (see Corollary 6.19).

Organization of the paper. In Section 1 we present the topological and combinatorial preliminaries, including five canonical Rosenthal compacta (Section 1.4) and the role of NIP in model theory, combinatorics and learning (Section 1.5). In Section 2 we recall compositional computation structures, the source of the structural and model-theoretic viewpoint of the paper, and their stable sets of computations. That section ends with two motivating examples, Newton’s method for square roots and Newton’s method on the Riemann sphere, whose different behaviors the classification is designed to separate (Section 2.5). Section 3 contains the classification of deep computations and the main theorems. In Section 4 we work out Newton’s method on the Riemann sphere and on the real projective line. Next, in Section 5, we place classical algorithms at every level of the hierarchy and beyond it. In Section 6 we replace topology by measure and study deep computations that are universally measurable on each type shard. The output features of such deep computations are therefore, outside sets of arbitrarily small probability, uniform limits of polynomials in the input features. We intend to develop this measure-theoretic viewpoint further in future research. Section 7 contains the questions that remain open. In Appendix A we prove that the shift on the Thue–Morse subshift generates an action of $\mathbb{Z}$ that is not tame, a fact used in Section 5.6.

CONTENTS

Introduction	2
Organization of the paper	7
1. Topological preliminaries:	
From continuity to Baire class 1	8
1.1. From Rosenthal’s dichotomy to the Bourgain–Fremlin–Talagrand dichotomy to Shelah’s NIP	10
1.2. Rosenthal compacta	12
1.3. The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with $\mathcal{P}$ countable	12
1.4. Five canonical Rosenthal compacta	14
1.5. NIP in model theory, combinatorics and learning: polynomial versus exponential growth	17
Historical remarks	18
2. Compositional Computation Structures:	
A structural approach to arbitrary-precision computations	19
2.1. Computation States Structures	19
2.2. Compositional Computation Structures	21
2.3. Computability of deep computations and the Extendibility Axiom	22
2.4. Stability	23
2.5. Two motivating examples	24
3. Classifying deep computations	24
3.1. NIP, Rosenthal compacta, and deep computations	25
3.2. Todorćević’s trichotomy and levels of NIP	26
3.3. The Argyros–Dodos–Kanellopoulos heptachotomy and binary search	28
4. Examples of Compositional Computation Structures	30
4.1. Example: The Newton CCS for the cubic $q(z) = z^3 - 2z + 2$	31
4.2. Example: The Newton CCS for the cubic $u(z) = z^3 - 1$	32
4.3. Example: Square roots	33
5. Computations across the hierarchy	34

5.1. Iterative algorithms and tame dynamics	34
5.2. Stable and metrizable	36
5.3. Hereditarily separable but not metrizable	38
5.4. First countable but not hereditarily separable	40
5.5. NIP but not first countable: the Euclidean algorithm	42
5.6. Beyond NIP	49
5.7. Summary	54
6. Randomized versions of NIP and Monte Carlo computability of deep computations	54
6.1. NIP and Monte Carlo computability of deep computations	56
6.2. Talagrand stability and Monte Carlo computability of deep computations	59
7. Open questions	62
Appendix A. The Thue–Morse system is not tame	63
References	66

1. TOPOLOGICAL PRELIMINARIES: FROM CONTINUITY TO BAIRE CLASS 1

In this section we assemble the topology of function spaces on which the rest of the paper rests. The two themes of this paper are the approximation of a deep computation by a sequence of computations and the complexity of the set of all deep computations. Since deep computations are pointwise limits of continuous maps along ultrafilters, both themes belong to the theory of spaces of functions with the topology of pointwise convergence. We first recall Polish spaces, Baire category and the first Baire class; we then turn to Rosenthal’s dichotomy and the Bourgain–Fremlin–Talagrand dichotomy, from which the NIP emerges. The Rosenthal compacta come next, and five of them serve as yardsticks for the classification. The section closes with the role of the NIP in model theory, in combinatorics and in statistical learning. All spaces are assumed to be Hausdorff unless explicitly stated otherwise.

We write $\mathbb{N}$ for the set $\{0, 1, 2, \dots\}$ of natural numbers, so that the semigroups of iterates considered below contain the identity. A subset of a topological space is F_σ if it is a countable union of closed sets, and G_δ if it is a countable intersection of open sets. In a metric space every open set is F_σ , and, equivalently, every closed set is G_δ .

A *Polish space* is a separable topological space whose topology is induced by a complete metric. Countable products of Polish spaces, with the product topology, are Polish and furnish classical examples besides the real line. They include the Cantor space $2^\mathbb{N}$ of all infinite binary sequences, the Baire space $\mathbb{N}^\mathbb{N}$ of all infinite sequences of natural numbers, and the space $\mathbb{R}^\mathbb{N}$ of all sequences of real numbers.

Fact 1.1. [Eng89, 4.3.24] *A subset of a Polish space is itself Polish in the subspace topology if and only if it is a G_δ set. In particular, closed subsets and open subsets of Polish spaces are also Polish spaces.*

A subset A of a topological space X is *nowhere dense* if its closure has empty interior. The set A is of the *first category*, or *meager*, if it is contained in a countable

union of closed nowhere dense sets, and of the *second category* otherwise. The complements of meager sets are called *comeager*. Since the closure of a nowhere dense set is nowhere dense, the meager subsets of X form the σ -ideal generated by the nowhere dense subsets of X . A topological space X is *Baire* if every comeager subset of X is dense, or, equivalently, if every countable intersection of dense open sets is dense. The Baire category theorem states that compact Hausdorff spaces and completely metrizable spaces, Polish spaces in particular, are Baire.¹

For topological spaces X and Y , we denote by $C_p(X, Y)$ the set of all continuous functions $f: X \rightarrow Y$ with the topology of pointwise convergence, and we abbreviate $C_p(X, \mathbb{R})$ to $C_p(X)$. The topology of pointwise convergence is the one that $C_p(X, Y)$ inherits as a subspace of the product Y^X , in which X serves merely as an index set.

A function $f: X \rightarrow Y$ between topological spaces is of *Baire class 1* if it is the pointwise limit of a sequence $(f_n)_{n \in \mathbb{N}}$ of continuous functions, that is, $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ for all $x \in X$. When Y is metrizable, an elementary argument shows that a function of Baire class 1 is continuous at every point outside a set of the first category. We denote by $B_1(X, Y)$ the set of functions $X \rightarrow Y$ of Baire class 1 with the topology of pointwise convergence, and we write $B_1(X)$ for $B_1(X, \mathbb{R})$; the inclusions $C_p(X, Y) \subseteq B_1(X, Y) \subseteq Y^X$ are topological. Relatively compact subsets of $C_p(X)$ and of $B_1(X)$, especially for Polish X , are central objects of study in functional analysis and in topological dynamics.

A topological space X is *perfectly normal* if it is normal and every closed subset of X is a G_δ set (equivalently, every open subset of X is an F_σ set). Every metrizable space (hence, every Polish space) is perfectly normal.

The following characterization of Baire class 1 functions is classical; for a proof, see [Tod97, Section 10].

Fact 1.2 (Lebesgue, Hausdorff, Baire²). *If X is perfectly normal, then the following conditions are equivalent for a function $f: X \rightarrow \mathbb{R}$:*

- (1) *The function f is of Baire class 1, that is, it is a pointwise limit of continuous functions.*
- (2) *The preimage $f^{-1}[U]$ is an F_σ subset of X whenever $U \subseteq \mathbb{R}$ is open.*

If X is Polish, then (1) and (2) are equivalent to:

- (3) *For every nonempty closed $F \subseteq X$, the restriction $f|_F$ has a point of continuity.*

Moreover, if X is Polish and $f \notin B_1(X)$, then there exist countable $D_0, D_1 \subseteq X$ and reals $a < b$ such that

$$D_0 \subseteq f^{-1}(-\infty, a], \quad D_1 \subseteq f^{-1}[b, \infty), \quad \overline{D_0} = \overline{D_1}.$$

¹More generally, any Čech-complete space is Baire [Eng89, Theorem 3.9.3].

²The implication (1) $\Rightarrow$ (2) holds in every topological space. Indeed, if f is the pointwise limit of a sequence $(f_n)_{n \in \mathbb{N}}$ of continuous functions and a is a real number, then

$$f^{-1}(a, \infty) = \bigcup_{k \geq 1} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} f_n^{-1}[a + 1/k, \infty)$$

is a countable union of closed sets. The same is true of $f^{-1}(-\infty, a)$; since every open subset of $\mathbb{R}$ is a countable union of intervals $(a, b) = (a, \infty) \cap (-\infty, b)$, condition (2) follows. The reverse implication, (2) $\Rightarrow$ (1), was established by Lebesgue for real-valued functions of finitely many real variables [Leb04] and extended by Hausdorff to real-valued functions on metric spaces [Kec95, Theorem 24.10]. The proof of (2) $\Rightarrow$ (1) requires of X only that it be normal and that each of

1.1. From Rosenthal’s dichotomy to the Bourgain–Fremlin–Talagrand dichotomy to Shelah’s NIP. The question when a deep computation is accessible to approximation, that is, when it is the limit of a *sequence* of computations, is answered by Rosenthal’s dichotomy and the Bourgain–Fremlin–Talagrand dichotomy. We recall both dichotomies in this subsection; from them the NIP emerges in the theory of function spaces. In a metrizable space every point of the closure of a set is the limit of a sequence of points of the set. In spaces of functions with the topology of pointwise convergence, by contrast, the closure of a set need not admit this sequential description, and the two dichotomies determine when it does not.

On the Cantor space $X = 2^{\mathbb{N}}$, the n -th coordinate map $\pi_n: X \rightarrow \{0, 1\}$ is continuous for each $n \in \mathbb{N}$; however, the sequence $(\pi_n)_{n \in \mathbb{N}}$ has *no* convergent subsequence in $\mathbb{R}^X$ (see [Tod97, Chapter 1.1]). In a sense, this sequence exhibits the most extreme failure of sequential convergence. Indeed, the closure of $\{\pi_n : n \in \mathbb{N}\}$ in $\{0, 1\}^X$, or in $\mathbb{R}^X$, is homeomorphic to the Stone–Čech compactification $\beta\mathbb{N}$ of the discrete space of natural numbers. The space $\beta\mathbb{N}$ has cardinality $2^{2^{\aleph_0}}$, the largest possible for a separable Hausdorff space, and its points correspond to the ultrafilters on $\mathbb{N}$.³

In 1974 Haskell Rosenthal proved the following dichotomy for pointwise bounded sequences of continuous functions on a Polish space. His theorem shows that for these sequences a copy of $\beta\mathbb{N}$, such as the closure of the coordinate maps, is the only obstruction to the existence of a convergent subsequence.

Theorem 1.3 (Rosenthal’s dichotomy [Ros74]). *If X is a Polish space and $(f_n)_{n \in \mathbb{N}}$ is a pointwise bounded sequence in $C_p(X)$, then the sequence (f_n) has either a convergent subsequence or a subsequence whose closure in $\mathbb{R}^X$ is homeomorphic to $\beta\mathbb{N}$.*

Rosenthal’s dichotomy grew out of *Rosenthal’s ℓ_1 -theorem*, in whose statement ℓ_1 denotes the space of absolutely summable sequences. This theorem states that every bounded sequence in a Banach space has a subsequence that is either weakly Cauchy or equivalent to the unit vector basis of ℓ_1 . In particular, a Banach space contains no isomorphic copy of ℓ_1 if and only if every bounded sequence in it has a weakly Cauchy subsequence.

The Bourgain–Fremlin–Talagrand theorem, which concerns sets rather than sequences, characterizes the relative compactness in the larger space $B_1(X)$ of a pointwise bounded set of continuous functions on a Polish space X . For such a set, relative compactness in $B_1(X)$ is the property that every function in its pointwise

its open subsets be an F_σ set. These two requirements are precisely the defining properties of a perfectly normal space. The implication (1) $\Rightarrow$ (3) holds whenever X is Polish. If f satisfies (1) and F is a nonempty closed subset of X , then the restriction $f|_F$ is of Baire class 1. As noted before Fact 1.2, the points of discontinuity of $f|_F$ therefore form a meager subset of F . Since F is closed in a Polish space, it is itself Polish, hence a Baire space that this meager set cannot exhaust. Consequently, the restriction $f|_F$ has a point of continuity. The reverse implication (3) $\Rightarrow$ (1) is due to Baire, who proved it in his thesis for real-valued functions of real variables [Bai99]; for Polish spaces, see [Kec95, Theorem 24.15]. Finally, the last assertion of Fact 1.2 follows from the implication (3) $\Rightarrow$ (1) together with the Baire category theorem.

³Addition on $\mathbb{N}$ extends to an associative operation on $\beta\mathbb{N}$, whose idempotent elements, the idempotent ultrafilters, exist by the Ellis–Numakura lemma. Alva, Dueñez, Iovino and Walton [ADIW26] use idempotent ultrafilters to prove the existence of *deep equilibria* in compositional computation structures. Deep equilibria are limits of the iterates of a single computation, taken along an ultrafilter on $\mathbb{N}$, that are idempotent under composition.

closure be of Baire class 1. This theorem is a counterpart of Rosenthal's dichotomy, although it is usually stated as an equivalence rather than as a dichotomy (see Theorem 1.4 below).

The statement of the theorem requires the following notation. For $f \in \mathbb{R}^X$ and a real number a , define

$$X_{\leq a}^f := f^{-1}(-\infty, a], \quad X_{\geq a}^f := f^{-1}[a, +\infty).$$

For a set $A \subseteq \mathbb{R}^X$ and a real number a , define, by a slight abuse of notation,

$$X_{\leq a}^A := \bigcap_{f \in A} X_{\leq a}^f = \{x \in X : f(x) \leq a \text{ for all } f \in A\},$$

$$X_{\geq a}^A := \bigcap_{f \in A} X_{\geq a}^f = \{x \in X : f(x) \geq a \text{ for all } f \in A\};$$

in particular, we have $X_{\geq a}^\emptyset = X = X_{\leq a}^\emptyset$. For a sequence $(f_n)_{n \in \mathbb{N}}$ in $\mathbb{R}^X$ and a set $I \subseteq \mathbb{N}$, define $I^\complement := \mathbb{N} \setminus I$ and $f_I := \{f_i : i \in I\}$.

Theorem 1.4 (The Bourgain–Fremlin–Talagrand dichotomy [BFT78]). *Let X be a Polish space, and let $A \subseteq C_p(X)$ be pointwise bounded. The following are equivalent:*

- (i) *The set A is relatively compact in $B_1(X)$.*
- (ii) *For every $(f_n)_{n \in \mathbb{N}} \subseteq A$ and every $a < b$ there is $I \subseteq \mathbb{N}$ such that*

$$X_{\leq a}^{f_I} \cap X_{\geq b}^{f_{I^\complement}} = \emptyset.$$

(In this form, the Bourgain–Fremlin–Talagrand dichotomy is a special case of the equivalence (ii) $\Leftrightarrow$ (v) of Corollary 4G of Bourgain, Fremlin and Talagrand [BFT78].)

In condition (ii) of Theorem 1.4, the set $X_{\leq a}^{f_I}$ is defined by the inequalities $f_i(x) \leq a$ with $i \in I$, and the set $X_{\geq b}^{f_{I^\complement}}$ by the inequalities $f_j(x) \geq b$ with $j \in I^\complement$. Since I or $I^\complement$ is infinite, at least one of the two sets is defined by countably infinitely many inequalities. A computation, however, can check only finitely many inequalities at a time, and the following definitions accordingly impose only finitely many.

Definition 1.5. A set $A \subseteq \mathbb{R}^X$ of functions has the *finitary No-Independence Property (NIP)* if, for every sequence $(f_n)_{n \in \mathbb{N}}$ in A and all reals $a < b$, there exist finite disjoint sets $I, J \subseteq \mathbb{N}$ such that $X_{\leq a}^{f_I} \cap X_{\geq b}^{f_J} = \emptyset$. We say that finite disjoint sets I, J with this property *witness finitary NIP* for (f_n) and a, b .

A set $A \subseteq \mathbb{R}^X$ has the *finitary Independence Property (IP)* if it does not have finitary NIP. Explicitly, finitary IP means that there exist a sequence $(f_n)_{n \in \mathbb{N}}$ in A and reals $a < b$ such that $X_{\leq a}^{f_I} \cap X_{\geq b}^{f_J} \neq \emptyset$ for every pair of finite disjoint sets $I, J \subseteq \mathbb{N}$.

If the word “finite” is omitted in these definitions, we obtain *countable* NIP and *countable* IP. Countable NIP is weaker than finitary NIP, and countable IP is accordingly stronger than finitary IP. If, in addition, the witnesses are required to satisfy $J = I^\complement$, we obtain the properties that we call *BFT-NIP* and *BFT-IP*. Theorem 1.4 thus states that, on a Polish space X , a pointwise bounded set $A \subseteq C_p(X)$ is relatively compact in $B_1(X)$ if and only if it has BFT-NIP.

From now on, IP and NIP mean *finitary* IP and NIP unless otherwise specified.

Theorem 1.6. *Let X be a compact metrizable (hence Polish) space. For every pointwise bounded set $A \subseteq C_p(X)$, the following are equivalent:*

- (i) the set A is relatively compact in $B_1(X)$;
- (ii) the set A has BFT-NIP;
- (iii) the set A has countable NIP;
- (iv) the set A has finitary NIP.

(The equivalence $(ii) \Leftrightarrow (iv)$ holds for every compact space X , even without the assumption that A is pointwise bounded, and the equivalence $(ii) \Leftrightarrow (iii)$ holds for every set $A \subseteq \mathbb{R}^X$ on an arbitrary space X .)

Proof. The equivalence $(i) \Leftrightarrow (ii)$ is Theorem 1.4. The implications $(iv) \Rightarrow (iii)$ and $(ii) \Rightarrow (iii)$ are trivial. The converse $(iii) \Rightarrow (ii)$ holds because witnesses I and J of countable NIP satisfy $J \subseteq I^{\mathbb{C}}$, whence I and $I^{\mathbb{C}}$ witness BFT-NIP. To prove $(ii) \Rightarrow (iv)$, suppose that (f_n) and $a < b$ witness finitary IP, and let $I \subseteq \mathbb{N}$ be given. The closed sets $X_{\leq a}^{f_n}$ ($n \in I$) and $X_{\geq b}^{f_m}$ ($m \in I^{\mathbb{C}}$) have the finite intersection property; by compactness, therefore, we obtain $X_{\leq a}^{f_I} \cap X_{\geq b}^{f_{I^{\mathbb{C}}}} \neq \emptyset$. Since I is arbitrary, the set A does not have BFT-NIP, and (ii) fails. $\square$

Theorem 1.6 can be cast as a dichotomy: on a compact metrizable space X , every pointwise bounded set $A \subseteq C_p(X)$ either is relatively compact in $B_1(X)$ or has IP in the finitary and, equivalently, in the countable sense.

The independence property was isolated by Saharon Shelah [She71, Definition 4.1] as one of the dividing lines of his classification of first-order theories [She90]. We return to it in Section 1.5, where it appears together with its counterparts in combinatorics and in statistical learning.

1.2. Rosenthal compacta. The Bourgain–Fremlin–Talagrand dichotomy leads to the Rosenthal compacta, the compact spaces within which the deep computations of this paper are classified, and we recall them here. By Theorem 1.4, if a pointwise bounded set of continuous functions on a Polish space has BFT-NIP, then its pointwise closure is a compact set of functions of the first Baire class. Compact sets of this kind had appeared in Rosenthal’s work on the ℓ_1 -theorem, and in 1977 he made them the subject of a study of their own [Ros77]. Godefroy [God80] named them after Rosenthal in the following definition.

Definition 1.7. A *Rosenthal compactum* is a topological space homeomorphic to a compact subset of $B_1(X)$, with the topology of pointwise convergence, for some Polish space X .

Rosenthal compacta have strong tameness properties; for instance, every point in the closure of a subset of a Rosenthal compactum is the limit of a sequence of points of the subset [BFT78, Theorem 3F]. In this paper Rosenthal compacta serve as spaces of deep computations. Theorem 3.1 below shows that for countably many features the deep computations of an NIP set of computations, extended to the limit states, form a Rosenthal compactum on each type shard. Here a type shard is a compact set of states and limit states on which every feature obeys a prescribed bound (see Section 2.1.1).

1.3. The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with $\mathcal{P}$ countable. This subsection is devoted to reducing the study of maps with values in a product of copies of the real line to that of real-valued functions. The reduction is needed because a state of a computation is described by the values of its features. A computation is therefore a map

with values in the product of copies of the real line indexed by the features. Fix a countable set $\mathcal{P}$ (in this paper, “countable” means finite or countably infinite), whose elements we call *predicate symbols* or *formal predicates*. Our goal is to characterize the relatively compact subsets of $B_1(X, \mathbb{R}^{\mathcal{P}})$, where X is a perfectly normal space, typically a Polish space.

Whenever $\mathcal{P}$ is regarded as a topological space, it carries the discrete topology, and the product $X \times \mathcal{P}$ is regarded either as a set or as a topological product, depending on the context. The spaces $C_p(X, \mathbb{R}^{\mathcal{P}}) \subseteq B_1(X, \mathbb{R}^{\mathcal{P}})$ are subspaces of the ambient space $(\mathbb{R}^{\mathcal{P}})^X$, which can be described in two other ways, through the natural homeomorphisms

$$(\mathbb{R}^{\mathcal{P}})^X \cong \mathbb{R}^{X \times \mathcal{P}} \cong (\mathbb{R}^X)^{\mathcal{P}},$$

given by

$$\begin{aligned} \mathbb{R}^{X \times \mathcal{P}} &\rightarrow (\mathbb{R}^{\mathcal{P}})^X : \Phi \mapsto \check{\Phi} \\ \mathbb{R}^{X \times \mathcal{P}} &\rightarrow (\mathbb{R}^X)^{\mathcal{P}} : \Phi \mapsto \Phi^\wedge, \end{aligned}$$

where

$$\check{\Phi}(x) := \Phi(x, \cdot) \in \mathbb{R}^{\mathcal{P}}, \quad \Phi^\wedge(P) := \Phi(\cdot, P) \in \mathbb{R}^X.$$

These identifications treat X and $\mathcal{P}$ as mere sets; in particular, the topology of X plays no role in them.

The following proposition reduces the study of continuous functions, and of functions of Baire class 1, from X to $\mathbb{R}^{\mathcal{P}}$ to the case of real-valued functions.

Proposition 1.8. *The identification $(\mathbb{R}^{\mathcal{P}})^X \cong \mathbb{R}^{X \times \mathcal{P}} \cong (\mathbb{R}^X)^{\mathcal{P}}$ induces identifications*

$$C_p(X, \mathbb{R}^{\mathcal{P}}) \cong C_p(X \times \mathcal{P}) \cong C_p(X)^{\mathcal{P}}, \quad B_1(X, \mathbb{R}^{\mathcal{P}}) \cong B_1(X \times \mathcal{P}) \cong B_1(X)^{\mathcal{P}}.$$

(The cardinality of $\mathcal{P}$ plays no role here.)

Proof. The identification of the C_p -spaces follows from the definition of the product topology and the discreteness of $\mathcal{P}$. By the former, a continuous map $X \rightarrow \mathbb{R}^{\mathcal{P}}$ is precisely a $\mathcal{P}$ -indexed family of continuous functions $X \rightarrow \mathbb{R}$; by the latter, such families correspond to continuous functions $X \times \mathcal{P} \rightarrow \mathbb{R}$. For the spaces of functions of Baire class 1, recall that membership in the first Baire class is defined through pointwise limits of sequences of continuous functions. Since a sequence of $\mathcal{P}$ -indexed families of continuous functions is the same as a $\mathcal{P}$ -indexed family of sequences of continuous functions, the identification of these spaces follows as well. $\square$

For $A \subseteq Y^X$ and $K \subseteq X$, we write $A|_K := \{f|_K : f \in A\}$ for the set of restrictions to K of the functions in A . Similarly, for $A \subseteq (\mathbb{R}^{\mathcal{P}})^X$ and $P \in \mathcal{P}$, we write $\pi_P \circ A := \{\pi_P \circ f : f \in A\}$, where $\pi_P : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}$ is the projection onto the coordinate P . The following theorem extends Theorem 1.6 to $\mathbb{R}^{\mathcal{P}}$ -valued functions.

Theorem 1.9. *Assume that $\mathcal{P}$ is countable, X is a Polish space, and $A \subseteq C_p(X, \mathbb{R}^{\mathcal{P}})$ is pointwise bounded in the sense that the set $\pi_P \circ A \subseteq C_p(X)$ is pointwise bounded for every $P \in \mathcal{P}$. The following are equivalent for every compact $K \subseteq X$:*

- (i) *the set $A|_K$ is relatively compact in $B_1(K, \mathbb{R}^{\mathcal{P}})$;*
- (ii) *the set $\pi_P \circ A|_K$ has NIP for every $P \in \mathcal{P}$.*

Proof. By Proposition 1.8 and Tychonov's theorem, condition (i) holds if and only if $\pi_P \circ A|_K$ is relatively compact in $B_1(K)$ for every $P \in \mathcal{P}$. Since K is compact metrizable, Theorem 1.6 shows that this is equivalent to (ii). $\square$

In the applications, the NIP is verified on a set of states, whereas compactness is available only on the closure of that set. The following lemma shows that the NIP passes from a set to its closure.

Lemma 1.10. *Assume that X is Hausdorff and that $A \subseteq C_p(X)$. The following are equivalent for every $L \subseteq X$:*

- (i) *the set $A|_L$ satisfies the NIP;*
- (ii) *the set $A|_{\overline{L}}$ satisfies the NIP.*

Proof. The implication (ii) $\Rightarrow$ (i) is trivial. Conversely, suppose that a sequence (f_n) in A and reals $a < b$ witness the IP on $\overline{L}$, and choose a' and b' with $a < a' < b' < b$. For finite disjoint $I, J \subseteq \mathbb{N}$, the open set $\bigcap_{n \in I} f_n^{-1}(-\infty, a') \cap \bigcap_{n \in J} f_n^{-1}(b', \infty)$ contains a point of $\overline{L}$, namely a point at which $f_n \leq a$ for $n \in I$ and $f_n \geq b$ for $n \in J$; hence it meets L . Therefore the sequence (f_n) and the reals $a' < b'$ witness the IP on L . $\square$

1.4. Five canonical Rosenthal compacta. Here we fix the notation for five Rosenthal compacta, all built on the Cantor set, that are the canonical members of the classes of Todorćević's trichotomy (see Theorem 3.3 below). By the trichotomy, the space of deep computations of each non-metrizable NIP example of Section 5 contains a copy of one of these compacta. The five compacta are distinguished from one another by three topological properties. Recall that a topological space is *separable* if it has a countable dense subset and *hereditarily separable* if each of its subspaces is separable. Finally, a topological space is *first countable* if each of its points has a countable base of neighborhoods.

Let $2^{\mathbb{N}}$ carry the product topology and the lexicographic order $<$, whose order topology is the product topology. The set of finite binary words is denoted $2^{<\mathbb{N}}$, and the length of a word $t \in 2^{<\mathbb{N}}$ is denoted $|t|$. For $a \in 2^{\mathbb{N}}$ and $n \in \mathbb{N}$, we write $a \upharpoonright n$ for the word formed by the first n bits of a . Two words are *comparable* if one extends the other, and an *antichain* is a set of pairwise incomparable words. For a finite word $t \in 2^{<\mathbb{N}}$, let $t \frown 0^\infty$ and $t \frown 1^\infty$ be the sequences that begin with t and continue with all 0's and with all 1's, respectively. The set of sequences extending t is then the clopen interval $[t \frown 0^\infty, t \frown 1^\infty]$. Copies of the binary tree within itself index the dense subset of the last of the five compacta below, and the families of Section 3.3 are compared along these copies. These copies are the dyadic subtrees: a *dyadic subtree* of $2^{<\mathbb{N}}$ is a subset isomorphic to $2^{<\mathbb{N}}$ under the order of extension. We write a dyadic subtree as $\{w_t : t \in 2^{<\mathbb{N}}\}$, indexed through the canonical isomorphism, which preserves extension and the lexicographic order. A dyadic subtree is called *regular* if each of its levels is contained in a level of $2^{<\mathbb{N}}$. A sequence (u_k) of words *converges* to $a \in 2^{\mathbb{N}}$ if, for every $n \in \mathbb{N}$, all but finitely many of the words u_k extend $a \upharpoonright n$.

Double brackets denote truth values: $\llbracket \dots \rrbracket$ is 1 if the statement inside the brackets is true and 0 if it is false. We introduce the following functions $2^{\mathbb{N}} \rightarrow \{0, 1\}$:

- the *prefix test* $v_t := \llbracket |\cdot| = t \rrbracket$, the characteristic function of the set of sequences extending a fixed word $t \in 2^{<\mathbb{N}}$;

- the *equality test* $\delta_a := \llbracket \cdot = a \rrbracket$, the characteristic function of the singleton $\{a\}$ for fixed $a \in 2^{\mathbb{N}}$;
- the *threshold tests* $\chi_{[a]} := \llbracket a \leq \cdot \rrbracket$ and $\chi_{(a)} := \llbracket a < \cdot \rrbracket$, the characteristic functions of the final intervals $\{x : a \leq x\}$ (closed) and $\{x : a < x\}$ (open), for fixed $a \in 2^{\mathbb{N}}$; the special case $\mathbf{0} := \chi_{(1^\infty)}$ is the identically zero function.

Prefix tests are continuous, and so are the threshold tests $\chi_{[t \frown 0^\infty]}$ and $\chi_{(t \frown 1^\infty)}$, since each of these functions depends only on the first $|t|$ bits of its argument. By contrast, equality tests and threshold tests at all other points are discontinuous functions of Baire class 1. All closures below are taken in the topology of pointwise convergence.

- (1) The *Alexandroff compactification* $A(2^{\mathbb{N}}) = \{\delta_a : a \in 2^{\mathbb{N}}\} \cup \{\mathbf{0}\}$. Each δ_a is isolated, and every neighborhood of $\mathbf{0}$ contains all but finitely many of the functions δ_a . Thus the Alexandroff compactification is neither separable nor first countable at $\mathbf{0}$.
- (2) The *extended Alexandroff compactification* $\hat{A}(2^{\mathbb{N}}) = \overline{\{v_t : t \in 2^{<\mathbb{N}}\}} = \{v_t : t \in 2^{<\mathbb{N}}\} \cup A(2^{\mathbb{N}})$. As $n \rightarrow \infty$, the prefix tests $v_{a \upharpoonright n}$ converge to δ_a ; along an antichain they converge to $\mathbf{0}$. Hence the extended Alexandroff compactification is separable but not first countable.
- (3) The *split Cantor set* $S(2^{\mathbb{N}}) = \{\chi_{[a]}, \chi_{(a)} : a \in 2^{\mathbb{N}}\}$. The continuous thresholds $\chi_{[t \frown 0^\infty]}$, together with the isolated point $\mathbf{0} = \chi_{(1^\infty)}$, form a countable dense subset: thresholds approaching a from below converge to $\chi_{[a]}$, and from above to $\chi_{(a)}$. It follows that the split Cantor set is separable, first countable and hereditarily separable, but not metrizable.
- (4) The *Alexandroff duplicate* $D(2^{\mathbb{N}}) = \{(\mathbf{0}, a), (\delta_a, a) : a \in 2^{\mathbb{N}}\} \subseteq A(2^{\mathbb{N}}) \times 2^{\mathbb{N}}$. The points (δ_a, a) are isolated, whereas every neighborhood of $(\mathbf{0}, a)$ contains, for all b in some neighborhood of a , the point $(\mathbf{0}, b)$ and, with finitely many exceptions, the point (δ_b, b) . Consequently, the Alexandroff duplicate is first countable but not separable; it is contained in the separable compactum $\hat{D}(2^{\mathbb{N}}) = \overline{\{(v_t, t \frown 0^\infty) : t \in 2^{<\mathbb{N}}\}} = D(2^{\mathbb{N}}) \cup \{(v_t, t \frown 0^\infty) : t \in 2^{<\mathbb{N}}\}$.
- (5) The *extended duplicate of the split Cantor set*. Following Argyros, Dodos and Kanellopoulos [ADK08, 4.3.7], we define it along a regular dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$, fixed once and for all. This subtree is required to satisfy $s_t \frown 0^\infty \neq s_{t'} \frown 0^\infty$ and $s_t \frown 1^\infty \neq s_{t'} \frown 1^\infty$ for all distinct $t, t' \in 2^{<\mathbb{N}}$, and the extended duplicate is then

$$\hat{D}(S(2^{\mathbb{N}})) = \overline{\{(v_{s_t}, \chi_{[s_t \frown 0^\infty]}) : t \in 2^{<\mathbb{N}}\}} \subseteq \hat{A}(2^{\mathbb{N}}) \times S(2^{\mathbb{N}}).$$

Call a sequence $a \in 2^{\mathbb{N}}$ a *branch* of the subtree if each of its initial segments $a \upharpoonright n$ is extended by some word s_t ; the branches form a perfect subset of $2^{\mathbb{N}}$. The compactum $\hat{D}(S(2^{\mathbb{N}}))$ consists of the isolated points $(v_{s_t}, \chi_{[s_t \frown 0^\infty]})$ and, for each branch a , of the point $(\delta_a, \chi_{[a]})$ and of at most two points with first coordinate $\mathbf{0}$ [ADK08, 4.3.7]. The point $(\mathbf{0}, \chi_{[a]})$ belongs to the compactum exactly when some sequence (u_k) of pairwise incomparable words of the subtree converges to a with $u_k \frown 0^\infty$ increasing in k . Likewise, the point $(\mathbf{0}, \chi_{(a)})$ belongs to the compactum exactly when some such sequence converges to a with $u_k \frown 0^\infty$ decreasing in k . The points $(\delta_a, \chi_{[a]})$ are reached along the branches. Indeed, if a is a branch and $u_1, u_2, \dots$ are the words of the subtree that are initial segments of a , listed by increasing length,

then the sequences $u_k \frown 0^\infty$ increase strictly to a . The points $(v_{u_k}, \chi_{[u_k \frown 0^\infty)})$ then converge to $(\delta_a, \chi_{[a)})$ as $k \rightarrow \infty$. As the closure of a countable set, the extended duplicate of the split Cantor set is separable. This compactum is also first countable, since it embeds into the Helly space of nondecreasing functions from $[0, 1]$ into itself, which is first countable [ADK08, 4.3.7]. For each branch a , the point $(\delta_a, \chi_{[a)})$ is the only point $(\delta_b, \chi_{[b)})$, with b a branch, whose first coordinate takes the value 1 at a . The points $(\delta_a, \chi_{[a)})$, with a a branch, therefore form a discrete subspace, which has the cardinality of the continuum because the branches form a perfect set. Hence the extended duplicate of the split Cantor set is not hereditarily separable.

Figure 1 shows the five compacta as patterns of convergence of words of the binary tree $2^{<\mathbb{N}}$.

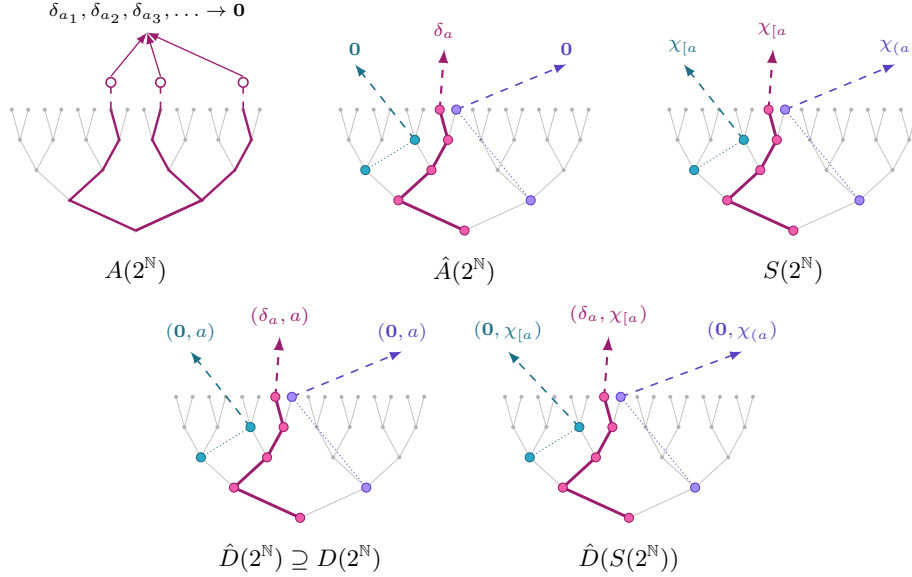


FIGURE 1. The five compacta of Section 1.4, drawn on the binary tree $2^{<\mathbb{N}}$. The tree grows upward, and the extension of a word by 0 lies to the left of its extension by 1. In the first tree, which stands for $A(2^{\mathbb{N}})$, the equality tests δ_{a_k} of distinct sequences $a_k \in 2^{\mathbb{N}}$ converge to the zero function $\mathbf{0}$. In the other trees, the thick path runs through the initial segments of a sequence $a \in 2^{\mathbb{N}}$, here $a = 0110 \dots$. The dots on the left of the path mark words t with $t \frown 0^\infty < a$, and those on the right mark words with $t \frown 0^\infty > a$. Both sequences of words converge to a , and the dashed arrows end at the limits. A word t stands for the prefix test v_t in $\hat{A}(2^{\mathbb{N}})$ and for the threshold test $\chi_{[t \frown 0^\infty)}$ in $S(2^{\mathbb{N}})$. In $\hat{D}(2^{\mathbb{N}})$, which contains the Alexandroff duplicate $D(2^{\mathbb{N}})$, the word t stands for the pair $(v_t, t \frown 0^\infty)$. In $\hat{D}(S(2^{\mathbb{N}}))$, the tree is the dyadic subtree of words s_t fixed in Section 1.4, and the node s_t stands for the pair $(v_{s_t}, \chi_{[s_t \frown 0^\infty)})$.

The constructions of the Alexandroff compactification and of the Alexandroff duplicate make sense with $2^{\mathbb{N}}$ replaced by any compact metric space K ; the resulting spaces are denoted $A(K)$ and $D(K)$, respectively. Since $A(K)$ is the one-point compactification of a discrete space of the cardinality of K , it depends only on that cardinality; for instance, $A([0, 1])$ is homeomorphic to $A(2^{\mathbb{N}})$.

1.5. NIP in model theory, combinatorics and learning: polynomial versus exponential growth. We now trace the NIP of Definition 1.5 to its counterparts in model theory, combinatorics and statistical learning, and we make precise how it is related to them. This relation is used in Section 5, where a bound on a Vapnik–Chervonenkis dimension yields the NIP of Bresenham’s algorithm. For families of $\{0, 1\}$ -valued functions (that is, of sets), a combinatorial dividing line closely related to the NIP appeared independently in three fields between 1968 and 1972, some years before Theorem 1.4. In model theory the line arose in the work of Saharon Shelah [She71, She90], and in combinatorics in the work of Norbert Sauer [Sau72] and, again, of Shelah [She72, She90]. In statistical learning, finally, the same line appeared in the work of Vladimir Vapnik and Alexey Chervonenkis [VC68, VC71, VC74]. In each of the three fields, the line separates the “tame” families of sets from the “wild” ones. The number of traces of a tame family on an n -element set (the intersections of the members of the family with the set) is bounded by a polynomial in n . In a wild family, by contrast, for every $n \in \mathbb{N}$ some n -element set has all 2^n possible traces.

In model theory: Among the dividing lines of Shelah’s classification theory [She90], the independence property is detected by counting types over finite sets of parameters. Let T be a complete first-order theory, and let $\varphi(x; y)$ be a formula of its language, in which x and y are finite tuples of variables of lengths $|x|$ and $|y|$. Finally, let B be a finite set of $|y|$ -tuples of elements of a model M of T . A φ -type over B is a maximal set of formulas $\varphi(x; b)$ and $\neg\varphi(x; b)$, with $b \in B$, that is consistent with the theory of M with constants for the elements of M . Since B is finite, every φ -type over B is realized by an $|x|$ -tuple a of elements of M , and the type is determined by the set $\{b \in B : M \models \varphi(a; b)\}$. Hence the set $S_\varphi(B)$ of φ -types over B is in bijection with the traces on B of the family of sets $\{b : M \models \varphi(a; b)\}$, with a ranging over $|x|$ -tuples of elements of M . This bijection gives the bound $|S_\varphi(B)| \leq 2^{|B|}$. The formula φ has the *independence property* (IP) if, for every $n \in \mathbb{N}$, some set B of n tuples satisfies $|S_\varphi(B)| = 2^n$. Otherwise the formula is *NIP*, or *dependent*, and the theory T is NIP if all its formulas are NIP. Since T is complete, these notions do not depend on the model M .

If φ is NIP, the Sauer–Shelah lemma below bounds $|S_\varphi(B)|$ by a polynomial in $|B|$. For a set A of parameters, we write $S_\varphi(A)$ for the set of φ -types over the set of $|y|$ -tuples of elements of A . If φ has IP, then for every n some set A of at most $n|y|$ parameters satisfies $|S_\varphi(A)| \geq 2^n$ [Sim15, Section 2.3.4]. Consequently, φ is NIP if and only if $|S_\varphi(A)|$ is bounded by a polynomial in $|A|$ as A ranges over the finite sets of parameters. Over infinite sets of parameters, counting φ -types does not separate NIP from IP without further set-theoretic hypotheses. Indeed, every unstable formula, NIP or not, already has 2^κ φ -types over some set of κ parameters whenever κ is infinite and $2^\kappa = \kappa^+$ [Sim15, Section 2.3.4].

In combinatorics: Sauer [Sau72] and Shelah [She72] proved independently the following dichotomy, now known as the *Sauer–Shelah lemma*. Let $\mathcal{F} = \{C_i : i \in I\}$ be a family of subsets of a set Ω , indexed by a set I . Then either for every $n \in \mathbb{N}$ there is a set $B \subseteq \Omega$ with $|B| = n$ such that $|\{C_i \cap B : i \in I\}| = 2^n$, or there exists $N \in \mathbb{N}$ such that, for every $B \subseteq \Omega$ with $|B| \geq N$,

$$|\{C_i \cap B : i \in I\}| \leq \sum_{j=0}^{N-1} \binom{|B|}{j}.$$

In the first case $\mathcal{F}$ is said to have “exponential complexity”, and in the second case “polynomial complexity”; every family of subsets of a *finite* set Ω has polynomial complexity. The name “polynomial complexity” is justified by the following estimate, valid for fixed $N > 0$ and for $|B| = m > 0$, which bounds the number of traces by a polynomial in m :

$$\sum_{j=0}^{N-1} \binom{m}{j} \leq \sum_{j=0}^{N-1} \frac{m^j}{j!} \leq \left(\sum_{j=0}^{N-1} \frac{1}{j!} \right) \cdot m^{N-1} < e \cdot m^{N-1}.$$

In machine learning: The Sauer–Shelah lemma is also the dichotomy that Vapnik and Chervonenkis had announced in 1968 [VC68], in their study of uniform convergence in statistics. Their proof of the dichotomy appeared in 1971 [VC71], a year before the papers of Sauer and Shelah (see also [VC74]). The largest size of a set $B \subseteq \Omega$ with $2^{|B|}$ traces $C_i \cap B$ is called the *VC dimension* of $\mathcal{F}$. Equivalently, the VC dimension is the integer $N-1$ for the least integer N as in the Sauer–Shelah lemma. If no such integer N exists, the VC dimension of $\mathcal{F}$ is infinite. Through the Vapnik–Chervonenkis theorem, the Sauer–Shelah lemma bounds the number of data points (the sample size) needed to “learn” a concept class $\mathcal{F}$ of finite VC dimension in a statistical sense. Here “learning” is understood up to a preset error and with a preset confidence. The Fundamental Theorem of Statistical Learning concerns PAC-learnability, where PAC stands for “Probably Approximately Correct” [Val84]. This theorem states that a class of $\{0, 1\}$ -valued functions satisfying a mild measurability condition is PAC-learnable if and only if its VC dimension is finite [BEHW89]; see also [SSBD14, Theorem 6.7].

For first-order formulas the three dichotomies coincide. To see this, fix a formula $\varphi(x; y)$ and consider the family of sets $\{b : M \models \varphi(a; b)\}$, with a ranging over the $|x|$ -tuples of elements of a model M of T . By the description of φ -types given above, the formula φ has IP exactly when this family has infinite VC dimension. For an NIP formula, moreover, the Sauer–Shelah lemma gives the polynomial bound on the number of φ -types. For an arbitrary family of sets, finiteness of the VC dimension implies that the characteristic functions of its members have the NIP of Definition 1.5. The NIP does not, however, imply finite VC dimension: the finite subsets of $\mathbb{N}$ form a family of infinite VC dimension whose characteristic functions have the NIP.

Historical remarks. The Baire hierarchy of functions was introduced by René-Louis Baire in his 1899 doctoral thesis, *Sur les fonctions de variables réelles*. The theory of spaces of continuous functions with the topology of pointwise convergence, C_p -theory, was pioneered by A. V. Arhangel’skiĭ and his students in the 1970s and

1980s [Ark92]. This theory has since found numerous applications in model theory and in functional analysis; Tkachuk’s problem book [Tka11] develops the subject systematically, through problems with complete solutions.

2. COMPOSITIONAL COMPUTATION STRUCTURES:

A STRUCTURAL APPROACH TO ARBITRARY-PRECISION COMPUTATIONS

The theory of function spaces assembled in Section 1 is brought to bear on computation through the framework of compositional computation structures, to which the present section is devoted. Within this framework, deep computations become points of a function space. A computation on real-valued data is observed through its features, each known to finite precision. A deep computation is a limit of computations along an ultrafilter in the topology of this observation, namely the topology of pointwise convergence. Following the first paper of this series [ADIW26], we first describe states by their types, the tuples of values of all their features. Computations act on these states and are composed; deep computations are the pointwise limits of computations along ultrafilters, taken in the space of types. Finally, we recall the stable sets of computations, studied in that paper, and close the section with two examples that motivate the classification of Section 3.

2.1. Computation States Structures. The first ingredient of the framework, which we introduce here, is the set of states on which computations act, together with the real-valued features through which the states are observed. A *computation states structure* is a pair $(L, \mathcal{P})$ consisting of a set L of *states* and a collection $\mathcal{P}$ of real-valued functions on L that we call *predicates* or *features*. Intuitively, the set L consists of the states of a computation, and the predicates are the primitives that are given and accepted as computable. For $P \in \mathcal{P}$ and $v \in L$, the number $P(v)$ is the value of the (observable) feature P at the state v . Maps $f: L \rightarrow L$ are called *transitions* of $(L, \mathcal{P})$; the computations of Definition 2.4 below are transitions.

The *type* of a state $v \in L$ is the indexed family⁴

$$\text{tp}(v) = (P(v))_{P \in \mathcal{P}} \in \mathbb{R}^{\mathcal{P}}$$

of its features.

We assume that each state $v \in L$ is determined by its type $\text{tp}(v)$. Under this assumption we may identify L with its image $\text{tp}[L] \subseteq \mathbb{R}^{\mathcal{P}}$ under the type map and endow L with the subspace topology. This topology is the topology of pointwise convergence induced by the product topology of $\mathbb{R}^{\mathcal{P}}$; in particular, for each $P \in \mathcal{P}$, the map $v \mapsto P(v)$ is continuous on L .

The principal examples are the state space $L = \mathbb{R}^{\mathbb{N}}$ with the predicates $P_i(v) = v_i$ for $i \in \mathbb{N}$ and, for a positive integer n , the state space $L = \mathbb{R}^n$ with the predicates $P_i(v) = v_i$ for $0 \leq i < n$. Their subspaces are the state spaces of primary interest in this paper. Indeed, through the identification of each state with its type, every

⁴As in Proposition 1.8, where $\mathcal{P}$ is merely a set of symbols, we regard each predicate as named by a formal label P , a *predicate symbol*. The meaning of this symbol in the computation states structure is the real function $P(\cdot): v \mapsto P(v)$ on L . The symbol P thus names an observable feature, whose value at a state $v \in L$ is $P(v)$. Whenever $\mathcal{P}$ is used as an index set, as in the product $\mathbb{R}^{\mathcal{P}}$ below, it is regarded as a set of symbols. The distinction between a symbol and its meaning is usually suppressed in mathematics, but the model-theoretic framework requires awareness of it. When necessary, we write $P(\cdot)$ for the function on L and P for the symbol; when no confusion can arise, we also write P for the function.

computation states structure with countably many predicates becomes a subspace of $\mathbb{R}^{\mathbb{N}}$ or of some $\mathbb{R}^n$ once its predicates are enumerated.

Definition 2.1. Let $(L, \mathcal{P})$ be a computation states structure. The type $\text{tp}(v) \in \mathbb{R}^{\mathcal{P}}$ of a state $v \in L$ is called a *realized* (state) type. The closure of the set of realized types in $\mathbb{R}^{\mathcal{P}}$, with the topology of pointwise convergence, is called the *space of (state) types* of $(L, \mathcal{P})$ and is denoted $\mathcal{L}$. Its elements $\mathbf{v} \in \mathcal{L}$ are called *state types*, and those in $\mathcal{L} \setminus L$ are called *unrealized* state types.

State types formalize the intuitive notion of a limit state, and the predicates extend to them continuously. Indeed, for each $P \in \mathcal{P}$, the restriction to $\mathcal{L}$ of the P -th coordinate map $\mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}$ given by $\mathbf{v} = (\mathbf{v}_Q)_{Q \in \mathcal{P}} \mapsto \mathbf{v}_P$ is a continuous extension of the predicate $P(\cdot)$ on L . Abusing notation, we still denote this extension by $P(\cdot)$, or just P .

Remark 2.2. One may regard unrealized state types as “hyper-states”, and those among them that are obtained as limits of sequences (or as ultralimits) of realized state types as “deep states”. For readers versed in model theory we note that, at least for realizing state types with parameters in L , the type space $(\mathcal{L}, \mathcal{P})$ serves as a “poor man’s” ultraproduct of $(L, \mathcal{P})$. Indeed, by the very definition of $\mathcal{L}$, every type $\mathbf{v}$ is realized in this type space.

2.1.1. Bounds on precision and magnitude. Shards. The analysis of this paper takes place on shards, the sets of states on which each feature obeys a prescribed bound; before defining them, we explain why actual computation imposes such bounds. Computation states structures are designed to model actual computation, in which operations on real-valued quantities can be carried out only approximately, to a preset degree of precision. A bound on the magnitude of the quantities involved is no less important in practice, for no machine can store, let alone compute with, arbitrarily large numbers. An actual computation therefore fixes at the outset bounds for its errors and for the absolute magnitudes of the quantities it handles. These bounds are fixed even when the collection $\mathcal{P}$ of predicates is infinite, possibly uncountable. Moreover, different parts of a computation may require different error bounds and may deal with quantities of very different sizes. Focusing on magnitudes, we consider sets of computations for which a bound for each predicate is fixed in advance. We are thereby led to the notion of *shard*, introduced in the first paper of this series [ADIW26].

Definition 2.3. Let $(L, \mathcal{P})$ be a computation states structure, with L identified with the set $\text{tp}[L] \subseteq \mathbb{R}^{\mathcal{P}}$ of its realized types. A *sizer* is a family $r_{\bullet} = (r_P)_{P \in \mathcal{P}}$ of positive real numbers indexed by $\mathcal{P}$. For a sizer $r_{\bullet}$, let $\mathbb{R}[r_{\bullet}] := \prod_{P \in \mathcal{P}} [-r_P, r_P]$, a compact space, and define the $r_{\bullet}$ -*shard* of the state space L by

$$L[r_{\bullet}] = L \cap \mathbb{R}[r_{\bullet}].$$

The $r_{\bullet}$ -*type shard* is the closure $\mathcal{L}[r_{\bullet}] = \overline{L[r_{\bullet}]}$, a closed and hence compact subset of $\mathbb{R}[r_{\bullet}]$.

The union of all type shards, as the sizer $r_{\bullet}$ varies, is denoted $\mathcal{L}_{\text{sh}}$ and called the space of *shard-supported types*.

We have $\mathcal{L}_{\text{sh}} \subseteq \mathcal{L}$, and the inclusion may be proper; equality holds in the important special case when $\mathcal{P}$ is countable (see [ADIW26]).

2.2. Compositional Computation Structures. In this subsection we recall the compositional computation structures of the first paper of this series [ADIW26], within which deep computations are defined, beginning with the ingredients of their definition. Computations transform states, and two computations performed one after the other form a computation. The computations of a system therefore form a semigroup under composition, which we take to contain the identity, the computation that does nothing. Deep computations are the limits of computations along ultrafilters in the topology of pointwise convergence, taken in the space of types. Since an actual computation keeps the magnitudes that it handles within bounds, the shards of Section 2.1.1 enter the definition through the notion of confinement. The following definition records these ingredients.

Definition 2.4. A *compositional computation structure* (CCS) is a triple $(L, \mathcal{P}, \Gamma)$, where

- the pair $(L, \mathcal{P})$ is a computation states structure, and
- the set $\Gamma \subseteq L^L$ is a semigroup under composition that contains the identity map id of L , that is, a monoid of transformations of L .

The elements of Γ are called the *computations* of $(L, \mathcal{P}, \Gamma)$. The state space L is identified with a subset of the type space $\mathcal{L} \subseteq \mathbb{R}^{\mathcal{P}}$ of $(L, \mathcal{P})$.

We topologize Γ as a subset of the product L^L . In this product, as usual, the exponent L serves merely as an index set, whereas the base $L \subseteq \mathbb{R}^{\mathcal{P}}$ carries the topology of pointwise convergence. The semigroup Γ is thus a subset of the product $\mathcal{L}^L \subseteq (\mathbb{R}^{\mathcal{P}})^L$, and its closure $\bar{\Gamma}$ is contained in $\mathcal{L}^L$. The elements $\xi \in \bar{\Gamma}$ are called (real-valued) *deep computations*, and the elements of the closure $\bar{\Delta}$ of a set $\Delta \subseteq \Gamma$ are called the deep computations of Δ .⁵

A collection $\mathcal{R}$ of sizers is *exhaustive* if $L = \bigcup_{r_{\bullet} \in \mathcal{R}} L[r_{\bullet}]$. A computation $\gamma \in \Gamma$ is $\mathcal{R}$ -*confined* if it maps $L[r_{\bullet}]$ into itself for every $r_{\bullet} \in \mathcal{R}$, and a subset $\Delta \subseteq \Gamma$ is $\mathcal{R}$ -*confined* if each $\gamma \in \Delta$ is.

Unlike its subspace L^L , which is a semigroup under composition, the space $\mathcal{L}^L$ has no natural semigroup structure; deep computations are therefore not composable in general.

Proposition 2.5. *If $\Delta \subseteq \Gamma$ is confined by an exhaustive collection of sizers, then its closure $\bar{\Delta}$ is a compact subset of $\mathcal{L}_{\text{sh}}^L$.*

Proof. Let $\mathcal{R}$ be an exhaustive collection of sizers that confines Δ , and choose, for each $v \in L$, a sizer $r_{\bullet}^{(v)} \in \mathcal{R}$ with $v \in L[r_{\bullet}^{(v)}]$. Since every $\gamma \in \Delta$ maps v into $L[r_{\bullet}^{(v)}]$, the set Δ lies in the compact space $\prod_{v \in L} \mathcal{L}[r_{\bullet}^{(v)}] \subseteq \mathcal{L}_{\text{sh}}^L$, which therefore contains $\bar{\Delta}$. $\square$

For a CCS $(L, \mathcal{P}, \Gamma)$ we regard the elements of Γ as “standard” finitary computations and the elements of $\bar{\Gamma}$, i.e., deep computations, as possibly infinitary limits of standard computations.

⁵In the first paper of this series [ADIW26], the deep computations of Δ are the accumulation points of Δ that satisfy a confinement condition. Counting the members of Δ among them affects none of the properties studied in this paper. Indeed, under the Extendibility Axiom of Section 2.3, the members of Δ are continuous on the type shards, and those that are not accumulation points of Δ are isolated points of $\bar{\Delta}$.

Remarks 2.6. (For readers versed in model theory.)⁶

- (1) Compositional computation structures, as defined above, are not literally real-valued structures, because nothing in the structure itself forces Γ to be realized as a subset of the product space L^L . However, as in their original definition [ADIW26], they may be regarded as real-valued structures $(L, \mathcal{P}, \Gamma, \circ, \text{ev})$, where
 - the pair $(L, \mathcal{P})$ is a computation states structure,
 - the pair $(\Gamma, \circ)$ is a semigroup,
 - the map $\text{ev} : \Gamma \times L \rightarrow L : (\gamma, v) \mapsto \text{ev}(\gamma, v)$ is a semigroup action of Γ on L (i.e., $\text{ev}(\gamma \circ \delta, v) = \text{ev}(\gamma, \text{ev}(\delta, v))$ for $\gamma, \delta \in \Gamma$ and $v \in L$).

The two definitions are reconciled by identifying each $\gamma \in \Gamma$ with the map $v \mapsto \text{ev}(\gamma, v)$.

- (2) Regarded as real-valued structures in this sense, the compositional computation structures with a fixed collection $\mathcal{P}$ of predicate symbols form an elementary class. The notion of state type introduced above is essentially that of a quantifier-free type with parameters of sort L only. This notion may be enriched by further formulas, for instance with parameters in Γ or with quantifiers, and extended to types of sort Γ (“transform-types”). As usual, all types over a structure $\underline{C} = (L, \mathcal{P}, \Gamma, \circ, \text{ev})$ are realized in some ultrapower $\underline{C}^{(U)}$ of $\underline{C}$, constructed as in the references above.⁷ In Banach space theory, this approach to spaces of types goes back to Krivine’s real-valued languages [Kri72, Kri74] and to the work of Krivine and Maurey on stable Banach spaces [KM81].

2.3. Computability of deep computations and the Extendibility Axiom.

The rest of the paper depends on two notions, the computability of transforms and the Extendibility Axiom, both of which we recall in this subsection. The Extendibility Axiom requires every computation to have an extension that is continuous on each type shard. As for computability, a machine evaluates only sums and products of finitely many observed features, and only to within a preset error on a preset range of magnitudes. For this reason, computability is defined by uniform approximation on shards by polynomials in the features. Let $(L, \mathcal{P})$ be a computation states structure. A map $f : L \rightarrow \mathcal{L}$ from its state space to its space of types is called a *transform*. In the two definitions that follow, and in the characterization of computability after them, the collection $\mathcal{P}$ of predicates is assumed to be *countable*.

- (1) A *computable predicate* is a function from L to $\mathbb{R}$ whose restriction to each shard is a uniform limit of polynomials, with real coefficients, in the features $P(\cdot)$.
- (2) A transform $f : L \rightarrow \mathcal{L}$ is *computable* if, for each $Q \in \mathcal{P}$, the output feature $Q \circ f : L \rightarrow \mathbb{R}$ is a computable predicate.

Alva, Dueñez, Iovino and Walton [ADIW26] showed that a transform $f : L \rightarrow \mathcal{L}$ is computable if and only if it extends to a continuous map $\tilde{f} : \mathcal{L} \rightarrow \mathcal{L}$. This characterization motivates the following definition, in which continuity is required on each type shard.

⁶The real-valued structures used here are those of Keisler’s model theory for real-valued structures [Kei23]; see also [DnI17, Section 6] for an introduction.

⁷By contrast, no ultrapower is needed to realize state types in a computation states structure (see Remark 2.2).

A computation $\gamma \in \Gamma$ is *extendable* if it has an extension $\tilde{\gamma}: \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ such that, for every sizer $r_\bullet$, there is a sizer $s_\bullet$ for which the restriction $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]}: \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$ is continuous. Such an extension is unique, since it is continuous on each type shard $\mathcal{L}[r_\bullet]$, in which $L[r_\bullet]$ is dense. For a set $\Delta \subseteq \Gamma$ of extendable computations, we write $\tilde{\Delta}$ for the set $\{\tilde{\gamma}: \gamma \in \Delta\}$.

Extendibility Axiom. *Every computation $\gamma \in \Gamma$ is extendable.*

The Extendibility Axiom is discussed at length in the first paper of this series [ADIW26]. *From now on, all compositional computation structures are assumed to satisfy the Extendibility Axiom.*

The following lemma extends deep computations to the type shards; it is used in Theorems 3.1 and 6.7.

Lemma 2.7. *Let $\mathcal{R}$ be an exhaustive collection of sizers and let $\Delta \subseteq \Gamma$ be $\mathcal{R}$ -confined. Every deep computation $f \in \bar{\Delta}$ has an extension $\tilde{f}: \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ that, for every $r_\bullet \in \mathcal{R}$, maps $\mathcal{L}[r_\bullet]$ into itself and satisfies $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$.*

Proof. Write $f = \lim_{i, \mathcal{U}} \gamma_i$, where $(\gamma_i)_{i \in I}$ is a family in Δ and $\mathcal{U}$ is an ultrafilter on the index set I , and let $r_\bullet \in \mathcal{R}$ be given. Each γ_i maps $L[r_\bullet]$ into itself and is continuous on $\mathcal{L}[r_\bullet]$; hence it maps the closure $\mathcal{L}[r_\bullet]$ of $L[r_\bullet]$ into itself. Since $\mathcal{L}[r_\bullet]$ is compact, the limit $\tilde{f}(v) := \lim_{i, \mathcal{U}} \gamma_i(v)$ exists and lies in $\mathcal{L}[r_\bullet]$ for every $v \in \mathcal{L}[r_\bullet]$; in other words, the restriction $\tilde{f}|_{\mathcal{L}[r_\bullet]}$ equals $\lim_{i, \mathcal{U}} \gamma_i|_{\mathcal{L}[r_\bullet]}$. Since $\mathcal{R}$ is exhaustive, this defines $\tilde{f}$ on a set that contains L , and there $\tilde{f}$ agrees with f ; at the remaining points of $\mathcal{L}_{\text{sh}}$ we let $\tilde{f}$ be the identity. $\square$

Lemma 2.7 connects deep computations with the second notion of limit introduced in the first paper of this series [ADIW26]. In that paper, this second notion is formulated in the space of maps from the space of types to itself, endowed with the topology of pointwise convergence. An *ultracomputation* of a set $\Delta \subseteq \Gamma$ of extendable computations is an accumulation point of $\tilde{\Delta}$ in that space, subject to a confinement condition. Thus a deep computation sends states to types, whereas an ultracomputation sends types to types. Every deep computation is the restriction to L of an ultracomputation [ADIW26]. On the type shards $\mathcal{L}[r_\bullet]$ with $r_\bullet \in \mathcal{R}$, the ultralimit $\lim_{i, \mathcal{U}} \gamma_i$ in the proof of Lemma 2.7 provides such an extension of the deep computation $f = \lim_{i, \mathcal{U}} \gamma_i$ considered there. The classification of Section 3 is based on the sets $\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$, whose members are the restrictions to $\mathcal{L}[r_\bullet]$ of such ultralimits of extended computations. Strictly speaking, the classification therefore concerns the restrictions of ultracomputations to type shards. For the dynamical compositional computation structures of Section 5.1, in which every type is realized and every computation is its own extension, deep computations and ultracomputations coincide.

2.4. Stability. The highest level of the hierarchy studied in this paper is occupied by the stable sets of computations, all of whose deep computations are computable when $\mathcal{P}$ is countable. We recall these sets here, before descending to the levels below in Section 3. By the theorem of Alva, Dueñez, Iovino and Walton recalled above, the computable transforms are exactly those that extend to continuous maps on the space of types. Accordingly, the stable sets are defined by the continuity of their deep computations on every type shard.

Definition 2.8. A set $\Delta \subseteq \Gamma$ of computations is *stable* if, for every sizer $r_\bullet$, every deep computation in $\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is continuous (that is, computable if $\mathcal{P}$ is countable).

Stable sets of computations, the subject of the first paper of this series [ADIW26], are characterized there by Grothendieck’s double limit criterion and identified with stability in the sense of model theory.

2.5. Two motivating examples. Before developing the classification, we describe two algorithms whose deep computations differ sharply in character, for the contrast between them indicates what the classification must distinguish. Both algorithms are treated in detail in Section 4.

Square roots. Newton’s method for $p(x) = x^2 - a$, with $a > 0$, acts on the real projective line $\mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$ by the map $N_p(x) = (x^2 + a)/(2x)$, with $N_p(0) = N_p(\infty) = \infty$. The iterates of N_p are continuous and converge pointwise to a map $\mathfrak{f}$. This map sends every positive number to $\sqrt{a}$, every negative number to $-\sqrt{a}$, and both 0 and ∞ to ∞ (see Figure 2). The deep computations are the iterates together with $\mathfrak{f}$. Since the map $\mathfrak{f}$ is discontinuous, it is not computable, and the set of iterates is not stable. Nevertheless, the map $\mathfrak{f}$ is the limit of a *sequence* of computations, and the space of deep computations is countable.

Newton’s method on the Riemann sphere. For $q(z) = z^3 - 2z + 2$, Newton’s method on the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ displays three distinct modes of behavior, each confined to its own part of the sphere (see Figure 3 in Section 4.1). The iterates converge on the basins of the three roots. On the basin of the cycle $0 \mapsto 1 \mapsto 0$, by contrast, the iterates do not converge, although the even iterates and the odd iterates converge there separately to different limits. Finally, on the common boundary of these basins, a fractal set, the iterates are chaotic. For the simpler polynomial $z^2 - 1$, chaos of this kind produces deep computations that are not even limits of sequences of computations (see Proposition 5.15 and Theorem 3.1).

The classification of Section 3 is designed to separate such behaviors. In Theorem 3.1 below we characterize, for countably many features, the sets of computations whose deep computations are of the first Baire class on every type shard. For these sets every deep computation is, on each shard, the limit of a sequence of computations, as in the case of square roots. Todorčević’s trichotomy, which we recall in Section 3.2, then measures the complexity of the space of such deep computations.

3. CLASSIFYING DEEP COMPUTATIONS

The main theorems of the paper, proved in this section with the tools of the topology of function spaces, classify sets of computations by the complexity of their deep computations. The resulting classification yields a new hierarchy of complexity classes. The examples of Section 2.5 show why this classification is needed. Indeed, some natural algorithms have discontinuous deep computations that are nevertheless limits of sequences of computations, whereas others have deep computations that are not even limits of sequences. The main topological tool is Todorčević’s trichotomy for Rosenthal compacta [Tod99], and our model is the work of Glasner and Megrelishvili on tame semigroups in topological dynamics [GM22].

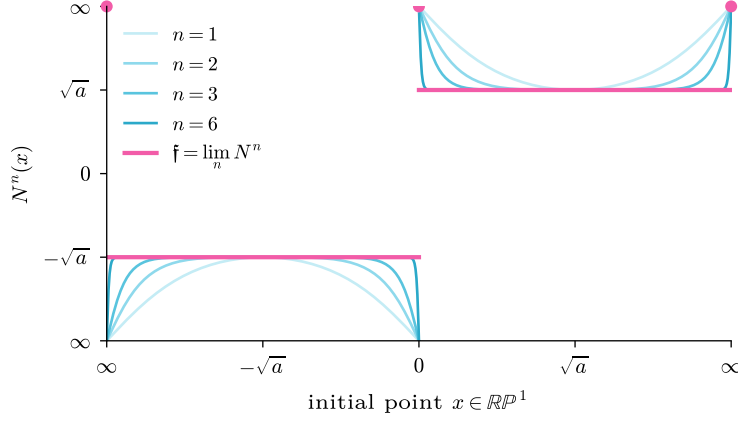


FIGURE 2. Newton iterates N_p^n for $p(x) = x^2 - a$, with $a > 0$, on the real projective line, drawn as graphs whose axes run from ∞ through 0 back to ∞ . The continuous maps N_p^n converge pointwise to the deep computation $\mathbf{f}$, which sends the positive half line to $\sqrt{a}$, the negative half line to $-\sqrt{a}$, and 0 and ∞ to ∞ .

3.1. NIP, Rosenthal compacta, and deep computations. The classification rests on Theorem 3.1, which we prove in this subsection. This theorem characterizes, by a finitary condition and for countably many features, the sets of computations whose deep computations are of the first Baire class on every type shard. Analogously, the sets of computations whose deep computations are continuous were characterized in the first paper of this series [ADIW26] by Grothendieck's double limit criterion.

Theorem 3.1. *Let $(L, \mathcal{P}, \Gamma)$ be a compositional computation structure (see Definition 2.4) satisfying the Extendibility Axiom, with $\mathcal{P}$ countable. Let $\mathcal{R}$ be an exhaustive collection of sizers, and suppose that $\Delta \subseteq \Gamma$ is $\mathcal{R}$ -confined. Then the following conditions are equivalent:*

- (i) *the inclusion $\overline{\Delta|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ holds for all $r_\bullet \in \mathcal{R}$;*
- (ii) *for all $P \in \mathcal{P}$ and $r_\bullet \in \mathcal{R}$, the set $\pi_P \circ \Delta|_{\mathcal{L}[r_\bullet]}$ satisfies the NIP, where $\pi_P: \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}$ is the projection onto the coordinate P . In other words, for all $P \in \mathcal{P}$ and $r_\bullet \in \mathcal{R}$, all reals $a < b$ and all sequences $(\gamma_n)_{n \in \mathbb{N}}$ in Δ , there are finite disjoint sets $I, J \subseteq \mathbb{N}$ such that*

$$L[r_\bullet] \cap \bigcap_{n \in I} (\pi_P \circ \gamma_n)^{-1}(-\infty, a] \cap \bigcap_{n \in J} (\pi_P \circ \gamma_n)^{-1}[b, \infty) = \emptyset.$$

Furthermore, if these conditions hold, then every deep computation $f \in \overline{\Delta}$ has an extension $\tilde{f}: \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$ such that $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ for all $r_\bullet \in \mathcal{R}$. In particular, on each shard, every deep computation is the pointwise limit of a sequence of computations.

Proof. Since $\mathcal{P}$ is countable, each type shard $\mathcal{L}[r_\bullet]$ is a compact metrizable, hence Polish, subspace of $\mathbb{R}^{\mathcal{P}}$. Let $r_\bullet \in \mathcal{R}$ be given; by the Extendibility Axiom and confinement, every $\tilde{\gamma}$ with $\gamma \in \Delta$ maps $\mathcal{L}[r_\bullet]$ continuously into itself (see the proof

of Lemma 2.7). Hence, for every $P \in \mathcal{P}$, the set $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ consists of continuous functions and is pointwise bounded. By Theorem 1.9 and Lemma 1.10, condition (ii) holds at $r_\bullet$ if and only if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathbb{R}^{\mathcal{P}})$, which (i) implies. Conversely, suppose that this inclusion holds. For a Polish space X , every compact subset of $B_1(X)$ is Fréchet–Urysohn (see [BFT78, Theorem 3F] or [Deb13, Theorem 4.1]). In a Fréchet–Urysohn space, every point in the closure of a subset is the limit of a sequence of points of that subset. By Proposition 1.8, the closure $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ may be regarded as a compact subset of $B_1(\mathcal{L}[r_\bullet] \times \mathcal{P})$, where the space $\mathcal{L}[r_\bullet] \times \mathcal{P}$ is Polish because $\mathcal{P}$ is countable and discrete. Every member of this closure is therefore the pointwise limit of a sequence of maps $\tilde{\gamma}_n|_{\mathcal{L}[r_\bullet]}$ with $\gamma_n \in \Delta$, which are continuous maps from $\mathcal{L}[r_\bullet]$ to itself; hence (i) holds at $r_\bullet$. Finally, if (i) holds, then the extension $\tilde{f}$ of Lemma 2.7 satisfies $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ for all $r_\bullet \in \mathcal{R}$. By the argument just given, it follows that every deep computation is, on each shard, the pointwise limit of a sequence of computations. $\square$

3.2. Todorčević’s trichotomy and levels of NIP. Todorčević’s trichotomy now enables us to stratify the NIP sets of computations into three levels. The trichotomy concerns separable Rosenthal compacta, and, by Theorem 3.1 and under its hypotheses, the deep computations of a countable NIP set of computations form such a compactum on each type shard. We measure the complexity of such a compactum by three topological properties, namely metrizability, hereditary separability and first countability. The three properties form a chain: every compact metrizable space is hereditarily separable, and R. Pol proved that every hereditarily separable Rosenthal compactum is first countable (see [Deb13, Section 10]). These properties define three levels of NIP.

Definition 3.2. Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom, with $\mathcal{P}$ countable, and let $\mathcal{R}$ be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be an $\mathcal{R}$ -confined countable set of computations satisfying the NIP on shards and features (condition (ii) in Theorem 3.1). We say that:

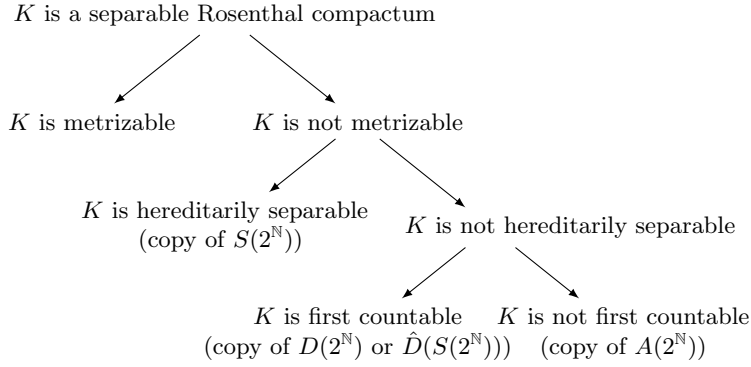
- (i) the set Δ is NIP_1 if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ is first countable for every $r_\bullet \in \mathcal{R}$;
- (ii) the set Δ is NIP_2 if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ is hereditarily separable for every $r_\bullet \in \mathcal{R}$;
- (iii) the set Δ is NIP_3 if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ is metrizable for every $r_\bullet \in \mathcal{R}$.

By the chain of properties above, the levels satisfy $\text{NIP}_3 \Rightarrow \text{NIP}_2 \Rightarrow \text{NIP}_1 \Rightarrow \text{NIP}$. Moreover, a countable stable set of computations (see Definition 2.8), which has the NIP by Theorem 3.1, is NIP_3 . Indeed, on each type shard the deep computations of such a set form a separable compact subset of a countable product of spaces $C_p(\mathcal{L}[r_\bullet])$, that is, a separable Eberlein compactum. Separable Eberlein compacta, in turn, are metrizable [Ark92]. The five compacta of Section 1.4 show that none of the implications between the three topological properties can be reversed for separable Rosenthal compacta. Specifically, the extended Alexandroff compactification $\hat{A}(2^{\mathbb{N}})$ is separable but not first countable. The extended duplicate $\hat{D}(S(2^{\mathbb{N}}))$ of the split Cantor set and the compactum $\hat{D}(2^{\mathbb{N}})$ that contains the Alexandroff duplicate are separable and first countable but not hereditarily separable. Finally, the split Cantor set $S(2^{\mathbb{N}})$ is hereditarily separable but not metrizable. Todorčević proved that these examples are canonical [Tod99], in the following sense.

Theorem 3.3 (Todorčević’s trichotomy [Tod99]; see also [ADK08, Theorem 3]). *Let K be a separable Rosenthal compactum.*

- (i) *If K is hereditarily separable but non-metrizable, then the split Cantor set $S(2^{\mathbb{N}})$ embeds into K .*
- (ii) *If K is first countable but not hereditarily separable, then either the Alexandroff duplicate $D(2^{\mathbb{N}})$ or the extended duplicate $\hat{D}(S(2^{\mathbb{N}}))$ of the split Cantor set embeds into K .*
- (iii) *If K is not first countable, then the Alexandroff compactification $A(2^{\mathbb{N}})$ embeds into K .*

Together with the chain of properties above, the trichotomy yields the following classification of separable Rosenthal compacta:



The three levels admit a second description, in terms of the number of tests required to distinguish deep computations from one another. Let Δ and $\mathcal{R}$ be as in Definition 3.2, fix $r_{\bullet} \in \mathcal{R}$, and write $K_{r_{\bullet}} = \overline{\Delta|_{\mathcal{L}[r_{\bullet}]}}$. We identify each $f \in K_{r_{\bullet}}$ with the real-valued function $(x, P) \mapsto P(f(x))$ on $\mathcal{L}[r_{\bullet}] \times \mathcal{P}$, whose points we call *tests*: the test (x, P) applies the feature P to the output on the input x . The topology of pointwise convergence on $K_{r_{\bullet}}$ is then the topology of pointwise convergence on tests. A set $\mathcal{T}$ of tests *separates* a set of deep computations if any two distinct members of the set differ on some test in $\mathcal{T}$. A member f of $K_{r_{\bullet}}$ is *countably determined* if some countable set $\mathcal{T}_f$ of tests separates f from every other member of $K_{r_{\bullet}}$. A set $E \subseteq K_{r_{\bullet}}$ is *finitely separated* if each $f \in E$ admits finitely many tests and some $\varepsilon_f > 0$ such that every $g \in E \setminus \{f\}$ differs from f by at least ε_f on one of these tests.

Proposition 3.4. *Let Δ and $\mathcal{R}$ be as in Definition 3.2, let $r_{\bullet} \in \mathcal{R}$ be given, and let $K_{r_{\bullet}} = \overline{\Delta|_{\mathcal{L}[r_{\bullet}]}}$.*

- (i) *The compactum $K_{r_{\bullet}}$ is metrizable if and only if one countable set of tests separates $K_{r_{\bullet}}$.*
- (ii) *It is first countable if and only if every member of $K_{r_{\bullet}}$ is countably determined.*
- (iii) *It is hereditarily separable if and only if no finitely separated subset of $K_{r_{\bullet}}$, that is, no subset that is discrete in the topology of pointwise convergence on tests, has cardinality $2^{\aleph_0}$.*

Proof. (i) If a countable set $\mathcal{T}$ of tests separates $K_{r_{\bullet}}$, then restriction to $\mathcal{T}$ is a continuous injection of the compact space $K_{r_{\bullet}}$ into the metrizable space $\mathbb{R}^{\mathcal{T}}$. This injection is therefore a homeomorphism onto its image. Conversely, if $K_{r_{\bullet}}$ is metrizable, then

the square $K_{r_\bullet} \times K_{r_\bullet}$ is second countable, and hence its open subset $\{(f, g) : f \neq g\}$ is Lindelöf. This open set is covered by the open sets $\{(f, g) : f(t) \neq g(t)\}$ for tests t , and a countable subcover gives the required set $\mathcal{T}$. (ii) If f is countably determined, the singleton $\{f\}$ is a countable intersection of open sets, and in a compact Hausdorff space such a point has a countable base of neighborhoods. Conversely, let $(U_k)_{k \in \mathbb{N}}$ be a countable base of neighborhoods of f , and choose in each U_k a basic neighborhood of f defined by a finite set F_k of tests. The union $\mathcal{T}_f = \bigcup_k F_k$ then separates f from every other member of $K_{r_\bullet}$. (iii) A set is finitely separated exactly when it is a discrete subspace, and, by Theorem 3.1, the compactum $K_{r_\bullet}$ is a separable Rosenthal compactum. A hereditarily separable space has no uncountable discrete subspace, since a separable discrete space is countable. Conversely, suppose that $K_{r_\bullet}$ is not hereditarily separable; then Theorem 3.3 provides a copy of $D(2^\mathbb{N})$, of $\hat{D}(S(2^\mathbb{N}))$ or of $A(2^\mathbb{N})$ inside $K_{r_\bullet}$. Each of these three compacta contains a discrete subspace of cardinality $2^{\aleph_0}$ (see Section 1.4). For separable Rosenthal compacta, the existence of such a subspace in the absence of hereditary separability is due to R. Pol (see [ADK08] and [GM22, Theorem 0.4]). $\square$

Thus the set Δ is NIP_3 when, on each type shard, a single countable set of tests separates all the deep computations of Δ . It is NIP_2 when no type shard carries a finitely separated set of deep computations of the cardinality of the continuum. Finally, the set Δ is NIP_1 when, on each type shard, each deep computation is separated from all the others by countably many tests of its own.

In Section 5 we exhibit computations at each level of the hierarchy and show that every implication between the classes stable, NIP_3 , NIP_2 , NIP_1 and NIP is strict. The definitions of NIP_1 , NIP_2 and NIP_3 are topological, and the descriptions of Proposition 3.4 still involve infinitely many tests. Whether the three levels nevertheless admit a finitary, combinatorial characterization, like that of NIP in Theorem 3.1, is Question 7.1.

3.3. The Argyros–Dodos–Kanellopoulos heptachotomy and binary search. For families of computations indexed by the binary tree, the Argyros–Dodos–Kanellopoulos heptachotomy describes their closures more finely than Todorćević’s trichotomy, which concerns the closures of arbitrary countable sets of computations. Accordingly, we refine the classification, reading the minimal families of the heptachotomy as primitives of binary search. The countable dense subsets exhibited in the separable compacta of Section 1.4 are indexed by the binary tree $2^{<\mathbb{N}}$, and this indexing suits our computational emphasis. Indeed, the branches of the binary tree are the infinite binary sequences that represent real numbers, and its nodes are the finite bitstrings that approximate them. Recall from Section 1.4 that a dyadic subtree of $2^{<\mathbb{N}}$ is written $\{w_t : t \in 2^{<\mathbb{N}}\}$, indexed through the canonical isomorphism, which preserves extension and the lexicographic order.

Definition 3.5. Let X be a Polish space.

- (1) If I is countable and $\{f_i : i \in I\} \subseteq \mathbb{R}^X$, $\{g_i : i \in I\} \subseteq \mathbb{R}^X$ are two families indexed by I , we say that they are *equivalent* if the map $f_i \mapsto g_i$ extends to a homeomorphism from $\{f_i : i \in I\}$ to $\{g_i : i \in I\}$.
- (2) If $\{f_t : t \in 2^{<\mathbb{N}}\}$ is a pointwise bounded family, we say that it is *minimal* if, for every dyadic subtree $\{w_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$, the family $\{f_{w_t} : t \in 2^{<\mathbb{N}}\}$ is equivalent to $\{f_t : t \in 2^{<\mathbb{N}}\}$.

Argyros, Dodos and Kanellopoulos proved that, up to equivalence, there are exactly seven minimal families $\{f_t : t \in 2^{<\mathbb{N}}\}$ that are relatively compact in $B_1(X)$. The same authors showed that every relatively compact family $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$ contains, along a dyadic subtree, a family equivalent to one of these seven families (see Theorem 3.6 below). We describe the seven minimal families next, in the notation of these authors. The description is carried out along the regular dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ fixed in Section 1.4 for the definition of the extended duplicate of the split Cantor set. This subtree satisfies $s_t \frown 0^\infty \neq s_{t'} \frown 0^\infty$ and $s_t \frown 1^\infty \neq s_{t'} \frown 1^\infty$ whenever $t \neq t'$. The families are built from the prefix tests v_t and the threshold tests $\chi_{[a]}$, $\chi_{(a]}$ of Section 1.4. Two further conventions are taken from the paper of Argyros, Dodos and Kanellopoulos [ADK08]. First, a point $\sigma \in 2^\mathbb{N}$, identified with a point of the Cantor set in $[0, 1]$, is regarded as the constant function on $2^\mathbb{N}$ with value σ . Second, a pair (f, g) of functions on $2^\mathbb{N}$ is regarded as a single function on the disjoint union $2^\mathbb{N} \sqcup 2^\mathbb{N}$.

- (1) The family $D_1 = \{\frac{1}{|t|+1}v_t : t \in 2^{<\mathbb{N}}\}$ has closure $D_1 \cup \{\mathbf{0}\}$, the one-point compactification of a countable discrete set.
- (2) The family $D_2 = \{s_t \frown 0^\infty : t \in 2^{<\mathbb{N}}\}$ is a discrete subset of its closure $\overline{D_2}$, which Argyros, Dodos and Kanellopoulos denote by $2^{\leq \mathbb{N}}$. This closure is compact and metrizable, and its accumulation points form a copy of $2^\mathbb{N}$.
- (3) The family $D_3 = \{\chi_{[s_t \frown 0^\infty]} : t \in 2^{<\mathbb{N}}\}$ is a discrete subset of its closure $\overline{D_3}$, a closed subspace of $S(2^\mathbb{N})$.
- (4) The family $D_4 = \{\chi_{(s_t \frown 1^\infty)} : t \in 2^{<\mathbb{N}}\}$ is a discrete subset of its closure $\overline{D_4}$, a closed subspace of $S(2^\mathbb{N})$.
- (5) The family $D_5 = \{v_t : t \in 2^{<\mathbb{N}}\}$ is a discrete subset of its closure $\overline{D_5} = \hat{A}(2^\mathbb{N})$.
- (6) The family $D_6 = \{(v_t, t \frown 0^\infty) : t \in 2^{<\mathbb{N}}\}$ is a discrete subset of its closure $\overline{D_6} = \hat{D}(2^\mathbb{N})$.
- (7) The family $D_7 = \{(v_{s_t}, \chi_{[s_t \frown 0^\infty]}) : t \in 2^{<\mathbb{N}}\}$ is a discrete subset of its closure $\overline{D_7} = \hat{D}(S(2^\mathbb{N}))$.

Theorem 3.6 (Heptachotomy of minimal families [ADK08, Theorem 2]). *Let X be a Polish space. For every relatively compact family $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$, there exist $i \in \{1, 2, \dots, 7\}$ and a regular dyadic subtree $\{w_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$ such that $\{f_{w_t} : t \in 2^{<\mathbb{N}}\}$ is equivalent to D_i . Moreover, the families D_i are minimal and pairwise non-equivalent.*

For deep computations, the building blocks of the families $D_1, \dots, D_7$ have a direct reading as primitives of binary search on $2^\mathbb{N}$, the space of binary expansions. A node $t \in 2^{<\mathbb{N}}$ determines the dyadic interval $[t \frown 0^\infty, t \frown 1^\infty]$, the set of inputs that round to t at precision $|t|$. The prefix test v_t asks whether the input rounds to t , and the point $t \frown 0^\infty$ is the rounded value itself. The comparison $\chi_{[t \frown 0^\infty]}$ asks whether the input lies at or to the right of the left endpoint of the interval. Likewise, the comparison $\chi_{(t \frown 1^\infty)}$ asks whether the input lies strictly to the right of the right endpoint. The resulting reading of the seven families is recorded in Table 1.

Let $\{\gamma_t : t \in 2^{<\mathbb{N}}\}$ be a family of computations of a CCS with countably many features. Suppose that this family is confined by an exhaustive collection $\mathcal{R}$ of sizers and satisfies the NIP on shards and features (condition (ii) of Theorem 3.1). By Theorem 3.1, for every feature P and every $r_\bullet \in \mathcal{R}$, the functions $P \circ \tilde{\gamma}_t|_{\mathcal{L}[r_\bullet]}$ form a relatively compact subset of $B_1(\mathcal{L}[r_\bullet])$. The heptachotomy (Theorem 3.6)

Family	Function at node t	Limits
D_1	prefix test weighted by $1/(t + 1)$	the test that never succeeds: confidence vanishes with depth
D_2	return the rounded value $t \frown 0^\infty$	the exact input: successive approximation
D_3	compare with the left endpoint	comparison with a limit of left endpoints, with ties broken upward and, at limits approached from above, also downward
D_4	compare with the right endpoint	comparison with a limit of right endpoints, with ties broken downward and, at limits approached from below, also upward
D_5	prefix test	exact equality, or no match
D_6	prefix test and rounded value	a dictionary lookup that returns its key
D_7	prefix test and comparison	a three-way comparison: below, inside or above

TABLE 1. The minimal families of Argyros, Dodos and Kanellopoulos as primitives of binary search.

then shows that the family of these functions contains, along a dyadic subtree, a family equivalent to one of the seven families $D_1, \dots, D_7$. In the reading of Table 1, the family $\{\gamma_t : t \in 2^{<\mathbb{N}}\}$ thus contains, on each type shard and for each feature, one of the seven primitives of binary search. In each case, the primitive occurs up to equivalence and along a dyadic subtree.

4. EXAMPLES OF COMPOSITIONAL COMPUTATION STRUCTURES

From the general theory we turn to the examples that motivate it, among them the two announced in Section 2.5, and describe the deep computations of three compositional computation structures. These three structures are Newton’s method on the Riemann sphere for two cubic polynomials (Sections 4.1 and 4.2) and Newton’s method for square roots on the real projective line (Section 4.3). In Section 5 we place the last of these, together with other classical algorithms, at the levels of the hierarchy of Section 3.

We begin by formalizing the Newton–Raphson root-finding method as a CCS [Gal00], [BFB16, Section 2.3]. Let $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the extended complex plane, that is, the Alexandroff (one-point) compactification of $\mathbb{C}$. We identify $\hat{\mathbb{C}}$ with the unit sphere $S^2 \subseteq \mathbb{R}^3$, the Riemann sphere, through the inverse stereographic projection. This projection is the homeomorphism $\hat{\mathbb{C}} \rightarrow S^2$ given by $z \mapsto (P_1(z), P_2(z), P_3(z))$, where $(P_1(\infty), P_2(\infty), P_3(\infty)) = (0, 0, 1)$ and, for $z \neq \infty$,

$$P_1(z) = \frac{2 \Re z}{|z|^2 + 1}, \quad P_2(z) = \frac{2 \Im z}{|z|^2 + 1}, \quad P_3(z) = \frac{|z|^2 - 1}{|z|^2 + 1}.$$

The extended complex plane $\hat{\mathbb{C}}$ thus carries the real predicates P_1, P_2, P_3 with values in $[-1, 1]$. Through the same identification, the topology of $\hat{\mathbb{C}}$ is metrized by the pullback of the Euclidean metric of S^2 .

Newton CCSs: steps and fractals. Here we describe the compositional computation structure generated by the Newton map, the single step that Newton's method iterates to approximate a root of a function. The structure so obtained is the common setting of the three examples that follow. We then recall the Fatou and Julia sets, on which the iterates of the Newton map behave regularly and chaotically, respectively. Let $p = p(z)$ be a nonconstant rational function with complex coefficients, that is, a nonconstant meromorphic map $\hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$. The *Newton step*, or *Newton map*, of p is the rational map

$$(4.1) \quad N_p : z \mapsto z - \frac{p(z)}{p'(z)}.$$

The iterates of the Newton map, that is, its compositional powers N_p^n , may be regarded as the computations of a CCS with state space $L = \hat{\mathbb{C}}$. The natural predicates are the coordinates P_1, P_2, P_3 of the inverse stereographic projection, since the real and imaginary parts $z \mapsto \Re z$ and $z \mapsto \Im z$ are not defined at $z = \infty$.

Let $\Gamma_p := \{N_p^n : n \in \mathbb{N}\}$ be the semigroup of iterates of the Newton map (4.1). The CCS $(\hat{\mathbb{C}}, (P_1, P_2, P_3), \Gamma_p)$ is called *Newton's method for $p(z)$* , or the *Newton CCS for $p(z)$* . Since the state space $\hat{\mathbb{C}}$ is compact and the Newton map N_p is continuous on it, every Newton CCS satisfies the Extendibility Axiom.

Newton's method succeeds at an initial estimate $z_0 \in \hat{\mathbb{C}}$ when the sequence $(z_n)_{n \in \mathbb{N}}$ of Newton steps, where $z_n := N_p^n(z_0)$, converges to a root of p .

Fatou and Julia sets. Recall that the *Fatou set* $\mathcal{F}_g$ of a rational map $g : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ is the set of points near which the iterates of g behave regularly. More precisely, a point belongs to $\mathcal{F}_g$ if it has a neighborhood on which every sequence of iterates of g has a subsequence that converges uniformly with respect to the spherical metric. The complement $\mathcal{J}_g := \hat{\mathbb{C}} \setminus \mathcal{F}_g$ of the Fatou set is the *Julia set* of g , a closed set that is typically a fractal. The Julia set of the Newton map N_p is called the *Newton fractal* of p .

4.1. Example: The Newton CCS for the cubic $q(z) = z^3 - 2z + 2$. Our first example is the second motivating example of Section 2.5, a cubic polynomial whose Newton map has, besides the three roots, an attracting cycle. Newton's method therefore fails on an open set of initial estimates, and the iterates of the Newton map do not converge. For the same reason, two distinct limits of the iterates along ultrafilters are among the deep computations of this method. One is taken along an ultrafilter that contains the even numbers, and the other along an ultrafilter that contains the odd numbers. The polynomial $q(z) = z^3 - 2z + 2$ has Newton map

$$N_q : z \mapsto z - \frac{z^3 - 2z + 2}{3z^2 - 2} = \frac{2z^3 - 2}{3z^2 - 2}.$$

Denote the real root of q by r_0 , and the roots in the upper and the lower half plane by r_+ and r_- . Starting from $z_0 = 0$, the Newton steps alternate $0, 1, 0, 1, \dots$, and neither 0 nor 1 is a root of q .

Let B_0, B_+, B_- be the basins of the roots r_0, r_+, r_- , and let B_{01} be the basin of the superattracting cycle $\{0, 1\}$. The Fatou set of N_q is $B_{01} \cup B_0 \cup B_+ \cup B_-$, and each of these four sets is nonempty and open. Since B_{01} is nonempty, the iterates N_q^n do not converge pointwise. Fix nonprincipal ultrafilters $\mathcal{U}_0$ and $\mathcal{U}_1$ on $\mathbb{N}$ that contain the even and the odd numbers, respectively. The limits of the iterates N_q^n along $\mathcal{U}_0$ and along $\mathcal{U}_1$ are deep computations, which we denote by f_0 and f_1 ,

respectively. On the Fatou set the even and the odd iterates converge pointwise, whence the two limits, on that set, do not depend on the choice of the ultrafilters. Both limits send each basin of a root to that root; on B_{01} , by contrast, they take only the values 0 and 1 and exchange them. The exchange of the two values follows from the identity $f_1 = N_q \circ f_0$, which holds on the Fatou set by the continuity of N_q . In particular, the two deep computations are distinct.⁸ Figure 3 depicts f_0 and f_1 , together with the Newton fractal of q , on which the iterates behave chaotically.

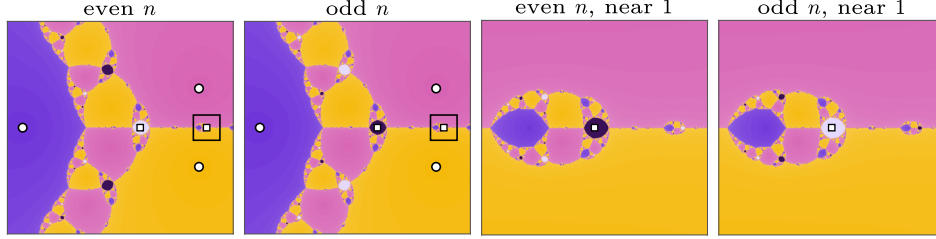


FIGURE 3. The deep computations of Newton's method for $q(z) = z^3 - 2z + 2$: the limits f_0 and f_1 of the iterates along $\mathcal{U}_0$ and along $\mathcal{U}_1$, on the whole plane and near the point 1. Points in the basins of the real root and of the roots in the upper and lower half planes (white circles) are purple, orchid and gold, paler where convergence is slower. The remaining open region is the basin of the superattracting cycle $0 \mapsto 1 \mapsto 0$ (white squares). A point there is lilac if the iterate in question is near 0 and plum if it is near 1, whence the two colors exchange between even and odd iterates. Accordingly, the two deep computations agree on the basins of the roots and differ on the basin of the cycle. The boundaries of the colored regions form the Newton fractal of q .

4.2. Example: The Newton CCS for the cubic $u(z) = z^3 - 1$. For contrast, we turn to $u(z) = z^3 - 1$, whose roots are the cube roots of unity $r_0 = 1$ and $r_{\pm} = \exp(\pm 2\pi i/3)$. The Newton map of u is

$$N_u : z \mapsto z - \frac{z^3 - 1}{3z^2} = \frac{2z^3 + 1}{3z^2}.$$

The critical points of this rational map of degree three are the three roots of u , which are superattracting fixed points, and the point 0. The map N_u sends the point 0 to ∞ , a repelling fixed point of multiplier $3/2$. The immediate basin of every attracting or parabolic cycle contains a critical point whose orbit converges to that cycle [Mil06, Sections 8 and 10]. Unlike N_q , the map N_u therefore has no attracting cycle other than its three attracting fixed points, the roots of u , and it has no parabolic cycle. The boundary of a Siegel disk or of a Herman ring, moreover, is infinite and lies in the closure of the critical orbits [Mil06, Sections 11

⁸The point ∞ is a repelling fixed point of N_q : if $|z|$ is large, then $N_q(z) = \frac{2}{3}z + O(1)$. Hence the point ∞ lies on the Newton fractal of q , and the two deep computations satisfy $f_0(\infty) = \infty = f_1(\infty)$.

and 15]. Since every critical orbit of N_u is finite, the map N_u has no such domain. Sullivan's theorem excludes wandering domains, and the classification of periodic Fatou components then shows that the Fatou set of N_u is the union of the basins of the three roots [Mil06, Section 16 and Appendix F]. On each of these basins, the iterates of N_u converge to the corresponding root. Figure 4 depicts Newton's method for u . To render the iterates themselves, and not merely their limits, we color each point z_0 by blending the colors of the three roots with relative weights $|z_n - r|^{-3}$ for $r \in \{r_0, r_+, r_-\}$, where $z_n := N_u^n(z_0)$. This coloring is a continuous function of z_0 , blurred for small n . As n increases, the coloring sharpens into large regions of nearly constant color, separated by narrow regions in which the color varies so rapidly that the coloring is nearly discontinuous.

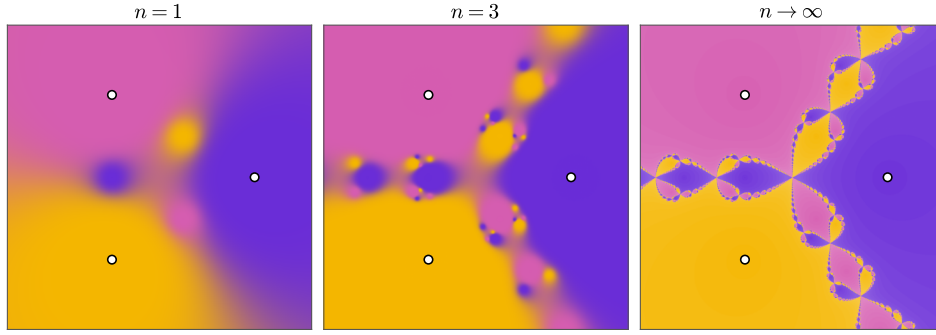


FIGURE 4. Newton's method for $u(z) = z^3 - 1$ after one and three iterations, and the limit of its iterates on the Fatou set. The coloring blends the colors of the three cube roots of unity (white circles) according to the distance of $N_u^n(z)$ to each of them. This limit is constant on each basin and changes value across the Newton fractal; in the right panel, points are paler where convergence is slower.

The chaos of Newton's method on its Julia set is visible to the predicates P_1, P_2, P_3 . In Propositions 5.15 and 5.16 below we show that, for the polynomials $z^2 - 1$, q and u , this chaos prevents the Newton CCS on the Riemann sphere from being NIP.

4.3. Example: Square roots. We return to the first motivating example of Section 2.5, Newton's method on the real projective line for the polynomial $p(x) = x^2 - a$ with real $a > 0$, whose iterates approximate the square roots $\pm\sqrt{a}$. The deep computations of this method, which we now work out, place it at the level NIP_3 (see Section 5.2). As for every rational function with real coefficients, the Newton map N_p maps the real projective line $\mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$ into itself. The Newton CCS for p may therefore be regarded as having state space $\mathbb{RP}^1$.

Let $\mathbb{R} \rightarrow S^1$, $x \mapsto (P_1(x), P_3(x))$, be the inverse stereographic projection of the real line onto $S^1 \subseteq \mathbb{R}^2$, given by

$$P_1(x) = \frac{2x}{x^2 + 1}, \quad P_3(x) = \frac{x^2 - 1}{x^2 + 1},$$

and extended to a homeomorphism of the real projective line $\mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$ onto S^1 by $\infty \mapsto (P_1(\infty), P_3(\infty)) = (0, 1)$. These coordinates are the restrictions to $\mathbb{RP}^1$

of the predicates P_1 and P_3 of the Riemann sphere, whose names they keep. As predicates, the two coordinates make $(\mathbb{RP}^1, (P_1, P_3))$ a computation states structure with a compact state space. The Newton map N_p is

$$x \mapsto x - \frac{x^2 - a}{2x} = \begin{cases} \frac{x^2 + a}{2x}, & x \neq 0, \infty; \\ \infty, & x = 0, \infty. \end{cases}$$

From every initial point $x_0 \neq 0, \infty$, the Newton steps converge to one of the square roots $\pm\sqrt{a}$. More precisely, the iterates of N_p converge pointwise to the map $\mathfrak{f}: \mathbb{RP}^1 \rightarrow \mathbb{RP}^1$ given by

$$\mathfrak{f}(x) = \lim_{n \rightarrow \infty} N_p^n(x) = \begin{cases} \sqrt{a}, & x \in (0, \infty); \\ -\sqrt{a}, & x \in (-\infty, 0); \\ \infty, & x = 0, \infty. \end{cases}$$

The map $\mathfrak{f}$ is therefore discontinuous at 0 and at ∞ (see Figure 2).

5. COMPUTATIONS ACROSS THE HIERARCHY

The hierarchy of Section 3 descends from the stable sets of computations through the classes NIP_3 , NIP_2 and NIP_1 to the class NIP . The implications of the hierarchy hold for countable sets of computations in compositional computation structures with countably many features. Although its levels are defined through the topology of the space of deep computations, each of them is attained by classical algorithms; some familiar algorithms lie beyond the hierarchy. In this section we establish both facts, and with them the strictness of every implication. In each example the computations are compositions of the steps of an algorithm. At all levels but one, these computations are the powers of a single step map. At the level NIP but not NIP_1 they form instead a group generated by two steps, both in the Euclidean algorithm and in free reduction in the free group on two generators. Moreover, a countable family of elements of the group generated by the steps of the Euclidean algorithm provides a second example at the level NIP_1 but not NIP_2 . The level of each algorithm is then a theorem about the closure of its computations, and for such computations the hierarchy coincides with the Glasner–Megrelishvili hierarchy of tame dynamical systems [GM22]. We begin with this coincidence (see Proposition 5.1); Table 2 at the end of the section summarizes the examples.

5.1. Iterative algorithms and tame dynamics. The algorithms of this section act by continuous steps on a compact space of states. Within this setting, which we now isolate, our hierarchy coincides with the Glasner–Megrelishvili hierarchy of tame dynamical systems. The coincidence allows the results of Glasner and Megrelishvili to be brought to bear on the examples that follow. A *dynamical CCS* is a compositional computation structure $(L, \mathcal{P}, \Gamma)$ in which L is a compact metrizable space and the predicates in $\mathcal{P}$ are continuous and separate the points of L . Moreover, the family $\mathcal{P}$ and the semigroup Γ are countable, and the members of Γ are continuous maps $L \rightarrow L$, among them the identity. A dynamical CCS is a *cascade CCS* if $\Gamma = \{\gamma^n : n \in \mathbb{N}\}$ is the semigroup of iterates of a single continuous map $\gamma: L \rightarrow L$. Since $\mathcal{P}$ separates the points of the compact space L , the map $v \mapsto \text{tp}(v)$ is a homeomorphism of L onto the space of types. Consequently a single shard covers L , every type is realized, and the Extendibility Axiom holds

with $\tilde{\gamma} = \gamma$ for every $\gamma \in \Gamma$. The deep computations of a dynamical CCS form the pointwise closure

$$E(L, \Gamma) = \overline{\Gamma} \subseteq L^L,$$

which is the *enveloping semigroup*, introduced by Ellis, of the dynamical system (L, Γ) ; for a cascade CCS we write $E(L, \gamma)$. The system (L, Γ) is called *tame* if, for every continuous function $f: L \rightarrow \mathbb{R}$, the family $\{f \circ \gamma : \gamma \in \Gamma\}$ has the NIP of Definition 1.5 [Köh95, Gla06]. We say that a dynamical CCS is *stable* when its set Γ of computations has the property of Definition 2.8. Similarly, the CCS is called NIP when Γ satisfies condition (ii) of Theorem 3.1 with respect to a single sizer $r_\bullet$ with $L[r_\bullet] = L$. With respect to the same sizer, it is called NIP_1 , NIP_2 or NIP_3 when Γ has the corresponding property of Definition 3.2.

A map $\xi: L \rightarrow L$ is of *Baire class 1* if $f \circ \xi$ is of Baire class 1 for every continuous $f: L \rightarrow \mathbb{R}$. It suffices that $P \circ \xi$ be of Baire class 1 for every $P \in \mathcal{P}$. Indeed, since the type map embeds L into $\mathbb{R}^{\mathcal{P}}$ with $\mathcal{P}$ countable, every open subset of L is a countable union of finite intersections of sets $P^{-1}(U)$ with $P \in \mathcal{P}$ and $U \subseteq \mathbb{R}$ open. By Fact 1.2, the preimage of every open set under ξ is then an F_σ set, and so is the preimage of every open set under $f \circ \xi$. A second appeal to Fact 1.2 shows that the function $f \circ \xi$ is of Baire class 1. In general, the definition of Baire class 1 given above for maps from L to L is less restrictive than the one adopted in Section 1. The latter requires a map from L to L to be the pointwise limit of a sequence of continuous maps from L to L . For example, if $L = [0, 1] \cup \{2\} \subseteq \mathbb{R}$, then the map that sends $[0, 1/2]$ to 0 and the rest of L to 2 is of Baire class 1 in the sense defined above. This map, however, is not the pointwise limit of a sequence of continuous maps from L to L , because each such map sends the connected set $[0, 1]$ into $[0, 1]$ or to 2. For the deep computations of a dynamical CCS, by contrast, the two definitions agree. If every deep computation is of Baire class 1 in the sense defined above, then the CCS is NIP, as we show in Proposition 5.1 below. In that case the last assertion of Theorem 3.1, applied to the single sizer $r_\bullet$ with $L = L[r_\bullet]$, shows that every deep computation is the pointwise limit of a sequence of computations. Since computations are continuous maps from L to L , every deep computation is then of Baire class 1 in the sense of Section 1 as well.

Proposition 5.1. *A dynamical CCS is NIP if and only if the dynamical system (L, Γ) is tame, and if and only if every deep computation is of Baire class 1. In that case the CCS is NIP_1 , NIP_2 or NIP_3 if and only if $E(L, \Gamma)$ is first countable, hereditarily separable or metrizable, respectively.*

Proof. The single sizer $r_\bullet$ with $L = L[r_\bullet]$ forms an exhaustive collection that confines Γ . If (L, Γ) is tame, the CCS is NIP: take $f = P$ for $P \in \mathcal{P}$. If the CCS is NIP, then Theorem 3.1 and Proposition 1.8 show that the function $P \circ \xi$ is of Baire class 1 for every $\xi \in E(L, \Gamma)$ and $P \in \mathcal{P}$. Every deep computation is therefore of Baire class 1. Finally, suppose that every deep computation is of Baire class 1, and let $f: L \rightarrow \mathbb{R}$ be continuous. The map $\xi \mapsto f \circ \xi$ is continuous on L^L , so the closure of $\{f \circ \gamma : \gamma \in \Gamma\}$ is the compact set $\{f \circ \xi : \xi \in E(L, \Gamma)\}$, which lies in $B_1(L)$. By Theorem 1.6, the family $\{f \circ \gamma : \gamma \in \Gamma\}$ has the NIP, so (L, Γ) is tame. The last statement is Definition 3.2, since $\overline{\Gamma}|_L = E(L, \Gamma)$. $\square$

For metrizable systems, Glasner and Megrelishvili call a tame system *tame₁* when its enveloping semigroup is first countable, *tame₂* when it is hereditarily separable, and *hereditarily non-sensitive* when it is metrizable. They prove the

inclusions between these classes [GM22, Proposition 0.6] and show by examples that the inclusions are proper. It follows from Proposition 5.1 that a dynamical CCS is NIP_1 , NIP_2 or NIP_3 exactly when its dynamical system is tame₁, tame₂ or hereditarily non-sensitive, respectively.

The simplest iterative methods are those whose iterates converge pointwise. In Proposition 5.2 we show that all of them are NIP_3 , and that the continuity of the limit decides whether they are stable.

Proposition 5.2. *Let $(L, \mathcal{P}, \Gamma)$ be a cascade CCS, with $\Gamma = \{\gamma^n : n \in \mathbb{N}\}$ for a continuous map $\gamma : L \rightarrow L$, and suppose that the iterates γ^n converge pointwise to a map $\gamma_\infty : L \rightarrow L$. Then $E(L, \gamma) = \{\gamma^n : n \in \mathbb{N}\} \cup \{\gamma_\infty\}$, and the CCS is NIP_3 . It is stable if and only if γ_∞ is continuous.*

Proof. A convergent sequence together with its limit is a compact subset of the Hausdorff space L^L , hence closed, and the enveloping semigroup $E(L, \gamma)$ therefore consists of the iterates and γ_∞ . For each $P \in \mathcal{P}$, the map $P \circ \gamma_\infty$ is the pointwise limit of the continuous functions $P \circ \gamma^n$ and is therefore of Baire class 1; by Proposition 5.1, the CCS is NIP . The space $E(L, \gamma)$ is a countable compact Hausdorff space, hence metrizable, so the CCS is NIP_3 . Its members other than γ_∞ are continuous, which gives the statement about stability. $\square$

5.2. Stable and metrizable. Our descent through the hierarchy starts at its top two levels, stable and NIP_3 , at which the deep computations form a metrizable space. At these two levels we place a rotation of the circle, Newton's method for square roots, the power method and the logistic map for $0 < r \leq 3$. By Proposition 5.2, every convergent method lies at one of these two levels. Convergence, however, is not necessary for stability, as the first example shows.

A stable example: rotation. Let the space of states be the circle $L = \mathbb{T} = \mathbb{R}/\mathbb{Z}$, with the predicates $\cos 2\pi x$ and $\sin 2\pi x$, and let the step be the rotation $\gamma(x) = x + \alpha$ by an irrational number α . Since a pointwise limit of rotations is a rotation and the points $n\alpha$ with $n \in \mathbb{N}$ are dense in $\mathbb{T}$, the deep computations are all the rotations $x \mapsto x + c$ with $c \in \mathbb{T}$. These rotations are continuous, so the rotation CCS is stable, and its space of deep computations is the circle group, a metrizable compactum.

Square roots: NIP_3 but not stable. For $a > 0$, the iterates of the Newton map N_p of the polynomial $p(x) = x^2 - a$ on the real projective line converge pointwise to the map $\mathfrak{f}$ defined in Section 4.3. Since this limit is discontinuous at 0 and ∞ , Proposition 5.2 shows that the square-root CCS is NIP_3 but not stable.

The power method: NIP_3 but not stable. The power method computes a dominant eigenvector of a matrix by repeated multiplication and normalization. Let $d \geq 2$ be an integer, and let $A = \text{diag}(\lambda_1, \dots, \lambda_d)$ be a diagonal matrix with $\lambda_1 > \lambda_2 > \dots > \lambda_d > 0$. The space of states is the real projective space $L = \mathbb{RP}^{d-1}$ of lines through the origin in $\mathbb{R}^d$, in which $[v]$ denotes the line spanned by a nonzero vector v . The family $\mathcal{P}$ consists of the predicates $P_{ij}([v]) = v_i v_j / \|v\|^2$ with $1 \leq i, j \leq d$, which are the entries of the orthogonal projection onto the line $[v]$. The step of the algorithm is $\gamma([v]) = [Av]$. For $[v] \in \mathbb{RP}^{d-1}$, let $i(v)$ be the least index i with $v_i \neq 0$, and put $\ell([v]) = [e_{i(v)}]$, where $e_1, \dots, e_d$ is the standard basis of $\mathbb{R}^d$.

Proposition 5.3. *The deep computations of the power method are $E(L, \gamma) = \{\gamma^n : n \in \mathbb{N}\} \cup \{\ell\}$. Hence the power method is NIP_3 . Its deep computation ℓ , the only*

one that is not an iterate, is of Baire class 1 and discontinuous, so the power method is not stable.

Proof. For $i = i(v)$, we have $\lambda_i^{-n} A^n v = v_i e_i + \sum_{j>i} (\lambda_j/\lambda_i)^n v_j e_j \rightarrow v_i e_i$, so $\gamma^n \rightarrow \ell$ pointwise, and Proposition 5.2 applies. The map ℓ is discontinuous at every $[v]$ with $i(v) > 1$, since $[v + te_1] \rightarrow [v]$ as $t \rightarrow 0$ while $\ell([v + te_1]) = [e_1] \neq \ell([v])$. $\square$

The deep computation ℓ sends each input to the eigenline of the largest eigenvalue whose eigenvector has a nonzero component in the input. Being discontinuous on the space of types L , the map ℓ is not computable in the sense of Section 2. This deep computation is nevertheless the limit of a sequence of computations, and the closure in which it lives is countable and metrizable.

Both levels: the logistic map. The logistic map $f_r(x) = rx(1-x)$, with $0 < r \leq 4$, maps $[0, 1]$ into itself. Its CCS has states $L = [0, 1]$, the single predicate $x \mapsto x$, and the semigroup of iterates of f_r (see Figure 5). For $0 < r \leq 1$, let ℓ_r be the constant map 0; for $1 < r \leq 3$, let ℓ_r be the map with $\ell_r(x) = 1 - 1/r$ for $0 < x < 1$ and $\ell_r(0) = \ell_r(1) = 0$.

Proposition 5.4. *For $0 < r \leq 3$, the iterates of f_r converge pointwise to ℓ_r . Consequently the logistic CCS is stable for $0 < r \leq 1$, and NIP_3 but not stable for $1 < r \leq 3$.*

Proof. Write $f = f_r$ and $x^* = 1 - 1/r$, and note that $f(x)/x = r(1-x)$ for $x > 0$. If $r \leq 1$, then $f(x) < x$ for $0 < x \leq 1$, so every orbit decreases to a fixed point of f , and 0 is the only fixed point in $[0, 1]$.

Suppose that $1 < r \leq 2$, in which case $0 < x^* \leq \frac{1}{2}$. The inequality $f(x) > x$ holds for $0 < x < x^*$, and the reverse inequality for $x^* < x \leq 1$. For $0 < x < 1$ we have $f(x) \in (0, \frac{r}{4}] \subseteq (0, \frac{1}{2}]$, and the function f , which fixes x^* , is increasing on $[0, \frac{1}{2}]$. Hence the orbit of $f(x)$ is increasing and bounded by x^* if $f(x) < x^*$, and decreasing and bounded below by x^* if $f(x) > x^*$. In both cases, as well as trivially when $f(x) = x^*$, the orbit of $f(x)$ converges to a fixed point of f in $(0, \frac{1}{2}]$, namely to x^* .

Suppose that $2 < r \leq 3$, and put $c = \frac{r}{4}$, the maximum of f , and $b = f(c)$. We have $x^* < c$, since $c - x^* = (r-2)^2/(4r)$, and therefore $b < f(x^*) = x^*$, since f is strictly decreasing on $[\frac{1}{2}, 1]$. We also have $b \geq 1/r$, because $b - 1/r = (r^3(4-r) - 16)/(16r)$ and the polynomial $r^3(4-r)$ increases from 16 on $[2, 3]$. The concave function f takes the value x^* at $1/r$ and at x^* , so $f \geq x^*$ on $[1/r, x^*]$. Hence the function f maps $J_- = [b, x^*]$ into $J_+ = [x^*, c]$, and, being decreasing on J_+ , it maps J_+ onto J_- . Fix x with $0 < x < 1$; the orbit of $f(x)$ then lies in $(0, c]$, and it meets $J_- \cup J_+ = [b, c]$. Indeed, if the orbit did not meet $[b, c]$, all its points would lie in $(0, b)$, where $f(y) > y$ because $b < x^*$. The orbit would then increase to a fixed point of f in $(0, b]$. No such fixed point exists, since the only positive fixed point of f is x^* , which exceeds b . Since f maps $J_- \cup J_+$ into itself, it remains to be shown that the orbit of every $y \in J_- \cup J_+$ converges to x^* . The points of period 2 of f that are not fixed are the roots of $r^2 y^2 - r(r+1)y + (r+1)$, whose discriminant $r^2(r+1)(r-3)$ is negative for $r < 3$; for $r = 3$ the only root is x^* . Hence the function $f(f(y)) - y$ vanishes on $[0, 1]$ only at 0 and x^* , and it is negative on $(x^*, 1]$, since it is negative at 1. For $y \in J_+$ with $y > x^*$, the point $f(f(y))$ lies in J_+ , so $x^* \leq f(f(y)) < y$. The sequence $(f^{2n}(y))$ therefore decreases to a fixed point of $f \circ f$ in $[x^*, y]$, that is, to x^* , whence $f^{2n+1}(y) \rightarrow f(x^*) = x^*$. Every $y \in J_-$ satisfies $f(y) \in J_+$, so its orbit converges to x^* as well.

Finally, the equalities $f(0) = f(1) = 0$ account for the endpoints, and in all cases the iterates therefore converge pointwise to ℓ_r . The map ℓ_r is continuous for $r \leq 1$, and discontinuous at 0 for $r > 1$, so Proposition 5.2 gives the remaining statements. $\square$

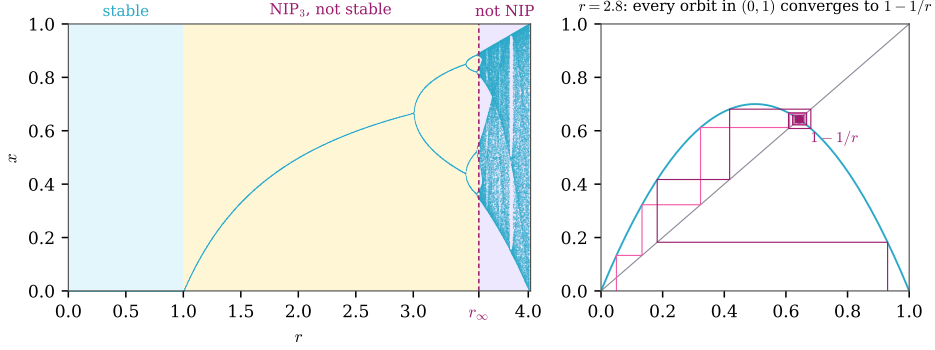


FIGURE 5. The logistic family $f_r(x) = rx(1 - x)$. Left: the long-run values of the orbit of 0.3 under f_r , for each r . The logistic CCS is stable for $r \leq 1$. It is NIP_3 but not stable for $1 < r \leq r_\infty$, and it is not NIP for $r_\infty < r \leq 4$ (see Propositions 5.4 and 5.18 and Remark 5.19). Here $r_\infty \approx 3.5699$ is the Feigenbaum parameter, at which the cascade of period doublings ends. Right: for $r = 2.8$, two orbits converge to the fixed point $1 - 1/r$. The deep computation sends 0 and 1 to 0 and every other state to $1 - 1/r$.

We return to the logistic family for $3 < r \leq 4$ at the end of the section, where we complete its classification (see Proposition 5.18 and Remark 5.19).

5.3. Hereditarily separable but not metrizable. Bresenham's line algorithm in the plane, which we analyze in this subsection, lies at the level NIP_2 but not NIP_3 , immediately below the metrizable level. Below that level, no countable set of tests separates the deep computations (see Proposition 3.4), a test being a pair of an input and a feature (see Section 3.2). Rounding gives rise to spaces of deep computations of this kind, since an algorithm that rounds at each step takes a new discrete decision at each step.

Drawing a line. Fix an irrational number α with $0 < \alpha < 1$. To draw the line $y = \alpha x + \rho$ on a screen, Bresenham's algorithm [Bre65] moves one pixel to the right at each step and decides whether to move one pixel up. The n -th decision is $b_n(\rho) = \lfloor (n+1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor = \llbracket \{n\alpha + \rho\} \in [1 - \alpha, 1) \rrbracket$, where $\{\cdot\}$ denotes the fractional part, and the state of the algorithm is the error term ρ modulo 1. The step $\rho \mapsto \rho + \alpha$ is a rotation of the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, but the decision bit is not a continuous function of ρ . The standard remedy [MH40, Lot02] is to take the decision sequences themselves as states.

Let $\mathcal{O} = \{-k\alpha : k \in \mathbb{Z}\} \subseteq \mathbb{T}$, a countable dense set containing both endpoints 0 and $1 - \alpha$ of the decision interval. The *split circle* $\mathbb{T}_{\mathcal{O}}$ is obtained from $\mathbb{T}$ by replacing each $t \in \mathcal{O}$ by two points $t^- < t^+$, with the circular order topology; it is a compact metrizable space, homeomorphic to the Cantor set. The rotation by

α preserves $\mathcal{O}$ and lifts to a homeomorphism R of $\mathbb{T}_{\mathcal{O}}$ that preserves the labels $\pm$. The decision interval becomes clopen in $\mathbb{T}_{\mathcal{O}}$, and the coding map $\mathbb{T}_{\mathcal{O}} \rightarrow 2^{\mathbb{Z}}$, $y \mapsto (\llbracket R^n y \in [1 - \alpha, 1) \rrbracket)_{n \in \mathbb{Z}}$, is a homeomorphism onto the *Sturmian shift* X_{α} that conjugates R to the shift σ . The *Bresenham CCS* has states $L = X_{\alpha}$, the coordinate predicates $P_j(y) = y_j$ for $j \in \mathbb{Z}$, and the semigroup $\Gamma = \{\sigma^n : n \in \mathbb{N}\}$.

A point of $\mathbb{T}_{\mathcal{O}}$ lies *over* $t \in \mathbb{T}$ if the natural projection $\mathbb{T}_{\mathcal{O}} \rightarrow \mathbb{T}$ maps it to t . For $c \in \mathbb{T}$, let R_c^+ (respectively R_c^-) be the map of $\mathbb{T}_{\mathcal{O}}$ that sends a point over t to $(t + c)^+$ (respectively $(t + c)^-$) when $t + c \in \mathcal{O}$, and to $t + c$ otherwise. Through the coding map, we regard R_c^+ and R_c^- as maps of X_{α} . The *double arrow circle* $\mathbb{D}$ is $\mathbb{T} \times \{-, +\}$ with the lexicographic circular order and its order topology, and we also write c^- and c^+ for the points $(c, -)$ and $(c, +)$ of $\mathbb{D}$. The double arrow circle is compact, first countable, hereditarily separable and not metrizable [Eng89, 3.10.C].

Theorem 5.5. *The Bresenham CCS is NIP. Its deep computations are*

$$E(X_{\alpha}, \sigma) = \{\sigma^n : n \in \mathbb{N}\} \cup \{R_c^+, R_c^- : c \in \mathbb{T}\},$$

where each σ^n is isolated and $c^{\pm} \mapsto R_c^{\pm}$ is a homeomorphism of $\mathbb{D}$ onto the remaining part. In particular, the Bresenham CCS is NIP_2 but not NIP_3 .

Proof. To prove that the CCS is NIP, note that the predicates $P_j \circ \sigma^n$ are $\{0, 1\}$ -valued. Moreover, the j -th coordinate of $\sigma^n y$ is 1 exactly when $R^{n+j} y$ lies in the clopen arc $[1 - \alpha, 1)$ of $\mathbb{T}_{\mathcal{O}}$, that is, when y lies in an arc of $\mathbb{T}_{\mathcal{O}}$. A family of arcs of a circularly ordered set has VC dimension at most 3. By the remark at the end of Section 1.5, the characteristic functions of the members of such a family therefore have the NIP.

For the deep computations, let γ be the limit of σ^{n_i} along an ultrafilter, and let $c = \lim n_i \alpha$ in $\mathbb{T}$. If n_i is eventually constant, then γ is a power of σ . Otherwise, since the points $n_i \alpha$ are distinct, the ultrafilter decides whether $n_i \alpha$ approaches c from the right or from the left. In the first case, for every point y over t , the points $R^{n_i} y$ over $t + n_i \alpha$ approach $t + c$ strictly from the right. These points converge in $\mathbb{T}_{\mathcal{O}}$ to $(t + c)^+$ if $t + c \in \mathcal{O}$ and to $t + c$ otherwise, whence $\gamma = R_c^+$; the second case gives $\gamma = R_c^-$. Conversely, for every N , every c is approached from both sides by the points $n \alpha$ with $n \geq N$; hence every $R_c^{\pm}$ is a deep computation. The maps R_c^+ and R_c^- are distinct, since they differ at the points over $d - c$ with $d \in \mathcal{O}$. The same argument shows that $c_i^{\varepsilon_i} \rightarrow c^+$ in $\mathbb{D}$ implies $R_{c_i}^{\varepsilon_i} \rightarrow R_c^+$ pointwise, and symmetrically for c^- . The map $c^{\pm} \mapsto R_c^{\pm}$ is therefore a continuous bijection from the compact space $\mathbb{D}$ onto a Hausdorff space, hence a homeomorphism. Finally, fix $d \in \mathcal{O}$ and consider the two points $(d - n \alpha)^{\pm}$ of $\mathbb{T}_{\mathcal{O}}$, which σ^n sends to d^- and d^+ . Every map $R_c^{\pm}$ sends them to the same point, and so does every pointwise limit of σ^{m_i} with $m_i \rightarrow \infty$, by the argument above. Hence no net of such maps, and no net of maps σ^m with $m \neq n$ and m bounded, converges to σ^n , so σ^n is isolated. Thus $E(X_{\alpha}, \sigma)$ is the union of a countable set and a copy of $\mathbb{D}$; it is hereditarily separable and not metrizable. $\square$

The two deep computations R_c^+ and R_c^- are the two ways of drawing the line infinitely far away, where the accumulated error has reached c . The two maps agree except at the pixels where the line passes exactly through a corner, and at those pixels they apply the two opposite rounding conventions. Figure 6 shows the two conventions.

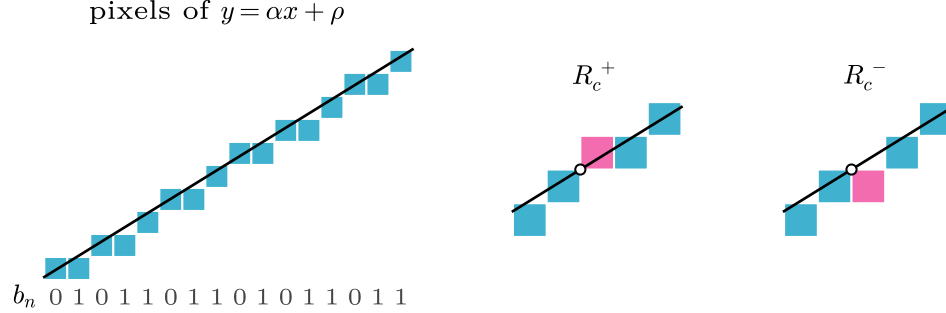


FIGURE 6. Left: the pixels drawn by Bresenham's algorithm for $\alpha = (\sqrt{5} - 1)/2$, with the decision bits b_n . Right: where the line passes exactly through a pixel corner, the deep computations R_c^+ and R_c^- break the tie in opposite directions; they agree on every other pixel.

The space $E(X_\alpha, \sigma)$ is the closure of a countable set, and it is therefore a separable Rosenthal compactum. Since this compactum is hereditarily separable and not metrizable, it contains, by Todorćević's trichotomy (see Theorem 3.3), a copy of the split Cantor set $S(2^{\mathbb{N}})$. For the action of $\mathbb{Z}$ generated by the shift, Glasner and Megrelishvili proved that Sturmian systems are tame₂ but not hereditarily non-sensitive [GM22, Corollary 8.3].

5.4. First countable but not hereditarily separable. When two rounding algorithms run in parallel, the computation descends a further level. We establish this descent here in the case of Bresenham's algorithm for lines in three-dimensional space, which lies at the level NIP_1 but not NIP_2 . At this level each deep computation is still separated from all the others by countably many tests of its own (see Proposition 3.4). The same proposition shows, however, that some finitely separated set of deep computations has the cardinality of the continuum.

Drawing a line in space. Let α and β be irrational numbers in $(0, 1)$ such that $1, \alpha, \beta$ are linearly independent over $\mathbb{Q}$. To draw a line in $\mathbb{Z}^3$ with direction $(1, \alpha, \beta)$, the three-dimensional Bresenham algorithm runs the planar algorithm in two coordinates at once. Its CCS has states $L = X_\alpha \times X_\beta$, the coordinate predicates of both factors, and the semigroup generated by $\sigma \times \sigma$.

The analysis of this algorithm uses a property of the double arrow circle $\mathbb{D}$ of Section 5.3 that keeps every compact metrizable subset of $\mathbb{D} \times \mathbb{D}$ countable. Recall that a *network* of a space is a family of subsets such that every neighborhood of a point contains a member of the family that contains the point. In particular, every base is a network.

Lemma 5.6. *Every subspace of the double arrow circle $\mathbb{D}$ that has a countable network, in particular every metrizable subspace, is countable.*

Proof. A metrizable subspace of $\mathbb{D}$ is separable, since $\mathbb{D}$ is hereditarily separable, and it therefore has a countable base, which is a countable network. Let $M \subseteq \mathbb{D}$ have a countable network $\{N_k : k \in \mathbb{N}\}$. For each $c \in \mathbb{T}$ with $c^+ \in M$, the open arc $[c^+, (c + \frac{1}{2})^-)$ is a neighborhood of c^+ in $\mathbb{D}$, and we may choose $k(c) \in \mathbb{N}$ with $c^+ \in N_{k(c)} \subseteq [c^+, (c + \frac{1}{2})^-)$. If $k(c) = k(c')$, then the point c'^+ lies in $[c^+, (c + \frac{1}{2})^-)$

and the point c^+ in $[c'^+, (c' + \frac{1}{2})^-]$, whence both $c' - c$ and $c - c'$ lie in $[0, \frac{1}{2})$ modulo 1, which forces $c = c'$. Thus the map $c \mapsto k(c)$ is injective, and the set M contains only countably many points c^+ . The same argument, applied to the arcs $((c - \frac{1}{2})^+, c^-]$, shows that M contains only countably many points c^- . $\square$

Corollary 5.7. *Every compact metrizable subspace of $\mathbb{D} \times \mathbb{D}$ is countable.*

Proof. Let $C \subseteq \mathbb{D} \times \mathbb{D}$ be compact and metrizable, and let $\mathcal{N}$ be a countable base of C . Since the two projections are continuous, the images of the members of $\mathcal{N}$ under each projection form a countable network of the image of C . By Lemma 5.6, both images are countable, and so is C , which is contained in their product. $\square$

For $c' \in \mathbb{T}$, we also write $R_{c'}^+$ and $R_{c'}^-$ for the maps of X_β defined as in Section 5.3, with β in place of α .

Theorem 5.8. *The three-dimensional Bresenham CCS is NIP_1 but not NIP_2 . Its space of deep computations $E = E(X_\alpha \times X_\beta, \sigma \times \sigma)$ is the union of the set of iterates $(\sigma \times \sigma)^n$, $n \in \mathbb{N}$, and of the closed subspace*

$$\{(R_c^\varepsilon, R_{c'}^{\varepsilon'}) : c, c' \in \mathbb{T}, \varepsilon, \varepsilon' \in \{+, -\}\},$$

which is homeomorphic to $\mathbb{D} \times \mathbb{D}$. Each of these iterates is isolated in E . The space E contains a copy of $\hat{D}(S(2^\mathbb{N}))$, and it contains no copy of $D(2^\mathbb{N})$, of $\hat{D}(2^\mathbb{N})$ or of $A(2^\mathbb{N})$.

Proof. Each predicate is a coordinate of one factor, so the CCS is NIP by Theorem 5.5. Every deep computation is the limit ξ of the iterates $(\sigma \times \sigma)^n$ along an ultrafilter $\mathcal{U}$ on $\mathbb{N}$. The coordinates of ξ are then the limits along $\mathcal{U}$ of the iterates σ^n on X_α and on X_β . If $\mathcal{U}$ is principal, the deep computation ξ is an iterate; otherwise the proof of Theorem 5.5 shows that ξ is a pair $(R_c^\varepsilon, R_{c'}^{\varepsilon'})$. By Kronecker's theorem, the points $(n\alpha, n\beta)$ with $n \geq N$ are dense in $\mathbb{T}^2$ for every N . Every point (c, c') is therefore approached by these points from inside each of the four open quadrants at (c, c') . As in the proof of Theorem 5.5, it follows that every pair $(R_c^\varepsilon, R_{c'}^{\varepsilon'})$ with $\varepsilon, \varepsilon' \in \{+, -\}$ is a deep computation. By Theorem 5.5, applied to each factor, the set of these pairs is homeomorphic to $\mathbb{D} \times \mathbb{D}$, and it is therefore compact and closed in E . Each iterate $(\sigma \times \sigma)^n$ is isolated in E , since σ^n is isolated in $E(X_\alpha, \sigma)$ and is the first coordinate of no other deep computation. The space E is a closed subspace of $E(X_\alpha, \sigma) \times E(X_\beta, \sigma)$, which is first countable as a product of two first countable spaces. Hence the CCS is NIP_1 , and no copy of the space $A(2^\mathbb{N})$, which is not first countable, lies in E . Consider the pairs (R_c^+, R_{-c}^+) , for $c \in \mathbb{T}$. For $0 < \delta < \frac{1}{2}$, the set $[c^+, (c + \delta)^-] \times [(-c)^+, (-c + \delta)^-]$ is a neighborhood of $(c^+, (-c)^+)$ in $\mathbb{D} \times \mathbb{D}$ that contains $(s^+, (-s)^+)$ only if $c \leq s < c + \delta$ and $-c \leq -s < -c + \delta$, that is, only if $s = c$. These pairs therefore form an uncountable discrete subspace, and the space of deep computations is not hereditarily separable; thus the CCS is not NIP_2 . A compact metrizable subspace of E meets the copy of $\mathbb{D} \times \mathbb{D}$ in a compact metrizable set, which is countable by Corollary 5.7, and the subspace itself is therefore countable. Since the non-isolated points of $D(2^\mathbb{N})$ form a copy of the Cantor set, neither $D(2^\mathbb{N})$ nor the compactum $\hat{D}(2^\mathbb{N})$ that contains it embeds into E . Finally, by Theorem 3.1 the space E , being the closure of a countable set, is a separable Rosenthal compactum that is first countable but not hereditarily separable. By Theorem 3.3, this compactum contains a copy of $D(2^\mathbb{N})$ or of $\hat{D}(S(2^\mathbb{N}))$, necessarily of the latter. $\square$

The uncountable discrete set consists of the deep computations that break ties upward in both coordinates at points where the two accumulated errors are opposite. Each of the two planar algorithms is NIP_2 , and drawing in two directions at once leaves the class. In just the same way, the Sorgenfrey plane shows that a product of two hereditarily separable spaces need not be hereditarily separable. Glasner and Megrelishvili show similarly that a product of two tame_2 systems need not be tame_2 [GM22, Proposition 4.1].⁹ A second example at this level arises from the Euclidean algorithm (see Proposition 5.13). This example is a countable family of elements of the group generated by the steps of the Euclidean algorithm, and its deep computations include a copy of $D(2^{\mathbb{N}})$. Whether a single iterated map can realize the level with such a copy is Question 7.3.

Remark 5.9. A copy of $\hat{D}(S(2^{\mathbb{N}}))$ in E whose dense set of isolated points consists of iterates can be exhibited explicitly. For $t \in 2^{<\mathbb{N}}$, let x_t be the right endpoint of the closed interval of length $3^{-|t|}$ indexed by t in the construction of the middle-thirds Cantor set, and put $y_t = \frac{3}{2}3^{-|t|} - x_t$. By Kronecker's theorem, there are pairwise distinct $n_t \in \mathbb{N}$ such that, modulo 1, the point $n_t\alpha$ lies in $(x_t, x_t + 3^{-|t|-1})$ and the point $n_t\beta$ in $(y_t, y_t + 3^{-|t|-1})$. For $b \in 2^{\mathbb{N}}$, let x_b be the point of the Cantor set coded by b , the limit of the points $x_{b|n}$. As t runs through the initial segments of b , the iterates $(\sigma \times \sigma)^{n_t}$ converge to $(R_{x_b}^+, R_{-x_b}^+)$. Along words t that converge to b , are not initial segments of b and satisfy $t \frown 0^\infty < b$, the iterates converge to $(R_{x_b}^-, R_{-x_b}^+)$. Finally, along words that converge to b and satisfy $t \frown 0^\infty > b$, the iterates converge to $(R_{x_b}^+, R_{-x_b}^-)$. The three convergences just described are the convergence patterns of the family D_7 of Section 3.3. Since the iterates are isolated in E , a lemma of Argyros, Dodos and Kanellopoulos [ADK08, Lemma 19] shows that the family $\{(\sigma \times \sigma)^{n_t} : t \in 2^{<\mathbb{N}}\}$ is equivalent to D_7 and therefore has a closure homeomorphic to $\hat{D}(S(2^{\mathbb{N}}))$. These three patterns are drawn in the left panel of Figure 8; its right panel shows the copy of $\hat{D}(2^{\mathbb{N}})$ that maps built from the Euclidean algorithm produce in Section 5.5.

5.5. NIP but not first countable: the Euclidean algorithm. The last tame level, NIP but not NIP_1 , is reached by the Euclidean algorithm, acting on the real projective line through the group generated by its steps. At this level some deep computation is not determined by countably many tests of its own, that is, by countably many pairs of an input and a feature (see Proposition 3.4). A constant deep computation has this property whenever uncountably many exact equality tests, maps each of which agrees with the constant map at every input but one, are also deep computations. Indeed, countably many tests involve only countably many inputs and therefore cannot distinguish the constant map from an exact equality

⁹The double arrow space is the set of the points $p^- < p^+$, for $p \in [0, 1]$, with its order topology. Glasner and Megrelishvili [GM22, Remark 4.2] assert that the square of this space contains a copy of $D(2^{\mathbb{N}})$, namely the set of the points $(p^+, (1-p)^+)$ and $(p^-, (1-p)^+)$ with p in the Cantor set. The set of these points is not closed, and hence not compact. For instance, if p is the limit of a decreasing sequence $(x_k)_{k \in \mathbb{N}}$ of points of the Cantor set, then the points $(x_k^-, (1-x_k)^+)$ of this set converge to $(p^+, (1-p)^-)$, a point outside it. Moreover, the square of the double arrow space contains no copy of $D(2^{\mathbb{N}})$ at all. Indeed, with intervals in place of arcs, the arguments of Lemma 5.6 and Corollary 5.7 apply to the double arrow space, and the non-isolated points of $D(2^{\mathbb{N}})$ form an uncountable compact metrizable set. Proposition 4.1 of Glasner and Megrelishvili does not depend on their Remark 4.2.

test whose exceptional input is not among those inputs. The Euclidean algorithm gives rise to equality tests of precisely this kind among its deep computations.

Let $L = \mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$, with the predicates P_1, P_3 of Section 4.3, and let $\Gamma = \text{PGL}(2, \mathbb{Z})$ be the group of Möbius maps $x \mapsto (ax + b)/(cx + d)$ with integer coefficients and $ad - bc = \pm 1$. It is generated by the steps $x \mapsto x + 1$ and $x \mapsto 1/x$ of the Euclidean algorithm, and the maps that the continued fraction algorithm composes are elements of Γ . We call $(\mathbb{RP}^1, (P_1, P_3), \Gamma)$ the *Euclidean CCS*. It is a dynamical CCS whose semigroup is a group, and its deep computations form $E(\mathbb{RP}^1, \Gamma)$.

For an irrational number $\eta = [a_0; a_1, a_2, \dots]$, let p_k/q_k be its convergents, put $p_{-1} = 1$ and $q_{-1} = 0$, and let $\eta_{k+1} = [a_{k+1}; a_{k+2}, \dots]$ be its complete quotients. Let M_k be the Möbius map $t \mapsto (p_k t + p_{k-1})/(q_k t + q_{k-1})$, which belongs to Γ because $p_k q_{k-1} - p_{k-1} q_k = (-1)^{k+1}$. Writing $A_j(t) = a_j + 1/t$, we have $M_k = A_0 \circ A_1 \circ \dots \circ A_k$, and hence the inverse map is $M_k^{-1} = A_k^{-1} \circ \dots \circ A_0^{-1}$, with $A_j^{-1}(t) = 1/(t - a_j)$. Thus M_k^{-1} is the map that the first $k+1$ steps of the continued fraction algorithm for η apply. In these terms, the classical identity $\eta = M_k(\eta_{k+1})$ states that these steps send η to its complete quotient η_{k+1} .

Lemma 5.10. *Let η be an irrational number, and let $x \in \mathbb{RP}^1$ be distinct from η . Then $M_k^{-1}(x) + q_{k-1}/q_k \rightarrow 0$ as $k \rightarrow \infty$, while $M_k^{-1}(\eta) = \eta_{k+1} > 1$.*

Proof. Solving $M_k(t) = x$ gives $M_k^{-1}(\infty) = -q_{k-1}/q_k$ and, for real x ,

$$M_k^{-1}(x) = \frac{q_{k-1}x - p_{k-1}}{p_k - q_k x} = -\frac{q_{k-1}}{q_k} \cdot \frac{x - p_{k-1}/q_{k-1}}{x - p_k/q_k}$$

for $k \geq 1$ (for such k we have $q_{k-1} \geq 1$), where the value is ∞ if $x = p_k/q_k$. The convergents of an irrational number are pairwise distinct, so x equals p_k/q_k for at most one k ; for all larger k the formula gives a real number. Both convergents tend to $\eta \neq x$, so the last fraction tends to 1, and $0 < q_{k-1}/q_k \leq 1$ for $k \geq 1$; hence $M_k^{-1}(x) + q_{k-1}/q_k \rightarrow 0$. The identity $M_k^{-1}(\eta) = \eta_{k+1}$ is the classical one, and $\eta_{k+1} > a_{k+1} \geq 1$ because η_{k+1} is irrational. $\square$

Let $\varphi = (1 + \sqrt{5})/2$ be the golden ratio. Let G be the set of irrational numbers $\eta = [0; a_1, a_2, \dots]$ whose sequence of partial quotients is a concatenation $1^1 d_1 1^2 d_2 1^3 d_3 \dots$, where 1^n denotes n consecutive ones and each $d_n \in \{1, 2\}$. The block 1^n occupies the indices $s_n, \dots, s_n + n - 1$ with $s_n = n(n+1)/2$. Distinct choices of (d_n) give distinct numbers, by the uniqueness of continued fraction expansions, so G has the cardinality of the continuum.

Proposition 5.11. (i) *For every $\eta \in G$, the map δ_η that sends η to φ and every other point of $\mathbb{RP}^1$ to $-1/\varphi$ is a deep computation of the Euclidean CCS.*

(ii) *The constant map c with value $-1/\varphi$ is a deep computation, and it has no countable base of neighborhoods in $E(\mathbb{RP}^1, \Gamma)$.*

Consequently $E(\mathbb{RP}^1, \Gamma)$ is not first countable.

Proof. (i) Let $\eta \in G$ be given; for each n , let $k_n = s_n + \lfloor n/2 \rfloor$ be the middle index of the block 1^n , and put $m_n = \lfloor n/2 \rfloor - 1$. Then $a_{k_n} = a_{k_n-1} = \dots = a_{k_n-m_n} = 1$ and $a_{k_n+1} = \dots = a_{k_n+m_n} = 1$, while $m_n \rightarrow \infty$. The two quantities that occur in Lemma 5.10 are continued fractions: $q_{k-1}/q_k = [0; a_k, a_{k-1}, \dots, a_1]$ and $\eta_{k+1} = [a_{k+1}; a_{k+2}, \dots]$. Suppose that two continued fractions, finite or infinite, $[c_0; c_1, \dots]$ and $[c'_0; c'_1, \dots]$ satisfy $c_j = c'_j$ for $j < m$, with $c_1 = \dots = c_{m-1} = 1$.

Then they share the convergents of index at most $m - 1$, and both lie between the last two of them (or equal one of them). It follows that the two continued fractions differ by at most $1/(q_{m-2}q_{m-1}) \leq 1/(F_{m-1}F_m)$, whatever their remaining partial quotients are. Here F_j is the j th Fibonacci number, with $F_1 = F_2 = 1$. The hypothesis holds with $m = m_n$ for the pair $[0; a_{k_n}, \dots, a_1]$, $[0; 1, 1, \dots]$ and for the pair $[a_{k_n+1}; a_{k_n+2}, \dots]$, $[1; 1, 1, \dots]$. Hence $q_{k_n-1}/q_{k_n} \rightarrow [0; 1, 1, \dots] = 1/\varphi$ and $\eta_{k_n+1} \rightarrow [1; 1, 1, \dots] = \varphi$, and Lemma 5.10 shows that the maps $M_{k_n}^{-1} \in \Gamma$ converge pointwise to δ_η , at $x = \infty$ as well.

(ii) Let (η_j) be a sequence of distinct elements of G . For each $x \in \mathbb{RP}^1$, at most one j has $\eta_j = x$, so $\delta_{\eta_j}(x) = -1/\varphi$ for all but at most one j . Hence $\delta_{\eta_j} \rightarrow c$ pointwise, and $c \in E(\mathbb{RP}^1, \Gamma)$ because this set is closed. Now let $(U_i)_{i \in \mathbb{N}}$ be open neighborhoods of c in $E(\mathbb{RP}^1, \Gamma)$. Each U_i contains a basic neighborhood $\{f : d(f(x), c(x)) < \varepsilon_i \text{ for } x \in F_i\}$ with $F_i \subseteq \mathbb{RP}^1$ finite, $\varepsilon_i > 0$ and d a metric on $\mathbb{RP}^1$. Since the set $F = \bigcup_i F_i$ is countable, some $\eta \in G$ lies outside F ; the map δ_η then agrees with c on F and hence lies in $\bigcap_i U_i$, although $\delta_\eta \neq c$. Thus no countable family of neighborhoods of c has intersection $\{c\}$. In a Hausdorff space, the intersection of a base of neighborhoods of c is $\{c\}$, so c has no countable base of neighborhoods. $\square$

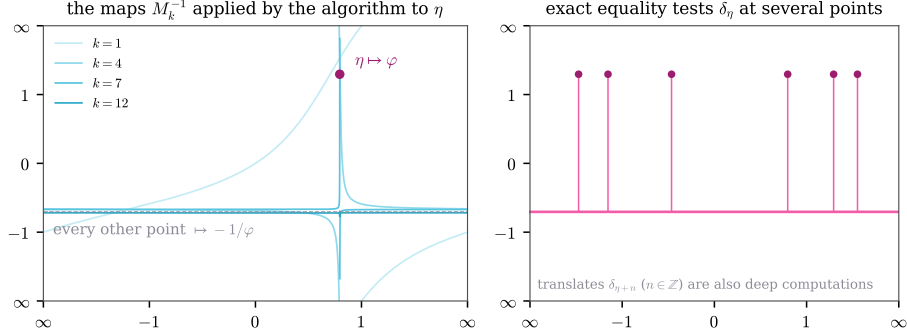


FIGURE 7. The Euclidean algorithm, drawn on the projective line (each axis runs from ∞ through 0 back to ∞). Left: the maps M_k^{-1} that the continued fraction algorithm applies to a number η whose partial quotients contain longer and longer runs of ones. As k grows along the middles of the runs, these maps send every point other than η near $-1/\varphi$, and η itself near φ . Right: the limits, exact equality tests δ_η , at several points. Every neighborhood of the constant map $-1/\varphi$ contains all but finitely many of the uncountably many tests δ_η , so the space of deep computations is not first countable (see Proposition 5.11).

Corollary 5.12. *The Euclidean CCS is NIP but not NIP_1 . Its space of deep computations contains the copy $\{\delta_\eta : \eta \in G\} \cup \{c\}$ of the Alexandroff compactification $A(2^{\mathbb{N}})$.*

Proof. Ellis proved that the action of $\text{GL}(2, \mathbb{R})$ on the projective line is tame [Ell93] (see [GM22, Example 5.7]). For every continuous $f : \mathbb{RP}^1 \rightarrow \mathbb{R}$, the family $\{f \circ \gamma :$

$\gamma \in \Gamma$ is contained in the corresponding family for $\mathrm{GL}(2, \mathbb{R})$, so it has the NIP. Hence the action of Γ is tame, and, by Proposition 5.1, the Euclidean CCS is NIP. By Proposition 5.11, it is not NIP_1 . Each δ_η is isolated in $\{\delta_\eta : \eta \in G\} \cup \{c\}$, since δ_η is the only member of this set that sends η near φ . Moreover, as in the proof of Proposition 5.11, every neighborhood of c contains all but finitely many of the maps δ_η . The set $\{\delta_\eta : \eta \in G\} \cup \{c\}$ is therefore the one-point compactification of a discrete set of cardinality $2^{\aleph_0}$, and hence it is homeomorphic to $A(2^{\mathbb{N}})$ (see Section 1.4). $\square$

The deep computation δ_η is an exact equality test (see Figure 7): it takes one value if the input is exactly η and another value otherwise. It is the pointwise limit of maps that the continued fraction algorithm applies to η , when these maps are applied to all inputs. The mechanism is that of convergence groups. These are groups of homeomorphisms of a compact space in which every sequence of distinct elements has a subsequence converging to a constant map uniformly on the compact subsets of the complement of some point. By Lemma 5.10, the maps M_k^{-1} draw the images of all points other than η together, near $-q_{k-1}/q_k \in [-1, 0)$, while sending η itself into $(1, \infty)$. In the language of convergence groups, every irrational number η is thus a conical limit point of Γ : some sequence in Γ sends η to one limit and every other point to another. The choice of G makes the two limits the same for uncountably many η . Glasner and Megrelishvili use the same device to prove that an infinite countable uniform convergence group acting on a compact metric space with infinite limit set is tame but not tame_1 [GM22, Proposition 6.1 and Theorem 6.5]. Their theorem does not cover Γ , which contains parabolic elements and is not a uniform convergence group on $\mathbb{RP}^1$. Ellis's example, the projective action of the whole group $\mathrm{GL}(2, \mathbb{R})$, is also tame but not tame_1 [GM22, Example 5.7]. The Euclidean CCS shows that the countable group generated by the two steps of the algorithm already reaches this level.

Iterating a single element of Γ cannot produce δ_η . For $g \in \Gamma$, the closure of $\{g^n : n \in \mathbb{Z}\}$ in $(\mathbb{RP}^1)^{\mathbb{RP}^1}$ is countable, hence metrizable, and it does not contain δ_η . Indeed, if g has finite order, the set of its powers is finite. If g is parabolic, with fixed point x_0 , then g^n converges pointwise to the constant map x_0 as $n \rightarrow \pm\infty$. If g is hyperbolic, with attracting and repelling fixed points x_a and x_r , then g^n converges as $n \rightarrow +\infty$ to the map that fixes x_r and sends every other point to x_a . As $n \rightarrow -\infty$, the two fixed points exchange their roles, and g^n converges to the map that fixes x_a and sends every other point to x_r . If g reverses orientation and has infinite order, then its trace t is nonzero, whence the square g^2 has trace $t^2 + 2 > 2$ and is a hyperbolic map with the same fixed points as g . The odd powers $g^{2m+1} = g \circ g^{2m}$ have the same two limits as the even ones. In every case the closure consists of the powers of g , which are homeomorphisms, together with at most two limit maps. Each of these limit maps is constant or fixes the unique point at which that map differs from a constant. The map δ_η is neither a power of g nor such a limit map, since it is not injective and $\delta_\eta(\eta) = \varphi \neq \eta$.

A copy of the Alexandroff duplicate. Among the deep computations of the Euclidean CCS, the device of Proposition 5.11 also produces a copy of the Alexandroff duplicate $D(2^{\mathbb{N}})$. With this copy comes a countable set of computations at the level NIP_1 but not NIP_2 . Instead of sending uncountably many irrational numbers to the same pair of values, we now let the value that a deep computation takes away from its exceptional point vary continuously with that point. Let $\mathcal{C}$ be the set of irrational

numbers $[0; \omega_1, \omega_2, \dots]$ with $\omega_n \in \{1, 2\}$ for every $n \geq 1$, a Cantor set homeomorphic to $\{1, 2\}^{\mathbb{N}}$ through continued fraction expansions. For $\omega = (\omega_1, \omega_2, \dots) \in \{1, 2\}^{\mathbb{N}}$ we write $[0; \omega]$ for the point $[0; \omega_1, \omega_2, \dots]$ of $\mathcal{C}$. For a finite or infinite sequence ω over the alphabet $\{1, 2\}$ and for each positive integer n not exceeding its length, let $\Lambda_n(\omega)$ be the block $(\omega_n, \omega_{n-1}, \dots, \omega_1, 1^n)$. This block consists of the first n terms of ω in reverse order, followed by n ones. For $\omega \in \{1, 2\}^{\mathbb{N}}$, let $\eta(\omega)$ be the irrational number in $(0, 1)$ whose partial quotients after the initial 0 form the concatenation $\Lambda_1(\omega)\Lambda_2(\omega)\Lambda_3(\omega)\dots$. Since the terms $\omega_1, \dots, \omega_n$ can be recovered from $\Lambda_n(\omega)$, the map $\omega \mapsto \eta(\omega)$ is injective. For a finite word $\mathbf{w} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$ of positive integers, let $M_{\mathbf{w}}$ be the Möbius map $x \mapsto [0; \mathbf{w}_1, \dots, \mathbf{w}_{k-1}, \mathbf{w}_k + 1/x]$, and let $M_{\emptyset}(x) = 1/x$ for the empty word. For an irrational number $[0; a_1, a_2, \dots]$ in $(0, 1)$, the map M_k introduced before Lemma 5.10 is then $M_{(a_1, \dots, a_k)}$. Write $\{1, 2\}^{<\mathbb{N}}$ for the set of finite words over $\{1, 2\}$. For a word $t = (t_1, \dots, t_m)$ in this set, let $t^\sharp$ be the word $\Lambda_1(t) \cdots \Lambda_{m-1}(t)(t_m, t_{m-1}, \dots, t_1)$ of length m^2 , and put $\gamma_t = M_{t^\sharp}^{-1} \in \Gamma$. Finally, for $\omega \in \{1, 2\}^{\mathbb{N}}$, let τ_ω be the map that sends $\eta(\omega)$ to φ and every other point of $\mathbb{RP}^1$ to $-[0; \omega]$, and let c_ω be the constant map with value $-[0; \omega]$. For the constant sequence $\omega = (1, 1, 1, \dots)$, the map c_ω is the map c of Proposition 5.11.

Proposition 5.13. (i) *The closure of the countable set $\Delta_{\mathcal{C}} = \{\gamma_t : t \in \{1, 2\}^{<\mathbb{N}}\}$ is $\Delta_{\mathcal{C}} \cup \{\tau_\omega : \omega \in \{1, 2\}^{\mathbb{N}}\} \cup \{c_\omega : \omega \in \{1, 2\}^{\mathbb{N}}\}$, and it is homeomorphic to $\hat{D}(2^{\mathbb{N}})$ through a homeomorphism that maps $\{\tau_\omega : \omega \in \{1, 2\}^{\mathbb{N}}\} \cup \{c_\omega : \omega \in \{1, 2\}^{\mathbb{N}}\}$ onto $D(2^{\mathbb{N}})$. In particular, $E(\mathbb{RP}^1, \Gamma)$ contains a copy of $D(2^{\mathbb{N}})$, and the set $\Delta_{\mathcal{C}}$ of computations of the Euclidean CCS is NIP_1 but not NIP_2 .*
(ii) *Every hereditarily separable compact subset of $E(\mathbb{RP}^1, \Gamma)$ is metrizable. In particular, $E(\mathbb{RP}^1, \Gamma)$ contains no copy of $S(2^{\mathbb{N}})$ and none of $\hat{D}(S(2^{\mathbb{N}}))$.*

Proof. (i) We use two facts from the proofs of Lemma 5.10 and Proposition 5.11. First, consider two continued fractions, finite or infinite, with positive integer partial quotients that agree in their terms of index less than m . Both lie between two consecutive convergents of index $m-2$ and $m-1$, and hence differ by at most $1/(F_{m-1}F_m)$. Second, the formula for M_k^{-1} in the proof of Lemma 5.10 uses only the terms $a_0, \dots, a_k$: if $k \geq 1$ and p_i/q_i are the convergents of $[a_0; a_1, \dots, a_k]$, the map $M(x) = [a_0; a_1, \dots, a_{k-1}, a_k + 1/x]$ satisfies

$$(5.14) \quad M^{-1}(x) = -\frac{q_{k-1}}{q_k} \cdot \frac{x - p_{k-1}/q_{k-1}}{x - p_k/q_k} \quad (x \in \mathbb{R}), \quad M^{-1}(\infty) = -\frac{q_{k-1}}{q_k},$$

where $q_{k-1}/q_k = [0; a_k, a_{k-1}, \dots, a_1]$.

Limits along initial segments. Let $\omega \in \{1, 2\}^{\mathbb{N}}$ be given, and put $\eta = \eta(\omega) = [0; a_1, a_2, \dots]$. For $m \geq 1$, the word $(\omega \upharpoonright m)^\sharp$ consists of the first m^2 partial quotients of η , whence $\gamma_{\omega \upharpoonright m} = M_{(\omega \upharpoonright m)^\sharp}^{-1}$. The terms $a_{m^2}, a_{m^2-1}, \dots, a_{m^2-m+1}$ are $\omega_1, \omega_2, \dots, \omega_m$, and $a_{m^2+1} = \dots = a_{m^2+m} = 1$. By the first fact, as $m \rightarrow \infty$, the quotient $q_{m^2-1}/q_{m^2} = [0; a_{m^2}, \dots, a_1]$ tends to $[0; \omega]$ and the complete quotient η_{m^2+1} tends to $[1; 1, 1, \dots] = \varphi$. Lemma 5.10 therefore shows that the maps $\gamma_{\omega \upharpoonright m}$ converge pointwise to τ_ω .

Limits along other words. Let $(t_j)_{j \in \mathbb{N}}$ be a sequence of words that converges to ω , none of which is an initial segment of ω . For each j , let n_j be the length of the longest common initial segment of t_j and ω , and observe that $n_j \rightarrow \infty$ because $t_j \rightarrow \omega$. To show that $\gamma_{t_j} \rightarrow c_\omega$ pointwise, we write $t_j^\sharp = (\mathbf{w}_1, \dots, \mathbf{w}_k)$, where $k = |t_j|^2$, and we let $\ell = \ell_j = n_j(n_j+1)+1$. The blocks $\Lambda_1(t_j), \dots, \Lambda_{n_j}(t_j)$ coincide

with $\Lambda_1(\omega), \dots, \Lambda_{n_j}(\omega)$, and the next term of $t_j^\sharp$ is the term of index $n_j + 1$ of t_j . Hence $w_i = a_i$ for $i < \ell$, while w_ℓ and a_ℓ are distinct elements of $\{1, 2\}$. Moreover, since $t_j^\sharp$ ends with the terms of t_j in reverse order, the terms $w_k, w_{k-1}, \dots$ begin with $\omega_1, \dots, \omega_{n_j}$, and $k - \ell \geq (n_j + 1)^2 - \ell = n_j$. Let $x \in \mathbb{RP}^1$ be distinct from η . The convergents of $[0; w_1, \dots, w_k]$ of index at least $\ell - 1$ agree with η in their terms of index less than ℓ . By the first fact, the convergents of index $k - 1$ and k therefore tend to η as $j \rightarrow \infty$, while $q_{k-1}/q_k = [0; w_k, \dots, w_1]$ tends to $[0; \omega]$. Formula (5.14) then shows that $\gamma_{t_j}(x) \rightarrow -[0; \omega]$ for every such x . At the point η we write γ_{t_j} as the composite of the map $x \mapsto 1/x$ followed by the maps $x \mapsto 1/(x - w_i)$, for $i = 1, \dots, k$ in this order. We then apply these $k + 1$ factors to η one at a time. The first ℓ factors send η to its complete quotient $\eta_\ell = [a_\ell; a_{\ell+1}, \dots]$, which lies in $(a_\ell, a_\ell + 1)$. The next factor sends η_ℓ to $z_j = 1/(\eta_\ell - w_\ell)$, which lies in $(\frac{1}{2}, 1)$ if $(a_\ell, w_\ell) = (2, 1)$ and in $(-\infty, -1)$ if $(a_\ell, w_\ell) = (1, 2)$; in either case $z_j < 1$. The remaining factors form the inverse of the map $M'(x) = [w_{\ell+1}; w_{\ell+2}, \dots, w_{k-1}, w_k + 1/x]$, and therefore $\gamma_{t_j}(\eta) = M'^{-1}(z_j)$. Let p'_i/q'_i be the convergents of $[w_{\ell+1}; w_{\ell+2}, \dots, w_k]$, the last of which has index $i = k - \ell - 1 \geq n_j - 1$. Every convergent of index at least 1 has the form $w_{\ell+1} + 1/y$ with $1 \leq y \leq 3$, since all its partial quotients lie in $\{1, 2\}$. Each of these convergents is therefore at least $\frac{4}{3}$, and its distance to z_j exceeds $\frac{1}{3}$. The convergents of index $i - 1$ and i differ by $1/(q'_{i-1}q'_i) \leq 1/(F_i F_{i+1})$, and $q'_{i-1}/q'_i = [0; w_k, \dots, w_{\ell+2}]$ tends to $[0; \omega]$, since its first $n_j - 1$ terms are $\omega_1, \dots, \omega_{n_j-1}$. By (5.14), applied to M' , we conclude that $\gamma_{t_j}(\eta) \rightarrow -[0; \omega]$, and hence that $\gamma_{t_j} \rightarrow c_\omega$.

The closure. By Corollary 5.12 and Theorem 3.1, the space $E(\mathbb{RP}^1, \Gamma)$ is a Rosenthal compactum. Consequently, every point of the closure of $\Delta_{\mathcal{C}}$ that does not belong to $\Delta_{\mathcal{C}}$ is the limit of a sequence (γ_{t_j}) with pairwise distinct words t_j [BFT78, Theorem 3F]. A compactness argument yields a subsequence of (t_j) that converges to some $\omega \in \{1, 2\}^{\mathbb{N}}$. A further subsequence consists either of initial segments of ω or of words that are not initial segments of ω . By the two preceding paragraphs, the corresponding maps converge to τ_ω in the first case and to c_ω in the second. Both kinds of limit occur for every ω , and the closure of $\Delta_{\mathcal{C}}$ is therefore the set described in (i).

The homeomorphism. Identify $\{1, 2\}^{\mathbb{N}}$ with $2^{\mathbb{N}}$ and $\{1, 2\}^{<\mathbb{N}}$ with $2^{<\mathbb{N}}$ by replacing the letters 1 and 2 with 0 and 1. Under this identification, define a map Θ on $\hat{D}(2^{\mathbb{N}})$ by $\Theta(v_t, t \frown 0^\infty) = \gamma_t$, $\Theta(\delta_\omega, \omega) = \tau_\omega$ and $\Theta(\mathbf{0}, \omega) = c_\omega$. The map Θ is surjective by the preceding paragraph. To see that Θ is injective, note first that the constant maps c_ω are pairwise distinct, because distinct sequences ω give distinct numbers $[0; \omega]$. For the same reason, the maps τ_ω are pairwise distinct, and none of them is constant. Neither kind of map is injective, whereas the maps γ_t are homeomorphisms. Finally, the maps γ_t are pairwise distinct, because $t^\sharp$ determines t and the Möbius map M_w determines the word w . Indeed, a rational number in $[0, 1]$ has at most two expansions $[0; w_1, \dots, w_k]$, whose lengths differ by one. The number $M_w(\infty) = [0; w_1, \dots, w_k]$ and the determinant $(-1)^{k+1}$ of a matrix that represents M_w therefore determine w . Since $\hat{D}(2^{\mathbb{N}})$ is first countable, the continuity of Θ may be tested on sequences. A sequence that converges to a point (δ_ω, ω) consists, apart from that point, of points $(v_t, t \frown 0^\infty)$ with t an initial segment of ω , whose lengths tend to infinity. By the paragraph on limits along initial segments, the image of this sequence under Θ converges to τ_ω . A sequence that converges to a point $(\mathbf{0}, \omega)$ consists, apart from that point, of points of at most three kinds. Points of the first kind have the form $(\mathbf{0}, \omega')$, and points of the second

kind the form $(\delta_{\omega'}, \omega')$, with $\omega' \rightarrow \omega$ and $\omega' \neq \omega$. Points of the third kind have the form $(v_t, t \frown 0^\infty)$, with t converging to ω and eventually not an initial segment of ω . The images of the points of the first kind converge to c_ω because $[0; \omega']$ depends continuously on ω' . The same holds for the images of the points of the third kind, by the paragraph on limits along other words. For the second kind, observe that the equality $\tau_{\omega'}(x) = -[0; \omega']$ fails at a given point $x \in \mathbb{RP}^1$ only if $\eta(\omega') = x$, which happens for at most one ω' . The images of the points of the second kind therefore converge to c_ω as well. Hence Θ is a continuous bijection of a compact space onto a Hausdorff space, and therefore a homeomorphism; it maps $D(2^\mathbb{N})$ onto $\{\tau_\omega : \omega \in \{1, 2\}^\mathbb{N}\} \cup \{c_\omega : \omega \in \{1, 2\}^\mathbb{N}\}$. The set $\Delta_{\mathcal{C}}$ is countable, and it has the NIP because Γ has it (see Corollary 5.12). The closure of $\Delta_{\mathcal{C}}$, a copy of $\hat{D}(2^\mathbb{N})$, is first countable but not hereditarily separable (see Section 3.2), and $\Delta_{\mathcal{C}}$ is therefore NIP_1 but not NIP_2 .

(ii) We first describe the deep computations that are not computations. For $y, z, w \in \mathbb{RP}^1$, let $\zeta_{y,z,w}$ be the map that sends z to w and every other point to y . Let $\xi \in E(\mathbb{RP}^1, \Gamma) \setminus \Gamma$ be given; since $E(\mathbb{RP}^1, \Gamma)$ is a Rosenthal compactum, the map ξ is the limit of a sequence $(g_n)_{n \in \mathbb{N}}$ of pairwise distinct elements of Γ . Represent each g_n by an integer matrix $\mathbf{G}_n$ with determinant ± 1 . Since only finitely many integer matrices have norm below a given bound, the norms $\|\mathbf{G}_n\|$ tend to infinity. Passing to a subsequence, we may assume that the normalized matrices $\mathbf{G}_n / \|\mathbf{G}_n\|$ converge to a matrix $\mathbf{G}$ of norm 1. Since $\det(\mathbf{G}_n / \|\mathbf{G}_n\|) = \pm \|\mathbf{G}_n\|^{-2} \rightarrow 0$, the matrix $\mathbf{G}$ has rank one, and $\mathbf{G}\mathbf{e} = \langle \mathbf{e}, \mathbf{e}' \rangle \mathbf{e}''$ for all $\mathbf{e} \in \mathbb{R}^2$, for some nonzero vectors $\mathbf{e}', \mathbf{e}'' \in \mathbb{R}^2$. If $x \in \mathbb{RP}^1$ is the line spanned by a vector $\mathbf{e}$ with $\langle \mathbf{e}, \mathbf{e}' \rangle \neq 0$, then $\mathbf{G}_n \mathbf{e} / \|\mathbf{G}_n\| \rightarrow \langle \mathbf{e}, \mathbf{e}' \rangle \mathbf{e}'' \neq 0$, and $g_n(x)$ converges to the line y spanned by $\mathbf{e}''$. Thus, if z is the line orthogonal to $\mathbf{e}'$ and $w = \xi(z)$, then $\xi = \zeta_{y,z,w}$.

Now let $K \subseteq E(\mathbb{RP}^1, \Gamma)$ be compact and hereditarily separable, and let Z_K be the set of points $z \in \mathbb{RP}^1$ for which some map $\zeta_{y,z,w}$ with $y \neq w$ belongs to K ; such a map determines z . Choose, for each $z \in Z_K$, one such map $\zeta_z = \zeta_{y_z, z, w_z} \in K$. We claim that the set $\{\zeta_z : z \in Z_K\}$ is discrete. Indeed, fix a metric d on $\mathbb{RP}^1$ and a point $z \in Z_K$, let $\varepsilon = d(y_z, w_z)/2$, and choose distinct points $x, x' \in \mathbb{RP}^1 \setminus \{z\}$. Consider the neighborhood of ζ_z formed by the maps g with $d(g(z), w_z) < \varepsilon$, $d(g(x), y_z) < \varepsilon$ and $d(g(x'), y_z) < \varepsilon$. Given $z' \in Z_K \setminus \{z\}$, one of the points x and x' differs from z' , and we may suppose that x does. Then $\zeta_{z'}(z) = \zeta_{z'}(x) = y_{z'}$, since neither z nor x equals z' . Were $\zeta_{z'}$ in the neighborhood of ζ_z just described, we would have $d(y_z, w_z) \leq d(y_z, y_{z'}) + d(y_{z'}, w_z) < 2\varepsilon$, which is impossible. Since K is hereditarily separable, the discrete set $\{\zeta_z : z \in Z_K\}$ is countable, and so is Z_K . Evaluation at the points of the countable set $Z_K \cup \mathbb{Q}$ separates the members of K . To see this, note first that two members of Γ that agree on the dense set $\mathbb{Q}$ agree everywhere. A member of Γ , moreover, is injective, whereas a map $\zeta_{y,z,w}$ takes the value y at all points of $\mathbb{Q}$ but at most one. Two maps $\zeta_{y,z,w}$ and $\zeta_{y',z',w'}$ with $y \neq y'$ differ at a point of $\mathbb{Q} \setminus \{z, z'\}$. Lastly, let $\zeta_{y,z,w}$ and $\zeta_{y,z',w'}$ be two distinct maps, labelled so that the first satisfies $w \neq y$. These two maps differ at the point $z \in Z_K$, at which the second takes the value y if $z' \neq z$ and the value $w' \neq w$ if $z' = z$. Hence restriction to $Z_K \cup \mathbb{Q}$ is a continuous injection of the compact space K into the metrizable space $(\mathbb{RP}^1)^{Z_K \cup \mathbb{Q}}$, and K is metrizable.

Finally, the split Cantor set $S(2^{\mathbb{N}})$ is hereditarily separable and not metrizable, and so is the closed subspace of $\hat{D}(S(2^{\mathbb{N}}))$ formed by its points with first coordinate $\mathbf{0}$. In view of the preceding paragraph, neither space embeds into $E(\mathbb{RP}^1, \Gamma)$. The subspace just described is hereditarily separable, since the projection onto the second coordinate maps this subspace homeomorphically onto a subspace of $S(2^{\mathbb{N}})$. To see that this subspace is not metrizable, let $\varpi \in 2^{\mathbb{N}}$ have infinitely many zeros, and let a be the branch of the subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ of Section 1.4 that extends the words $s_{\varpi \upharpoonright n}$. Let $u_1, u_2, \dots$ be the words $s_{(\varpi \upharpoonright n) \cup \{n\}}$ with $\varpi(n) = 0$, listed by increasing n . These words are pairwise incomparable and converge to a , and the sequences $u_k 0^\infty$ decrease, since the canonical isomorphism preserves the lexicographic order. By the description of $\hat{D}(S(2^{\mathbb{N}}))$ in Section 1.4, the point $(\mathbf{0}, \chi_{(a)})$ belongs to the subspace just described, and distinct sequences ϖ give distinct branches. In $S(2^{\mathbb{N}})$, each point $\chi_{(a)}$ has the neighborhoods $\{\chi_{(a)}\} \cup \{\chi_{[x]}, \chi_{(x)} : a < x < b\}$, for $b > a$, and the argument of Lemma 5.6 shows that a metrizable subspace of $S(2^{\mathbb{N}})$ contains only countably many of the points $\chi_{(a)}$. $\square$

Together with Corollary 5.12, Proposition 5.13 shows that the space of deep computations of the Euclidean CCS contains copies of $A(2^{\mathbb{N}})$ and of $D(2^{\mathbb{N}})$, but no copy of $S(2^{\mathbb{N}})$ or of $\hat{D}(S(2^{\mathbb{N}}))$. At the level NIP_1 but not NIP_2 , the set $\Delta_{\mathcal{C}}$ thus realizes the alternative of Todorćević's trichotomy opposite to the one realized by the three-dimensional Bresenham algorithm (see Theorem 5.8). The two alternatives are drawn side by side in Figure 8.

Free reduction. A second example at the level NIP but not NIP_1 is the free group F_2 on two generators a, b , acting on its boundary ∂F_2 . This boundary, a Cantor set, is the space of infinite reduced words in a, a^{-1}, b, b^{-1} . A generator acts on an infinite word by prepending a letter and cancelling it against the first letter of the word if the two letters are inverse, as a stack does. With the predicates $w \mapsto \llbracket w_n = x \rrbracket$, for $n \in \mathbb{N}$ and each letter x , the boundary ∂F_2 and the action of F_2 form a dynamical CCS. Glasner and Megrelishvili proved that this system is tame but not tame₁ [GM22, Corollary 6.6 and Example 6.7]. By Proposition 5.1, this CCS is therefore NIP but not NIP_1 .

An open question. Both examples at this level are actions of groups with two generators. Whether a single map can reach this level, that is, whether some cascade CCS is NIP but not NIP_1 , is Question 7.2. For actions of $\mathbb{Z}$, and more generally of abelian groups, the existence of a tame metrizable system whose enveloping semigroup is not first countable is Question 6.8 of Glasner and Megrelishvili [GM22]. To our knowledge, this question remains open.

5.6. Beyond NIP . Chaotic iteration leads beyond the hierarchy, to sets of computations with the independence property. In a set of computations with the independence property, the computations realize, for some feature and on suitable inputs, every finite pattern of high and low values. We close the section with examples of chaotic iteration, among them Newton's method on the Riemann sphere, the doubling map of the circle and the logistic map beyond the Feigenbaum parameter.

Newton's method on the Riemann sphere. On its Julia set the Newton map is chaotic, and the chaos is visible to the predicates of the CCS. We exhibit the independence in the simplest case, the square roots of 1.

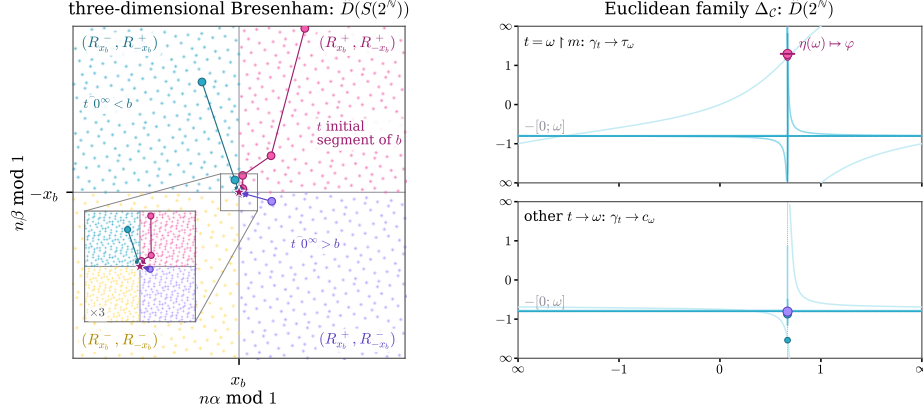


FIGURE 8. Case (ii) of Todorčević's trichotomy, realized in two ways (see Theorem 3.3). Left: the three-dimensional Bresenham algorithm, with α the reciprocal of the plastic number and $\beta = \alpha^2$. The small dots are the points $(n\alpha, n\beta)$ modulo 1 near $(x_b, -x_b)$, for $b = (1, 0, 1, 0, \dots)$ and $x_b = \frac{3}{4}$; the inset enlarges the center three times. Iterates $(\sigma \times \sigma)^n$ whose points approach $(x_b, -x_b)$ within one quadrant converge to the pair written there (see Theorem 5.8). Each entry $R^\pm$ of a pair is a deep computation of the planar algorithm; it rotates by its subscript and breaks ties upward (+) or downward (-) (see Theorem 5.5). The large dots mark the iterates of Remark 5.9, in the three patterns of the family D_7 . Right: the Möbius maps γ_t of Proposition 5.13 on the projective line, for $\omega = (1, 2, 1, 2, \dots)$. Above, the word t runs through initial segments of ω , and the maps converge to τ_ω . This limit sends a point $\eta(\omega)$ defined from ω to the golden ratio φ , and every other point to $-[0; \omega] = 1 - \sqrt{3}$. Below, the word t branches off from ω below or above it (colors as on the left), and the maps converge to the constant map c_ω with value $-[0; \omega]$. The large dots mark the values at $\eta(\omega)$. Thus the two sides of b have distinct limits, while the two sides of ω share one.

Proposition 5.15. *Let $p(z) = z^2 - 1$, let $N(z) = \frac{1}{2}(z + \frac{1}{z})$ be its Newton map on $\hat{\mathbb{C}}$, and let P_2 be the second coordinate of the inverse stereographic projection $\hat{\mathbb{C}} \rightarrow S^2$ of Section 4. The functions $P_2 \circ N^{3k}$, $k \in \mathbb{N}$, form an independent sequence: for all finite disjoint sets $I, J \subseteq \mathbb{N}$ there is $z \in \hat{\mathbb{C}}$ with $P_2(N^{3k}z) \geq \frac{1}{2}$ for $k \in I$ and $P_2(N^{3k}z) \leq -\frac{1}{2}$ for $k \in J$. In particular, the Newton CCS of p is not NIP.*

Proof. The Möbius transformation $\psi(z) = (z - 1)/(z + 1)$ conjugates N to the squaring map: $\psi(N(z)) = \psi(z)^2$. The map ψ sends the imaginary axis, together with ∞ , onto the unit circle, with $\psi^{-1}(e^{i\theta}) = i \cot(\theta/2)$. Hence, for $z = i \cot(\theta/2)$, we have $N^n(z) = i \cot(2^n\theta/2)$ and

$$P_2(N^n(z)) = \frac{2 \cot(2^{n-1}\theta)}{\cot^2(2^{n-1}\theta) + 1} = \sin(2^n\theta).$$

Writing $\theta = 2\pi \sum_{j \geq 1} x_j 2^{-j}$ with bits x_j , we find that the fractional part of $2^{3k}\theta/2\pi$ has binary expansion $0.x_{3k+1}x_{3k+2}x_{3k+3}\dots$. If $(x_{3k+1}, x_{3k+2}, x_{3k+3}) = (0, 1, 0)$, this fractional part lies in $[\frac{1}{4}, \frac{3}{8}]$, and hence $\sin(2^{3k}\theta) \geq \sin \frac{3\pi}{4} > \frac{1}{2}$. If the block is $(1, 1, 0)$, the fractional part lies in $[\frac{3}{4}, \frac{7}{8}]$, whence $\sin(2^{3k}\theta) < -\frac{1}{2}$. Choosing the block $(0, 1, 0)$ for $k \in I$ and $(1, 1, 0)$ for $k \in J$ gives the required z . $\square$

The proof also shows that the doubling map $x \mapsto 2x$ of the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, with the predicates $\cos 2\pi x$ and $\sin 2\pi x$, is not NIP: with $x = \theta/2\pi$, the functions $\sin(2\pi 2^{3k}x)$ form an independent sequence. The doubling map shifts the binary digits of x , and the independence comes from the freedom to choose those digits.

On the real projective line, by contrast, the same algorithm is tame. The Julia set of N , the imaginary axis together with ∞ , meets $\mathbb{RP}^1$ only at ∞ , and on $\mathbb{RP}^1$ the iterates converge to the deep computation of Section 4.3 for $a = 1$. By Proposition 5.2, the corresponding CCS is therefore NIP₃ (see Section 5.2). Whether Newton's method is tame is thus decided by the space of states on which it acts, and not by the formula that defines its step.

Rational maps of the sphere. The chaos exhibited in Proposition 5.15 is not peculiar to the square roots of 1. Chaos of this kind, and with it the independence property, is present on the Julia set of every rational map of degree at least 2. The proof of Proposition 5.16 below rests on two classical facts [Mil06, Sections 4 and 14] about the Julia set $\mathcal{J}_g$ of such a map g (see Section 4). The first is that $\mathcal{J}_g$ is a nonempty compact set without isolated points that satisfies $g^{-1}(\mathcal{J}_g) = \mathcal{J}_g$. The second is that, for every neighborhood U of a point of $\mathcal{J}_g$, the image $g^m(U)$ contains $\mathcal{J}_g$ for all sufficiently large m . This second fact is a consequence of Montel's theorem and of the density of repelling cycles in $\mathcal{J}_g$.

Proposition 5.16. *Let $g: \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ be a rational map of degree at least 2, and consider the CCS with states $\hat{\mathbb{C}}$, the predicates P_1, P_2, P_3 of Section 4, and the semigroup of iterates of g . There are $j \in \{1, 2, 3\}$, a positive integer n and reals $a < b$ such that the functions $P_j \circ g^{nk}$, $k \in \mathbb{N}$, form an independent sequence: for all finite disjoint sets $I, J \subseteq \mathbb{N}$ there is $z \in \hat{\mathbb{C}}$ with $P_j(g^{nk}(z)) \leq a$ for $k \in I$ and $P_j(g^{nk}(z)) \geq b$ for $k \in J$. In particular, the CCS is not NIP, and neither are the Newton CCSs for $q(z) = z^3 - 2z + 2$ and for $u(z) = z^3 - 1$.*

Proof. Choose distinct points $z_0, z_1 \in \mathcal{J}_g$ and an index j with $P_j(z_0) \neq P_j(z_1)$, say $P_j(z_0) < P_j(z_1)$, and reals $a < b$ strictly between these two values. By continuity, the points z_0 and z_1 have closed neighborhoods U_0 and U_1 such that $P_j \leq a$ on U_0 and $P_j \geq b$ on U_1 , and we may choose n with $g^n(U_0) \supseteq \mathcal{J}_g$ and $g^n(U_1) \supseteq \mathcal{J}_g$. Let $I, J \subseteq \mathbb{N}$ be finite and disjoint, and let m exceed all their elements; put $U^{(k)} = U_0$ for $k \in I$ and $U^{(k)} = U_1$ for $k \notin I$. Let $C_m = U^{(m)} \cap \mathcal{J}_g$ and, for $k = m-1, \dots, 0$, let $C_k = U^{(k)} \cap \mathcal{J}_g \cap g^{-n}(C_{k+1})$. Each C_k is nonempty: since $g^n(U^{(k)}) \supseteq \mathcal{J}_g \supseteq C_{k+1}$, some point of $U^{(k)}$ is mapped by g^n into C_{k+1} , and that point lies in $\mathcal{J}_g$ because $g^{-1}(\mathcal{J}_g) = \mathcal{J}_g$. Every $z \in C_0$ satisfies $g^{nk}(z) \in U^{(k)}$ for $k \leq m$, and hence $P_j(g^{nk}(z)) \leq a$ for $k \in I$ and $P_j(g^{nk}(z)) \geq b$ for $k \in J$. For the last assertion, note that the Newton maps $N_q(z) = (2z^3 - 2)/(3z^2 - 2)$ and $N_u(z) = (2z^3 + 1)/(3z^2)$ of Section 4 have degree 3, since in each quotient the numerator and the denominator have no common zero. $\square$

The logistic map at $r = 4$. The doubling map also governs the logistic map $f_4(x) = 4x(1 - x)$ of Section 5.2, as Figure 9 illustrates.

Proposition 5.17. *The functions f_4^{3k} , $k \in \mathbb{N}$, form an independent sequence on $[0, 1]$: for all finite disjoint sets $I, J \subseteq \mathbb{N}$ there is $x \in [0, 1]$ with $f_4^{3k}(x) \geq \frac{3}{4}$ for $k \in I$ and $f_4^{3k}(x) \leq \frac{1}{4}$ for $k \in J$. In particular, the logistic CCS for $r = 4$ is not NIP.*

Proof. Let $h(\theta) = \sin^2(\theta/2)$ for $\theta \in \mathbb{R}$, and note the identity $f_4(h(\theta)) = 4\sin^2(\theta/2)\cos^2(\theta/2) = \sin^2\theta = h(2\theta)$. Iterated, this identity yields $f_4^n(h(\theta)) = h(2^n\theta) = \sin^2(2^{n-1}\theta)$. Writing $\theta = 2\pi \sum_{j \geq 1} x_j 2^{-j}$ with bits x_j , we find that the fractional part of $2^{3k}\theta/2\pi$ has binary expansion $0.x_{3k+1}x_{3k+2}x_{3k+3}\dots$. Moreover, the angle $2^{3k-1}\theta$ differs from π times this fractional part by an integer multiple of π . If $(x_{3k+1}, x_{3k+2}, x_{3k+3}) = (0, 1, 1)$, this fractional part lies in $[\frac{3}{8}, \frac{1}{2}]$, and hence $f_4^{3k}(h(\theta)) = \sin^2(2^{3k-1}\theta) \geq \sin^2 \frac{3\pi}{8} > \frac{3}{4}$. If the block is $(0, 0, 0)$, the fractional part lies in $[0, \frac{1}{8}]$, whence $f_4^{3k}(h(\theta)) \leq \sin^2 \frac{\pi}{8} < \frac{1}{4}$. Choosing the block $(0, 1, 1)$ for $k \in I$ and $(0, 0, 0)$ for $k \in J$, and putting $x = h(\theta)$, gives the required point. $\square$

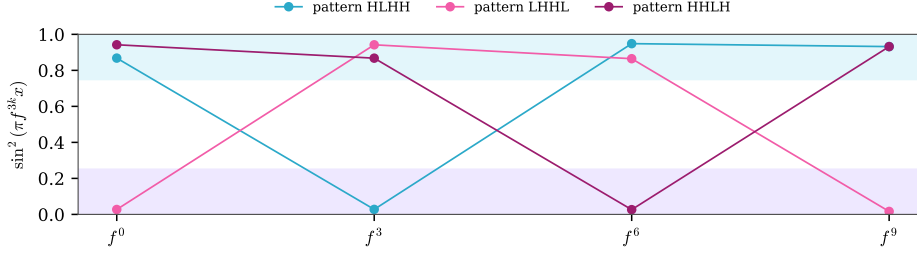


FIGURE 9. Independence for the doubling map $x \mapsto 2x$ of the circle and for the logistic map f_4 . Along an orbit of the doubling map, the values of $\sin^2(\pi x) = (1 - \cos 2\pi x)/2$ form an orbit of f_4 . At the times 0, 3, 6, 9, these values follow any prescribed pattern of high (H, at least $3/4$) and low (L, at most $1/4$) values. For this purpose, the binary digits of the initial state are chosen in blocks 011 for H and 000 for L.

Thus the logistic family, which is stable for $0 < r \leq 1$ and NIP₃ but not stable for $1 < r \leq 3$ (see Proposition 5.4), leaves the hierarchy at $r = 4$. The passage takes place at the Feigenbaum parameter $r_\infty \approx 3.5699$, the limit of the cascade of period doublings, and the family skips the intermediate levels. Indeed, in Proposition 5.18 below we show that the family remains NIP₃ below r_∞ and has the independence property beyond r_∞ , periodic windows included. Moreover, in Remark 5.19 we sketch the argument that the family is still NIP₃ at r_∞ . The arguments rest on the kneading theory of Milnor and Thurston [MT88] and on the theory of interval maps all of whose periodic points have periods that are powers of 2 [BC92, dMvS93].

Proposition 5.18. *Let r_∞ be the supremum of the parameters $r \in (0, 4]$ for which every periodic point of f_r has a power of 2 as its period.*

- (i) *For $3 < r < r_\infty$, the logistic CCS is NIP₃ but not stable.*
- (ii) *For $r_\infty < r \leq 4$, there are a positive integer n and reals $a < b$ such that the functions f_r^{nk} , $k \in \mathbb{N}$, form an independent sequence: for all finite disjoint*

sets $I, J \subseteq \mathbb{N}$ there is $x \in [0, 1]$ with $f_r^{nk}(x) \leq a$ for $k \in I$ and $f_r^{nk}(x) \geq b$ for $k \in J$. In particular, the logistic CCS is not NIP.

Proof. Writing f for f_r , we use three classical facts [MT88, BC92, dMvS93]. First, by the monotonicity of the kneading invariant in the logistic family, every periodic point of f has a power of 2 as its period if and only if $r \leq r_\infty$. Moreover, for $r < r_\infty$ only finitely many periods occur, whence f has only finitely many periodic points. Second, for $2 < r < 4$, every limit point of an orbit that starts in $(0, 1)$ lies in the interval $[f^2(\frac{1}{2}), f(\frac{1}{2})]$, which does not contain 0. Third, a continuous map of a compact interval with only finitely many periodic points has every orbit asymptotic to a periodic orbit [BC92].

(i) Suppose that $3 < r < r_\infty$, and let N be a common multiple of the periods of f , which are powers of 2. If the orbit of x is asymptotic to the periodic orbit $\mathcal{Q}$, then $f^{Nk}(x)$ converges to a point of $\mathcal{Q}$. Indeed, the map f^N fixes $\mathcal{Q}$ pointwise, and by uniform continuity the point of $\mathcal{Q}$ nearest to $f^i(x)$ is eventually carried by f to the point of $\mathcal{Q}$ nearest to $f^{i+1}(x)$. Let $\ell_r(x)$ be the limit of the sequence $(f^{Nk}(x))_{k \in \mathbb{N}}$. Then $f^{Nk+i} \rightarrow f^i \circ \ell_r$ pointwise for $0 \leq i < N$, and every limit of the iterates along a nonprincipal ultrafilter on $\mathbb{N}$ is one of these N maps. Each of the maps $f^i \circ \ell_r$ is of Baire class 1, being the pointwise limit of a sequence of iterates. Hence $E([0, 1], f)$ is countable and therefore metrizable, and by Proposition 5.1 the CCS is NIP₃. The map ℓ_r , which fixes 0 and maps $(0, 1)$ into the interval $[f^2(\frac{1}{2}), f(\frac{1}{2})]$, is thus discontinuous at 0, and the CCS is not stable.

(ii) Suppose that $r_\infty < r \leq 4$; by the first fact, the map f then has a periodic point whose period is not a power of 2. Hence, by a theorem of Block and Coppel [BC92, Chapter II], some iterate f^n is strictly turbulent. Strict turbulence of f^n means that there are disjoint closed intervals H_0 and H_1 , with H_0 to the left of H_1 , such that $f^n(H_0)$ and $f^n(H_1)$ both contain $H_0 \cup H_1$. Let $a = \max H_0$ and $b = \min H_1$, let $I, J \subseteq \mathbb{N}$ be finite and disjoint, and let m exceed all their elements; put $H^{(k)} = H_0$ for $k \in I$ and $H^{(k)} = H_1$ for $k \notin I$. Let $C_m = H^{(m)}$ and, for $k = m-1, \dots, 0$, let $C_k = H^{(k)} \cap f^{-n}(C_{k+1})$. Each set C_k is nonempty, since $f^n(H^{(k)}) \supseteq C_{k+1}$. Every $x \in C_0$ satisfies $f^{nk}(x) \in H^{(k)}$ for $k \leq m$, and hence $f^{nk}(x) \leq a$ for $k \in I$ and $f^{nk}(x) \geq b$ for $k \in J$. $\square$

Remark 5.19. At the Feigenbaum parameter itself the logistic CCS is again NIP₃ but not stable; we sketch the argument, which rests on the structure of infinitely renormalizable maps [BC92, dMvS93]. For $r = r_\infty$ and every $m \in \mathbb{N}$, the map $f = f_r$ has a cycle of 2^m pairwise disjoint closed intervals, permuted cyclically by f and contained in the intervals of the cycle of level $m-1$. Moreover, the lengths of the intervals of level m tend to 0 as $m \rightarrow \infty$. Finally, every orbit that starts in $(0, 1)$ either lands on a periodic orbit or enters the cycle of every level. Label the intervals of the cycle of level m by the residues modulo 2^m , in such a way that f adds 1 to the label. Let $\mathcal{U}$ be a nonprincipal ultrafilter on $\mathbb{N}$, and let ς be the limit of n along $\mathcal{U}$ in the ring $\mathbb{Z}_2$ of 2-adic integers. If the orbit of x lands on a periodic orbit, of period 2^m , then $f^n(x)$ is eventually determined by the residue of n modulo 2^m . Otherwise, for each m , the point $f^n(x)$ eventually lies in the interval of level m whose label differs from n by a certain residue modulo 2^m . This residue depends on x and m but not on n , and the intervals obtained for successive m shrink to a point. In either case, the points $f^n(x)$ converge along $\mathcal{U}$ to a limit $\ell^\varsigma(x)$ that depends only on ς . The map ℓ^ς is the pointwise limit of the iterates f^{n_j} along any sequence of natural numbers n_j that tends to infinity and converges to ς in $\mathbb{Z}_2$.

Each map ℓ^ς is therefore of Baire class 1. Hence $E([0, 1], f)$ consists of the iterates and the maps ℓ^ς , $\varsigma \in \mathbb{Z}_2$. The rational points of $[0, 1]$, together with the point $\frac{1}{2}$, separate the members of $E([0, 1], f)$: distinct iterates are continuous and differ at a rational point. Moreover, an iterate f^n and a map ℓ^ς differ at every rational point $x > 0$ so small that $f^n(x) < f^2(\frac{1}{2})$, since ℓ^ς maps $(0, 1)$ into $[f^2(\frac{1}{2}), f(\frac{1}{2})]$. Finally, two maps ℓ^ς and $\ell^{\varsigma'}$ with $\varsigma \neq \varsigma'$ differ at $\frac{1}{2}$, whose orbit lands on no periodic orbit. Indeed, whenever $\varsigma \not\equiv \varsigma'$ modulo 2^m , the points $\ell^\varsigma(\frac{1}{2})$ and $\ell^{\varsigma'}(\frac{1}{2})$ lie in distinct intervals of level m . As in the proof of Proposition 3.4(i), the space $E([0, 1], f)$ is therefore metrizable, and by Proposition 5.1 the CCS is NIP₃. The CCS is not stable, since each map ℓ^ς fixes 0 and is discontinuous there.

The Thue–Morse sequence. Even a sequence produced by a simple substitution rule can be wild. The Thue–Morse sequence t_n is the parity of the number of ones in the binary expansion of n . The same sequence is the limit of the words obtained from 0 by the substitution $0 \mapsto 01$, $1 \mapsto 10$. The two-sided sequences all of whose finite blocks occur in (t_n) form the Thue–Morse subshift $X_{TM} \subseteq 2^{\mathbb{Z}}$, on which the shift σ is a homeomorphism. Glasner and Megrelishvili observed that the Morse minimal system, which consists of the subshift X_{TM} and the action of $\mathbb{Z}$ generated by σ , is not tame [GM18]. In Appendix A we give a self-contained proof, which relies only on the structure theorem for tame minimal systems. By Proposition 5.1, the CCS with states X_{TM} , the coordinate predicates and the group $\{\sigma^n : n \in \mathbb{Z}\}$ is therefore not NIP. The cascade CCS of σ , whose computations are only the iterates σ^n with $n \in \mathbb{N}$, is not NIP either. Indeed, the reversal of two-sided sequences maps X_{TM} onto itself and conjugates σ to its inverse. The reversal therefore turns an independent sequence of functions $f \circ \sigma^{n_k}$ with negative exponents n_k into one with positive exponents. Both facts are stated, with complete proofs, in Proposition A.1.

5.7. Summary. Table 2 collects the examples of this section. The Euclidean algorithm and free reduction are actions of groups with two generators, and the family $\Delta_{\mathcal{C}}$ consists of elements of the first of these groups. Each of the remaining entries of the table is a single map iterated.

6. RANDOMIZED VERSIONS OF NIP AND MONTE CARLO COMPUTABILITY OF DEEP COMPUTATIONS

Replacing topology by measure, we dispense in this section with the countability of the set $\mathcal{P}$ of predicates, on which the theory of Sections 3–5 depends. In those sections we studied deep computations as *topological* limits of computations. There we showed that, for countable $\mathcal{P}$ and under the NIP, each deep computation is of the first Baire class on the type shards. We also showed that, on each shard, each such deep computation is the limit of a sequence of computations. That theory rests on the type shards being Polish, and hence on the countability of $\mathcal{P}$, which is used in the proof of Theorem 3.1. In place of membership in the first Baire class, we ask that a deep computation be *universally measurable*, that is, measurable with respect to every Radon probability measure on the type shards. For every such measure, each output feature of a universally measurable deep computation is then, by Luzin’s theorem, continuous on a closed subset of the shard whose complement has arbitrarily small measure. Hence, by the Stone–Weierstrass theorem, each such output feature is a uniform limit of polynomials in the features on that subset. We call a deep computation that is universally measurable in this sense *universally Monte*

Level	Computation	Deep computations; source
stable	rotation of the circle	the circle group
	logistic map, $0 < r \leq 1$	iterates and one continuous limit; Prop. 5.4
NIP ₃ , not stable	power method	countable; Prop. 5.3
	Newton's method for $x^2 - a$ on $\mathbb{RP}^1$	countable; Prop. 5.2
	logistic map, $1 < r \leq r_\infty$	countable for $r < r_\infty$, metrizable for $r = r_\infty$; Props. 5.4, 5.18, Rem. 5.19
NIP ₂ , not NIP ₃	Bresenham, plane (Sturmian shift)	isolated iterates and a double arrow circle; Thm. 5.5
NIP ₁ , not NIP ₂	Bresenham, space (two Sturmian shifts)	isolated iterates and a copy of $\mathbb{D} \times \mathbb{D}$; contains $\hat{D}(S(2^{\mathbb{N}}))$, not $D(2^{\mathbb{N}})$; Thm. 5.8
	the family $\Delta_{\mathbb{C}}$ in $\mathrm{PGL}(2, \mathbb{Z})$	a copy of $\hat{D}(2^{\mathbb{N}})$; Prop. 5.13
NIP, not NIP ₁	a single map, with $D(2^{\mathbb{N}})$	open; Question 7.3
	Euclidean algorithm: $\mathrm{PGL}(2, \mathbb{Z})$ on $\mathbb{RP}^1$ (group)	contains $A(2^{\mathbb{N}})$, of exact equality tests, and $D(2^{\mathbb{N}})$; Cor. 5.12, Prop. 5.13
	free reduction: F_2 on ∂F_2 (group)	not first countable; [GM22]
not NIP	a single map	open; Question 7.2
	doubling map of the circle	Prop. 5.15
	Newton's method for $z^2 - 1$ on $\hat{\mathbb{C}}$	Prop. 5.15
	Newton's method for q and u on $\hat{\mathbb{C}}$	Prop. 5.16
	logistic map, $r_\infty < r \leq 4$	Props. 5.17, 5.18
	Thue–Morse shift	Prop. A.1

TABLE 2. Computations at each level of the hierarchy and beyond it, with the cases left open.

Carlo computable (see Definition 6.6). The role of the Bourgain–Fremlin–Talagrand dichotomy is then taken by its measure-theoretic form (see Theorem 6.5). The role of NIP, for its part, is shared with *Talagrand stability*, a measure-theoretic condition introduced by Talagrand in his study of Pettis integration (see Definition 6.9).

The first main result of the section, Theorem 6.7, characterizes the NIP by the universal measurability of the restrictions of ultracomputations to the type shards. The theorem derives Monte Carlo computability from the NIP thus characterized; the second main result, Corollary 6.15, derives the same conclusion from universal Talagrand stability. By Theorem 6.18, these two hypotheses, the NIP and universal Talagrand stability, coincide for countable sets of computations (see Corollary 6.19).

6.1. NIP and Monte Carlo computability of deep computations. Of the two routes to universal Monte Carlo computability followed in this section, the first starts from the NIP and passes through universal measurability, which we introduce here. Universal measurability is the measure-theoretic counterpart of membership in the first Baire class. This first route leads to Theorem 6.7, by which the deep computations of a confined set of computations with the NIP on shards and features are universally Monte Carlo computable. By Fact 1.2, a function $f: X \rightarrow \mathbb{R}$ on a perfectly normal space X is of Baire class 1 if and only if $f^{-1}[U]$ is an F_σ subset of X for every open $U \subseteq \mathbb{R}$. The following definition replaces the F_σ condition by measurability with respect to every Radon measure.

Definition 6.1. Let X be a Hausdorff space, and let (Y, Σ) be a measurable space. A map $f: X \rightarrow Y$ is *universally measurable* (with respect to Σ) if $f^{-1}(E)$ is μ -measurable for every Radon measure μ on X and every $E \in \Sigma$. When $Y = \mathbb{R}$, we will always take $\Sigma = \mathcal{B}(\mathbb{R})$, the Borel σ -algebra of $\mathbb{R}$.

If X is a compact Hausdorff space, then Radon measures on X are finite, and every nonzero Radon measure on X becomes a probability measure after normalization. Moreover, the sets $E \subseteq \mathbb{R}$ whose preimages under a given function are μ -measurable form a σ -algebra, which contains the Borel sets as soon as it contains the open sets. Together, these two observations yield the following simplification of Definition 6.1 for compact spaces.

Remark 6.2. If X is compact, then a function $f: X \rightarrow \mathbb{R}$ is universally measurable if and only if $f^{-1}(U)$ is μ -measurable for every Radon probability measure μ on X and every open set $U \subseteq \mathbb{R}$.

Universal measurability thus asks for measurability with respect to every Radon measure on the domain, that is, with respect to every admissible way of measuring its subsets. An account of universally measurable sets and functions may be found in the *Handbook of measure theory* [Pap02, Chapters 1 and 2].

Notation 6.3. Following Bourgain, Fremlin and Talagrand [BFT78], we denote by $M_r(X)$ the collection of all universally measurable real-valued functions on X . For a Radon measure μ on X , we further denote by $\mathcal{M}^0(X, \mu)$ the collection of all μ -measurable real-valued functions on X .

The maps that arise in the study of deep computations, transitions and deep computations themselves among them, send a state space $L \subseteq \mathbb{R}^{\mathcal{P}}$ into $\mathbb{R}^{\mathcal{P}}$. The product space $\mathbb{R}^{\mathcal{P}}$ carries two natural σ -algebras: the Borel σ -algebra, generated by the open subsets of $\mathbb{R}^{\mathcal{P}}$, and the cylinder σ -algebra. The latter, generated by the subbasic open sets $\pi_P^{-1}(U)$ with $P \in \mathcal{P}$ and $U \subseteq \mathbb{R}$ open, is in general strictly smaller than the Borel σ -algebra, but the two σ -algebras coincide when $\mathcal{P}$ is countable. We use the cylinder σ -algebra to define universally measurable maps into $\mathbb{R}^{\mathcal{P}}$, because of the following characterization.

Proposition 6.4. Let X be a Hausdorff space and $Y = \prod_{i \in I} Y_i$ be any product of measurable spaces (Y_i, Σ_i) for $i \in I$. Let Σ_Y be the cylinder σ -algebra generated by the measurable spaces (Y_i, Σ_i) , and let $\pi_i: Y \rightarrow Y_i$ be the projections. The following are equivalent for $f: X \rightarrow Y$:

- (i) The map $f: X \rightarrow Y$ is universally measurable (with respect to Σ_Y).
- (ii) The map $\pi_i \circ f: X \rightarrow Y_i$ is universally measurable (with respect to Σ_i) for all $i \in I$.

Proof. The implication (i) $\Rightarrow$ (ii) holds because each projection π_i is measurable and $(\pi_i \circ f)^{-1}(E) = f^{-1}(\pi_i^{-1}(E))$ for every $E \in \Sigma_i$. To prove (ii) $\Rightarrow$ (i), fix a Radon measure μ on X , and note that the sets $C \subseteq Y$ for which $f^{-1}(C)$ is μ -measurable form a σ -algebra. This σ -algebra contains every measurable cylinder $C = \prod_{i \in J} C_i$: if J is the finite set of indices $i \in I$ with $C_i \neq Y_i$, then $f^{-1}(C) = \bigcap_{i \in J} (\pi_i \circ f)^{-1}(C_i)$ is μ -measurable by (ii). Since the measurable cylinders generate Σ_Y , the set $f^{-1}(C)$ is μ -measurable for every $C \in \Sigma_Y$. $\square$

By Proposition 6.4, the universal measurability of a map into $\mathbb{R}^{\mathcal{P}}$ can be checked feature by feature, just as membership in the first Baire class can (see Proposition 1.8).

We saw in Section 1 that the NIP corresponds to relative compactness in the space of functions of Baire class 1 (see Theorems 1.4, 1.6 and 1.9). In Section 3 we saw, moreover, that this correspondence extends to deep computations when the set $\mathcal{P}$ of predicates is countable (see Theorem 3.1). The following theorem, which holds on every Hausdorff space, characterizes the NIP in terms of universal measurability. Applied to the type shards, the theorem therefore requires no countability of $\mathcal{P}$.

Theorem 6.5 (Bourgain–Fremlin–Talagrand [BFT78, Theorem 2F]). *Let X be a Hausdorff space, and let $A \subseteq C_p(X)$ be pointwise bounded. The following are equivalent:*

- (i) *The closure $\overline{A}$ is contained in $M_r(X)$.*
- (ii) *For every compact $K \subseteq X$, the restriction $A|_K$ satisfies the NIP.*
- (iii) *For every Radon measure μ on X , the set A is relatively countably compact in $\mathcal{M}^0(X, \mu)$, i.e., every countable subset of A has a limit point in $\mathcal{M}^0(X, \mu)$.*

Theorem 6.5 leads to a measure-theoretic counterpart of computability, the universal Monte Carlo computability of Definition 6.6 below, whose motivation lies in two classical facts.

- (1) Littlewood’s second principle, that every Lebesgue measurable function is “nearly continuous”, is made precise by Luzin’s theorem. Let (X, Σ, μ) be a finite Radon measure space, and let Y be a second-countable topological space (e.g., $Y = \mathbb{R}^{\mathcal{P}}$ with $\mathcal{P}$ countable), equipped with its Borel σ -algebra. Then a function $f: X \rightarrow Y$ is measurable if and only if, for every $E \in \Sigma$ and every $\varepsilon > 0$, there is a closed set $F \subseteq E$ such that the restriction $f|_F$ is continuous and $\mu(E \setminus F) < \varepsilon$.
- (2) Computability of a transform is characterized by its continuous extendibility to the space of types (see Section 2.3); this characterization is at the core of the first paper of this series [ADIW26].

Taken together, the two facts justify Definition 6.6 below. Indeed, suppose that an extension of a transform is universally measurable on a type shard, and fix a Radon probability measure on that shard. By Luzin’s theorem, each output feature of that extension is then continuous on a closed subset of the shard whose measure is as close to 1 as desired. On that subset, by the Stone–Weierstrass theorem, the output feature in question is a uniform limit of polynomials in the features.

Definition 6.6. Let $(L, \mathcal{P}, \Gamma)$ be a CCS, and let $\mathcal{R}$ be a collection of sizers. A transform $f: L \rightarrow \mathcal{L}_{\text{sh}}$ is *universally Monte Carlo computable on $\mathcal{R}$* if it has an extension $\tilde{f}: \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ such that, for every $r_{\bullet} \in \mathcal{R}$, there is a sizer $s_{\bullet}$ for which the restriction $\tilde{f}|_{\mathcal{L}[r_{\bullet}]}: \mathcal{L}[r_{\bullet}] \rightarrow \mathcal{L}[s_{\bullet}]$ is universally measurable. Here universal

measurability means that, for every Radon probability measure μ on $\mathcal{L}[r_\bullet]$ and every $P \in \mathcal{P}$, the map $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow [-s_P, s_P]$ is μ -measurable.

The deep computations of a set of computations confined by an exhaustive collection of sizers are transforms $L \rightarrow \mathcal{L}_{\text{sh}}$ (see Proposition 2.5). Since each type shard $\mathcal{L}[r_\bullet]$ is compact, Remark 6.2 shows that Radon probability measures suffice in Definition 6.6, in place of general Radon measures.

Theorem 6.7. *Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom, let $\mathcal{R}$ be an exhaustive collection of sizers, and let $\Delta \subseteq \Gamma$ be $\mathcal{R}$ -confined. The following are equivalent:*

- (i) *for all $P \in \mathcal{P}$ and all $r_\bullet \in \mathcal{R}$, the set $\pi_P \circ \Delta|_{L[r_\bullet]}$ satisfies the NIP, where $\pi_P : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}$ is the projection onto the coordinate P ;*
- (ii) *for every $r_\bullet \in \mathcal{R}$, every member of $\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is universally measurable, as a map into $\mathbb{R}^{\mathcal{P}}$ with the cylinder σ -algebra.*

If these conditions hold, then every deep computation in $\overline{\Delta}$ is universally Monte Carlo computable on $\mathcal{R}$.

Proof. Fix $r_\bullet \in \mathcal{R}$ and $P \in \mathcal{P}$, and write $K = \overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$, a compact subset of $\mathcal{L}[r_\bullet]^{\mathcal{L}[r_\bullet]}$, and $A_P = \pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$. By the Extendibility Axiom and confinement, the set A_P consists of continuous functions on the compact space $\mathcal{L}[r_\bullet]$ and is pointwise bounded. Moreover, since the map $g \mapsto \pi_P \circ g$ is continuous, the closure of A_P is the compact set $\{\pi_P \circ g : g \in K\}$. Assume first that (i) holds; then the restriction of A_P to $L[r_\bullet]$, namely $\pi_P \circ \Delta|_{L[r_\bullet]}$, has the NIP. Since $L[r_\bullet]$ is dense in $\mathcal{L}[r_\bullet]$, Lemma 1.10 shows that the set A_P itself has the NIP, and so does its restriction to every compact subset of $\mathcal{L}[r_\bullet]$. Theorem 6.5 then gives the inclusion $\overline{A_P} \subseteq M_r(\mathcal{L}[r_\bullet])$, that is, every member of $\overline{A_P}$ is universally measurable. Since P is arbitrary, Proposition 6.4 shows that every member of K is universally measurable, and (ii) follows. Conversely, if (ii) holds, then every member of $\overline{A_P}$ is universally measurable by Proposition 6.4. By the implication (i) $\Rightarrow$ (ii) of Theorem 6.5, applied to the compact set $\mathcal{L}[r_\bullet]$ itself, the set A_P has the NIP. Consequently, so does its restriction $\pi_P \circ \Delta|_{L[r_\bullet]}$, and (i) holds. Finally, let $f \in \overline{\Delta}$, and let $\tilde{f}$ be the extension of Lemma 2.7. For every $r_\bullet \in \mathcal{R}$, the restriction $\tilde{f}|_{\mathcal{L}[r_\bullet]}$ belongs to $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$, and it is therefore universally measurable when (ii) holds. Since $\tilde{f}$ maps $\mathcal{L}[r_\bullet]$ into itself, the deep computation f satisfies Definition 6.6 with $s_\bullet = r_\bullet$ for every $r_\bullet \in \mathcal{R}$. $\square$

Remark 6.8. Condition (ii) of Theorem 6.7 concerns all the members of the sets $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$, that is, the restrictions of ultracomputations to the type shards. By contrast, the equivalence of the theorem fails if (ii) is weakened to the universal Monte Carlo computability of every deep computation, which requires only that some extension be universally measurable. To see this, let $L = \mathbb{Q}/\mathbb{Z} \subseteq \mathbb{T}$ carry the predicates $\cos 2\pi x$ and $\sin 2\pi x$, and let Γ consist of the iterates of the doubling map $x \mapsto 2x$. The space of types is the circle $\mathbb{T}$, and each computation extends to the continuous map $x \mapsto 2^n x$ of $\mathbb{T}$. Thus the Extendibility Axiom holds, and the single sizer with both bounds equal to 1 forms an exhaustive collection that confines Γ . By the paragraph that follows Proposition 5.15, the iterates of the doubling map of $\mathbb{T}$ do not have the NIP; by Lemma 1.10, neither do their restrictions to the dense set L . Every deep computation $f : L \rightarrow \mathbb{T}$, however, extends to the map that agrees

with f on L and takes the value 0 on $\mathbb{T} \setminus L$. Since L is countable, this extension is Borel, whence every deep computation is universally Monte Carlo computable.

6.2. Talagrand stability and Monte Carlo computability of deep computations. The second route to universal Monte Carlo computability starts from Talagrand stability, a notion that Talagrand introduced in his study of Pettis integration [Tal84] and that is closely related to the NIP. We follow this route here and compare the two routes at the end of the subsection. Let X be a compact Hausdorff space, let $A \subseteq \mathbb{R}^X$ be a set of functions, and let μ be a Radon probability measure on X . For a μ -measurable set $E \subseteq X$, a positive integer k and real numbers $a < b$, we write the points of E^{2k} as $x = (x_0, \dots, x_{2k-1})$ and define

$$D_k(A, E, a, b) = \bigcup_{f \in A} \{x \in E^{2k} : f(x_{2i}) \leq a, f(x_{2i+1}) \geq b \text{ for all } i < k\}.$$

Definition 6.9. The set A is *Talagrand μ -stable* if, for every μ -measurable set $E \subseteq X$ of positive measure and all reals $a < b$, there is an integer $k \geq 1$ such that

$$(\mu^{2k})^*(D_k(A, E, a, b)) < (\mu(E))^{2k}.$$

Here $(\mu^{2k})^*$ denotes the outer measure associated with the product measure μ^{2k} on X^{2k} .

The outer measure is needed in Definition 6.9 because $D_k(A, E, a, b)$ need not be measurable. The set $D_k(A, E, a, b)$ is measurable when A is a countable set of μ -measurable functions, since it is then a countable union of measurable sets.

Remarks 6.10. Let X , μ , $A \subseteq \mathbb{R}^X$ and $E \subseteq X$ be as above, and let $k \geq 1$ be an integer and $a < b$ be reals. The following hold.

- (1) If $A \subseteq B \subseteq \mathbb{R}^X$, then

$$D_k(A, E, a, b) \subseteq D_k(B, E, a, b).$$

- (2) If $a < a' < b' < b$, then

$$D_k(\overline{A}, E, a, b) \subseteq D_k(A, E, a', b').$$

[*Proof:* If $x \in D_k(\overline{A}, E, a, b)$, then there is $f \in \overline{A}$ such that $f(x_{2i}) \leq a < a'$ and $f(x_{2i+1}) \geq b > b'$ for all $i < k$. By the definition of the topology of pointwise convergence, there is $g \in A$ such that $g(x_{2i}) < a'$ and $g(x_{2i+1}) > b'$ for all $i < k$; hence $x \in D_k(A, E, a', b')$.]

- (3) Every subset of a Talagrand μ -stable set is Talagrand μ -stable, by (1).
(4) If A is Talagrand μ -stable, then so is $\overline{A}$, by (2).

A set of functions on a compact Hausdorff space is *universally Talagrand stable* if it is Talagrand μ -stable for every Radon probability measure μ on that space (see Definition 6.13 below). In Corollary 6.15 below we consider a confined set Δ of computations such that, for every feature P and every type shard $\mathcal{L}[r_\bullet]$, the set $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is universally Talagrand stable. We show there that the deep computations of every such set are universally Monte Carlo computable. The proof of the corollary rests on Lemma 6.12, by which the pointwise closure of a Talagrand μ -stable set consists of μ -measurable functions. The lemma, in turn, rests on the following characterization of measurable functions (compare Fact 1.2).

Fact 6.11 ([Fre03, Lemma 413G]). *Let (X, Σ, μ) be a measure space, and let $\mathcal{K} \subseteq \Sigma$ be a collection of measurable sets such that the following conditions hold:*

- The measure space (X, Σ, μ) is complete, i.e., for all $E \in \Sigma$ with $\mu(E) = 0$ and $F \subseteq E$ we have $F \in \Sigma$.
- It is semi-finite, i.e., for all $E \in \Sigma$ with $\mu(E) = \infty$ there exists $F \subseteq E$ such that $F \in \Sigma$ and $0 < \mu(F) < \infty$.
- It is saturated, i.e., a set $E \subseteq X$ belongs to Σ if and only if $E \cap F \in \Sigma$ for all $F \in \Sigma$ with $\mu(F) < \infty$.
- It is inner regular with respect to $\mathcal{K}$, i.e., for all $E \in \Sigma$

$$\mu(E) = \sup\{\mu(K) : K \in \mathcal{K} \text{ and } K \subseteq E\}.$$

Then a function $f: X \rightarrow \mathbb{R}$ is measurable if and only if, for every $K \in \mathcal{K}$ with $0 < \mu(K) < \infty$ and all reals $a < b$, either $\mu^*(K_{\leq a}) < \mu(K)$ or $\mu^*(K_{\geq b}) < \mu(K)$, where $K_{\leq a} = \{x \in K : f(x) \leq a\}$ and $K_{\geq b} = \{x \in K : f(x) \geq b\}$. For instance, let X be a compact Hausdorff space, and let μ be a Radon probability measure on X . If Σ is the completion by μ of the Borel σ -algebra of X and $\mathcal{K}$ is the collection of compact subsets of X , then all four conditions hold.

We first show that Talagrand stability passes to the pointwise closure and forces measurability.

Lemma 6.12. *Let X be a compact Hausdorff space, let μ be a Radon probability measure on X , and let $A \subseteq \mathbb{R}^X$ be a set of functions. If A is Talagrand μ -stable, then $\overline{A}$ is also Talagrand μ -stable and $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$.*

Proof. By Remarks 6.10, the closure $\overline{A}$ is Talagrand μ -stable, and hence so is $\{f\}$ for every $f \in \overline{A}$. It therefore suffices to show that if $\{f\}$ is Talagrand μ -stable, then f is μ -measurable. Suppose that f is not μ -measurable. Then Fact 6.11 gives a compact set E of positive measure and reals $a < b$ with $\mu^*(E_{\leq a}) = \mu^*(E_{\geq b}) = \mu(E)$, where $E_{\leq a} = \{x \in E : f(x) \leq a\}$ and $E_{\geq b} = \{x \in E : f(x) \geq b\}$. For every $k \geq 1$ we have $(E_{\leq a} \times E_{\geq b})^k \subseteq D_k(\{f\}, E, a, b)$, and this product has outer measure $(\mu(E))^{2k}$ (see [Tal84]); hence $\{f\}$ is not Talagrand μ -stable. $\square$

Definition 6.13. Let X be a compact Hausdorff space. A set $A \subseteq \mathbb{R}^X$ is *universally Talagrand stable* if it is Talagrand μ -stable for every Radon probability measure μ on X .

For pointwise bounded sets of continuous functions on a compact space, universal Talagrand stability implies the NIP.

Proposition 6.14. *Let X be a compact Hausdorff space, and let $A \subseteq C_p(X)$ be pointwise bounded. If A is universally Talagrand stable, then A satisfies the NIP.*

Proof. By Lemma 6.12, the closure $\overline{A}$ is contained in $\mathcal{M}^0(X, \mu)$ for every Radon probability measure μ on X , that is, $\overline{A} \subseteq M_r(X)$. By the implication (i) $\Rightarrow$ (ii) of Theorem 6.5, applied with $K = X$, the set A satisfies the NIP. $\square$

Corollary 6.15. *Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom, let $\mathcal{R}$ be an exhaustive collection of sizers, and let $\Delta \subseteq \Gamma$ be $\mathcal{R}$ -confined. If $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is universally Talagrand stable for all $P \in \mathcal{P}$ and all $r_\bullet \in \mathcal{R}$, then every deep computation in Δ is universally Monte Carlo computable on $\mathcal{R}$.*

Proof. Let $P \in \mathcal{P}$ and $r_\bullet \in \mathcal{R}$ be given. By the Extendibility Axiom and confinement, the set $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is a pointwise bounded family of continuous functions on the compact space $\mathcal{L}[r_\bullet]$. Proposition 6.14 therefore shows that this set has the

NIP, and hence so does its restriction $\pi_P \circ \Delta|_{L[r_\bullet]}$. Theorem 6.7 now completes the proof. $\square$

In the context of deep computations we have thus found two routes to universal Monte Carlo computability, the NIP and universal Talagrand stability, and the second implies the first (see Proposition 6.14). For countable sets of measurable functions, the following two results show that whether the NIP implies universal Talagrand stability depends on the axioms of set theory. On the one hand, it is consistent with the Zermelo–Fraenkel axioms of set theory with the axiom of choice (ZFC) that, for such sets, the NIP implies universal Talagrand stability.

Theorem 6.16 (Talagrand [Tal84, Theorem 9-3-1(a)]). *Let X be a compact Hausdorff space, and let $A \subseteq M_r(X)$ be countable and pointwise bounded. Assume that $[0, 1]$ is not the union of fewer than $2^{\aleph_0}$ closed sets of Lebesgue measure zero. If A satisfies the NIP, then A is universally Talagrand stable.*

(The hypothesis on $[0, 1]$ follows, for example, from the Continuum Hypothesis.)

On the other hand, for Lebesgue measure, the failure of this implication is also consistent.

Theorem 6.17 (Fremlin and Shelah [FS93]). *It is consistent with ZFC that there exists a countable pointwise bounded set of Lebesgue measurable functions with the NIP that is not Talagrand stable with respect to Lebesgue measure.*

Both results concern countable sets of measurable functions. For countable sets of *continuous* functions on a compact space, by contrast, the NIP implies universal Talagrand stability in ZFC, without any further set-theoretic hypothesis (see Theorem 6.18 below). Consequently, since the Extendibility Axiom makes computations continuous on the type shards, the two routes to universal Monte Carlo computability coincide for countable sets of computations. After a reduction to a compact metrizable space, this implication for countable sets of continuous functions can be derived from two results on the Glivenko–Cantelli property of a family of functions. The Glivenko–Cantelli property is the almost sure convergence of empirical means to expectations, uniformly over the family. The first result is a theorem of van Handel [vH13, Theorem 1.3], which concerns separable, uniformly bounded families of measurable functions on a standard Borel space. According to this theorem, such a family has the Glivenko–Cantelli property for every probability measure if and only if it contains no Boolean σ -independent sequence. The second result is a theorem of Talagrand [Tal87], which identifies, among uniformly bounded families, the Glivenko–Cantelli families for a measure μ with the Talagrand μ -stable ones (see also [Fre03, §465]). We give instead a direct proof, which reduces the question to a compact metrizable space and then applies Fact 1.2 to a cluster point of the reduced family.

Theorem 6.18. *Let X be a compact Hausdorff space, and let $A \subseteq C_p(X)$ be countable and pointwise bounded. If A satisfies the NIP, then A is universally Talagrand stable.*

Proof. Suppose that A satisfies the NIP, and let μ be a Radon probability measure on X . Towards a contradiction, suppose further that A is not Talagrand μ -stable. Then there are a μ -measurable set $E \subseteq X$ of positive measure and reals $a < b$ such that $\mu^{2^k}(D_k(A, E, a, b)) = (\mu(E))^{2^k}$ for every $k \geq 1$. The sets $D_k(A, E, a, b)$

are measurable, since A is countable and its members are continuous. We first transfer the situation to a compact metrizable space. The map $\Psi: X \rightarrow \mathbb{R}^A$ given by $\Psi(x) = (f(x))_{f \in A}$ is continuous, and its image $Y = \Psi(X)$ is compact and metrizable because A is countable. Moreover, each $f \in A$ factors as $f = \hat{f} \circ \Psi$, where $\hat{f}: Y \rightarrow \mathbb{R}$ is the restriction of the coordinate projection. The set $\hat{A} = \{\hat{f} : f \in A\} \subseteq C_p(Y)$ is pointwise bounded, and it satisfies the NIP, because $Y_{\leq a'}^{f_I} \cap Y_{\geq b'}^{f_J} = \Psi(X_{\leq a'}^{f_I} \cap X_{\geq b'}^{f_J})$ for every sequence $(f_n)_{n \in \mathbb{N}}$ in A , all finite sets $I, J \subseteq \mathbb{N}$ and all reals $a' < b'$. Let ν be the Borel probability measure on Y given by $\nu(B) = \mu(E \cap \Psi^{-1}(B)) / \mu(E)$. Since $D_k(A, E, a, b) = E^{2k} \cap (\Psi^{2k})^{-1}(D_k(\hat{A}, Y, a, b))$, where Ψ^{2k} applies Ψ to each coordinate, we obtain $\nu^{2k}(D_k(\hat{A}, Y, a, b)) = 1$ for every $k \geq 1$.

Now consider the product measure $\nu^{\mathbb{N}}$ on $Y^{\mathbb{N}}$, under which the coordinates y_n of a sequence $(y_n)_{n \in \mathbb{N}}$ are independent, each with distribution ν . Let Z denote the support of ν , which is a nonempty closed, and hence Polish, subspace of Y . For each $k \geq 1$, almost every sequence admits some $\hat{f} \in \hat{A}$ with $\hat{f}(y_{2i}) \leq a$ and $\hat{f}(y_{2i+1}) \geq b$ for all $i < k$, since this event has probability $\nu^{2k}(D_k(\hat{A}, Y, a, b)) = 1$. Moreover, since Y has a countable base and every basic open set that meets Z has positive measure, both $\{y_{2i} : i \in \mathbb{N}\}$ and $\{y_{2i+1} : i \in \mathbb{N}\}$ are dense subsets of Z for almost every sequence. Fix a sequence (y_n) with all these properties, and, for each $k \geq 1$, choose $\hat{f}_k \in \hat{A}$ with $\hat{f}_k(y_{2i}) \leq a$ and $\hat{f}_k(y_{2i+1}) \geq b$ for all $i < k$. Since $\hat{A}$ is pointwise bounded, the sequence $(\hat{f}_k)$ has a cluster point g in $\mathbb{R}^Y$, and $g \in B_1(Y)$ by Theorem 1.6. For every i , the inequalities $\hat{f}_k(y_{2i}) \leq a$ and $\hat{f}_k(y_{2i+1}) \geq b$ hold for all $k > i$, and therefore $g(y_{2i}) \leq a$ and $g(y_{2i+1}) \geq b$. By condition (3) of Fact 1.2, applied to the Polish space Y and its closed subset Z , the restriction $g|_Z$ has a point of continuity z . Every neighborhood of z in Z , however, contains points y_{2i} with $g(y_{2i}) \leq a$ and points y_{2i+1} with $g(y_{2i+1}) \geq b$; this contradicts the continuity of $g|_Z$ at z . $\square$

Corollary 6.19. *Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom, let $\mathcal{R}$ be an exhaustive collection of sizers, and let $\Delta \subseteq \Gamma$ be countable and $\mathcal{R}$ -confined. For all $P \in \mathcal{P}$ and $r_\bullet \in \mathcal{R}$, the set $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is universally Talagrand stable if and only if $\pi_P \circ \Delta|_{L[r_\bullet]}$ satisfies the NIP.*

Proof. By the Extendibility Axiom and confinement, the set $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is a countable, pointwise bounded family of continuous functions on the compact space $\mathcal{L}[r_\bullet]$, in which $L[r_\bullet]$ is dense. If this set is universally Talagrand stable, it has the NIP by Proposition 6.14, and so does its restriction to $L[r_\bullet]$. Conversely, suppose that $\pi_P \circ \Delta|_{L[r_\bullet]}$ has the NIP. By Lemma 1.10, so does the set $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$, to which Theorem 6.18 therefore applies. $\square$

For countable sets of computations, Theorem 6.18 thus supersedes Theorem 6.16 and its set-theoretic hypothesis. Accordingly, the example of Fremlin and Shelah, a set of measurable functions, has no counterpart among countable sets of computations.

7. OPEN QUESTIONS

We close with the questions that remain open, in the order in which they arise from the sections of the paper. The first concerns the hierarchy of Section 3.2,

whose levels were defined through the topology of the space of deep computations. Proposition 3.4 describes these levels by the number of tests needed to tell deep computations apart, but the descriptions so obtained still involve infinitely many tests.

Question 7.1. Is there a finitary, combinatorial characterization of NIP_1 , NIP_2 and NIP_3 , in the spirit of the characterization of NIP by finite patterns in Theorem 3.1?

In Section 5 we place a single iterated map at every level of the hierarchy except one. The exception is the level NIP but not NIP_1 , at which both of our examples are actions of groups with two generators (see Section 5.5).

Question 7.2. Is there a cascade CCS that is NIP but not NIP_1 ? Equivalently, is there a continuous map γ of a compact metrizable space L such that (L, γ) is tame and $E(L, \gamma)$ is not first countable?

For actions of $\mathbb{Z}$, and more generally of abelian groups, the analogous question is Question 6.8 of Glasner and Megrelishvili [GM22].

At the level NIP_1 but not NIP_2 , the deep computations of the three-dimensional Bresenham algorithm contain a copy of $\hat{D}(S(2^{\mathbb{N}}))$ and none of $D(2^{\mathbb{N}})$ (see Theorem 5.8). By contrast, the countable family Δ_e of elements of the group generated by the steps of the Euclidean algorithm realizes the other alternative of Todorćević's trichotomy (see Proposition 5.13).

Question 7.3. Is there a cascade CCS that is NIP_1 but not NIP_2 and whose space of deep computations contains a copy of $D(2^{\mathbb{N}})$?

APPENDIX A. THE THUE–MORSE SYSTEM IS NOT TAME

In this appendix we prove that the shift on the Thue–Morse subshift generates an action of $\mathbb{Z}$ that is not tame, a fact observed by Glasner and Megrelishvili [GM18]. From this fact we derive that the two dynamical CCSs defined by this shift, which we use in Section 5.6 as examples beyond the hierarchy, are not NIP . Our proof is self-contained except for the structure theorem for tame minimal systems under actions of abelian groups, due to Huang, to Kerr and Li, and to Glasner. We recall that theorem below and combine it with an elementary computation of the correlations of the Thue–Morse sequence.

We first recall the objects involved. The Thue–Morse sequence $(t_n)_{n \in \mathbb{N}}$ is defined by letting t_n be the parity of the number of ones in the binary expansion of n , whence $t_{2k} = t_k$ and $t_{2k+1} = 1 - t_k$ for every $k \in \mathbb{N}$. The same sequence is the limit of the words W_m , for $m \in \mathbb{N}$, obtained from 0 by m applications of the substitution $0 \mapsto 01$, $1 \mapsto 10$. The two-sided sequences all of whose finite blocks occur in (t_n) form the Thue–Morse subshift $X_{TM} \subseteq 2^{\mathbb{Z}}$, a closed subset of the compact metrizable space $2^{\mathbb{Z}}$. On this subshift the shift σ , defined by $\sigma(x)_n = x_{n+1}$, is a homeomorphism. As in Section 5.1, a system (L, Γ) consists of a compact metrizable space L and a countable semigroup Γ of continuous maps $L \rightarrow L$. Such a system is *tame*, in the sense of that section, if for every continuous function $f: L \rightarrow \mathbb{R}$ the family $\{f \circ \gamma: \gamma \in \Gamma\}$ has the NIP of Definition 1.5. We call a sequence $(f \circ \gamma_k)_k$ of members of that family *independent* if, for some reals $a < b$, it witnesses the finitary independence property of Definition 1.5. Hence the system (L, Γ) is tame exactly when no independent sequence exists. The CCS with states X_{TM} , the coordinate predicates $x \mapsto x_n$ for $n \in \mathbb{Z}$, and the group $\{\sigma^n: n \in \mathbb{Z}\}$ is a dynamical CCS

in the sense of Section 5.1. The cascade CCS of σ , in turn, is the dynamical CCS with the same states and predicates but with the semigroup $\{\sigma^n : n \in \mathbb{N}\}$. By Proposition 5.1, each of these two CCSs is NIP if and only if its dynamical system is tame.

That the dynamics of σ on X_{TM} is wild follows from the structure theory of tame minimal systems under actions of abelian groups. Huang [Hua06] and, independently, Kerr and Li [KL07] proved that every metrizable minimal tame system under an action of an abelian group is uniquely ergodic. Unique ergodicity means that the system carries exactly one invariant Borel probability measure. The same authors proved, in addition, that such a system is an almost one-to-one extension of its maximal equicontinuous factor. Glasner [Gla07] showed, moreover, that the factor map onto that factor is an isomorphism of measure-preserving systems. For a minimal homeomorphism, the maximal equicontinuous factor is a rotation $y \mapsto y + \alpha$ of a compact abelian group Y in which the multiples of α are dense, equipped with its Haar measure. The characters of Y are eigenfunctions of this rotation and span a dense subspace of the space of square-integrable functions on Y . Consider now a minimal system given by a homeomorphism σ that generates a tame action. The Koopman operator of this system sends each function that is square-integrable with respect to the invariant measure to its composite with σ . By the results just recalled, every such function is a limit in mean square of linear combinations of eigenfunctions of the Koopman operator. The Thue–Morse subshift fails this test, because the correlations of the Thue–Morse sequence are, on average, too small.

Proposition A.1. *Let $X_{TM} \subseteq 2^{\mathbb{Z}}$ be the Thue–Morse subshift, and let σ be the shift on X_{TM} . The action of $\mathbb{Z}$ on X_{TM} generated by σ is not tame. Consequently, neither the CCS with states X_{TM} , the coordinate predicates and the group $\{\sigma^n : n \in \mathbb{Z}\}$ nor the cascade CCS of σ is NIP.*

Proof. We first show that X_{TM} is minimal. Every finite block w of (t_n) occurs in some word W_m . If $\bar{W}_m$ denotes the word obtained from 1 by m applications of the substitution, then the words $W_{m+1} = W_m \bar{W}_m$ and $\bar{W}_{m+1} = \bar{W}_m W_m$ both contain W_m . Moreover, since $t_{2k} = t_k$ and $t_{2k+1} = 1 - t_k$, the block of (t_n) of length 2^{m+1} that begins at the index $2^{m+1}j$ is W_{m+1} or $\bar{W}_{m+1}$ according as t_j is 0 or 1. Hence the block w occurs in every block of length 2^{m+2} of (t_n) , and therefore in every point of X_{TM} ; every orbit of σ is thus dense in X_{TM} .

Suppose now that the action of $\mathbb{Z}$ generated by σ is tame. By the structure theorem recalled above, the subshift X_{TM} carries a unique σ -invariant Borel probability measure μ . The eigenfunctions of the unitary operator V on $L^2(\mu)$ that sends each function to its composite with σ span, by the same theorem, a dense subspace of $L^2(\mu)$. Let $\vartheta(x) = (-1)^{x_0}$ for $x \in X_{TM}$, and let $c(n) = \int \vartheta \cdot (\vartheta \circ \sigma^n) d\mu$ for $n \in \mathbb{N}$. By minimality and compactness, the subshift X_{TM} contains a point x with $x_j = t_j$ for every $j \geq 0$. Every weak* limit of the measures $\frac{1}{K} \sum_{k < K} \delta_{\sigma^k x}$ is σ -invariant and hence equal to μ , whence

$$c(n) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k < K} (-1)^{t_k + t_{k+n}}.$$

Splitting the sum according to the parity of k and using $t_{2k} = t_k$ and $t_{2k+1} = 1 - t_k$, we obtain $c(2n) = c(n)$ and $c(2n+1) = -\frac{1}{2}(c(n) + c(n+1))$ for every $n \in \mathbb{N}$, and in particular $c(1) = -\frac{1}{3}$. For $K \geq 1$, let $A(K) = \sum_{n < K} c(n)^2$ and

$B(K) = \sum_{n < K} c(n)c(n+1)$. For every $K \geq 1$, the recursions yield the identities and inequalities

$$B(2K) = -\frac{1}{2} \sum_{n < K} (c(n) + c(n+1))^2 \leq 0,$$

$$A(2K) = A(K) + \frac{1}{4} \sum_{n < K} (c(n) + c(n+1))^2 \leq \frac{3}{2}A(K) + \frac{1}{2}B(K),$$

where the last inequality follows by expanding the squares, because $c(0) = 1 \geq c(K)^2$. Hence we have $A(4K) \leq \frac{3}{2}A(2K)$ for every $K \geq 1$, and, since $A(2) = 1 + c(1)^2 \leq 2$, it follows that $A(2^k) \leq 2(\frac{3}{2})^{k-1}$ for every $k \geq 1$. Since the function A is nondecreasing, we conclude that $A(K)/K \rightarrow 0$ as $K \rightarrow \infty$ (see Figure 10). Now let λ be a complex number of modulus 1, and let Π_λ be the orthogonal projection of $L^2(\mu)$ onto the kernel of $V - \lambda I$. This kernel consists of the eigenfunctions of V with eigenvalue λ and the function 0. By von Neumann's mean ergodic theorem, applied to the unitary operator $\bar{\lambda}V$, the averages $\frac{1}{K} \sum_{n < K} \bar{\lambda}^n \vartheta \circ \sigma^n$ converge to $\Pi_\lambda \vartheta$ in $L^2(\mu)$. Consequently, since $c(n)$ is the inner product of $\vartheta \circ \sigma^n$ and ϑ , the Cauchy-Schwarz inequality yields

$$\|\Pi_\lambda \vartheta\|^2 = \lim_{K \rightarrow \infty} \left| \frac{1}{K} \sum_{n < K} \bar{\lambda}^n c(n) \right| \leq \lim_{K \rightarrow \infty} \left(\frac{A(K)}{K} \right)^{1/2} = 0.$$

Thus the function ϑ is orthogonal to every eigenfunction of V , and therefore to every function in $L^2(\mu)$, since these eigenfunctions span a dense subspace of $L^2(\mu)$. This conclusion is absurd, since $|\vartheta| = 1$. The action is therefore not tame, and, by Proposition 5.1, the CCS with the group $\{\sigma^n : n \in \mathbb{Z}\}$ is not NIP.

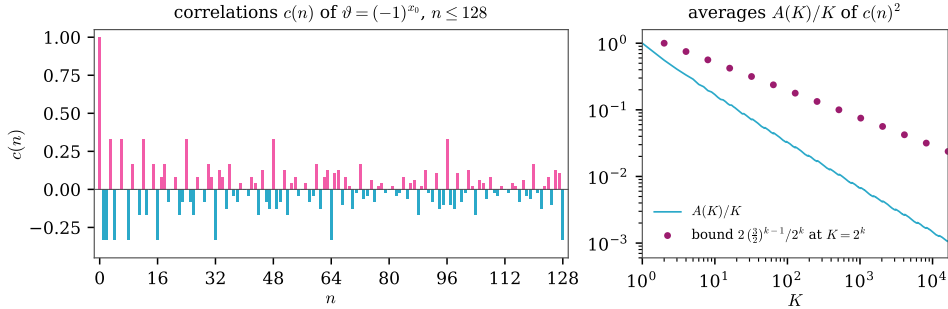


FIGURE 10. The correlations $c(n)$ of the function $\vartheta(x) = (-1)^{x_0}$ on the Thue–Morse subshift. They are computed exactly from the recursions $c(0) = 1$, $c(2n) = c(n)$ and $c(2n+1) = -\frac{1}{2}(c(n) + c(n+1))$, which give $c(1) = -\frac{1}{3}$. The left panel shows the values $c(n)$ for $0 \leq n \leq 128$. The right panel shows, on logarithmic scales, the averages $A(K)/K$ for $K \leq 2^{14}$, where $A(K) = \sum_{n < K} c(n)^2$. The dots mark the bound $2(\frac{3}{2})^{k-1}/2^k$ on $A(2^k)/2^k$. The decay of these averages is the key step in the proof of Proposition A.1.

For the cascade CCS of σ , we use a symmetry of the Thue–Morse sequence. The words W_{2^m} are palindromes, and every finite block of (t_n) occurs in one of them. Consequently, the reversal ι of two-sided sequences, defined by $\iota(x)_n = x_{-n}$, maps

X_{TM} onto itself and satisfies $\iota \circ \sigma = \sigma^{-1} \circ \iota$. Since the action of $\mathbb{Z}$ is not tame, there are a continuous function $f: X_{TM} \rightarrow \mathbb{R}$ and pairwise distinct integers n_k such that the sequence $(f \circ \sigma^{n_k})_k$ is independent. Moreover, infinitely many of the indices n_k have the same sign. If these indices are positive, the corresponding terms already form an independent sequence in the cascade. If these indices are negative, the functions $f \circ \sigma^{n_k} \circ \iota = (f \circ \iota) \circ \sigma^{-n_k}$ form an independent sequence with positive exponents, because the reversal ι is a bijection of X_{TM} . In either case the cascade of σ is not tame, and, by Proposition 5.1, its CCS is not NIP. $\square$

REFERENCES

- [ADIW26] Samson Alva, Eduardo Dueñez, José Iovino, and Claire Walton, *Approximability of deep computations*, Mathematical Structures in Computer Science (2026), To appear; arXiv:2409.06064.
- [ADK08] Spiros A. Argyros, Pandelis Dodos, and Vassilis Kanellopoulos, *A classification of separable Rosenthal compacta and its applications*, Dissertationes Mathematicae **449** (2008), 1–55.
- [Ark92] A. V. Arkhangel'skii, *Topological function spaces*, Mathematics and its Applications (Soviet Series), vol. 78, Kluwer Academic Publishers Group, Dordrecht, 1992, Translated from the Russian by R. A. M. Hoksbergen. MR 1144519
- [Bai99] René Baire, *Sur les fonctions de variables réelles*, Annali di Matematica Pura ed Applicata (3) **3** (1899), 1–123.
- [BC92] Louis S. Block and William A. Coppel, *Dynamics in one dimension*, Lecture Notes in Mathematics, vol. 1513, Springer, Berlin, 1992.
- [BEHW89] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth, *Learnability and the Vapnik–Chervonenkis dimension*, J. Assoc. Comput. Mach. **36** (1989), no. 4, 929–965.
- [BFB16] Richard L. Burden, J. Douglas Faires, and Annette M. Burden, *Numerical analysis*, 10th ed., Cengage Learning, Boston, MA, 2016.
- [BFT78] J. Bourgain, D. H. Fremlin, and M. Talagrand, *Pointwise compact sets of Baire-measurable functions*, American Journal of Mathematics **100** (1978), no. 4, 845–886.
- [BKK] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun, *Deep equilibrium models*, Preprint. <https://arxiv.org/abs/1909.01377>, <https://implicit-layers-tutorial.org/>.
- [Bre65] Jack E. Bresenham, *Algorithm for computer control of a digital plotter*, IBM Systems Journal **4** (1965), no. 1, 25–30.
- [CRBD] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud, *Neural ordinary differential equations*, Preprint. <https://arxiv.org/abs/1806.07366>.
- [Deb13] Gabriel Debs, *Descriptive aspects of Rosenthal compacta*, Recent Progress in General Topology III, Springer, 2013, pp. 205–227.
- [dMvS93] Wellington de Melo and Sebastian van Strien, *One-dimensional dynamics*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 25, Springer, Berlin, 1993.
- [DnI17] Eduardo Dueñez and José N. Iovino, *Model theory and metric convergence I: Metastability and dominated convergence*, Beyond First Order Model Theory, Volume I, CRC Press, Boca Raton, FL, 2017, pp. 131–187. MR 3729326
- [Ell93] Robert Ellis, *The enveloping semigroup of projective flows*, Ergodic Theory and Dynamical Systems **13** (1993), no. 4, 635–660.
- [Eng89] Ryszard Engelking, *General topology*, second ed., Sigma Series in Pure Mathematics, vol. 6, Heldermann Verlag, Berlin, 1989, Translated from the Polish by the author. MR MR1039321 (91c:54001)
- [Fre03] David H. Fremlin, *Measure theory, volume 4: Topological measure spaces*, Torres Fremlin, Colchester, UK, 2003.
- [FS93] David H. Fremlin and Saharon Shelah, *Pointwise compact and stable sets of measurable functions*, Journal of Symbolic Logic **58** (1993), no. 2, 435–455.
- [Gal00] A. Galántai, *The theory of Newton's method*, Journal of Computational and Applied Mathematics **124** (2000), no. 1, 25–44, Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

- [Gla06] Eli Glasner, *On tame dynamical systems*, Colloquium Mathematicum **105** (2006), no. 2, 283–295.
- [Gla07] ———, *The structure of tame minimal dynamical systems*, Ergodic Theory and Dynamical Systems **27** (2007), no. 6, 1819–1837.
- [GM18] Eli Glasner and Michael Megrelishvili, *More on tame dynamical systems*, Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics, Lecture Notes in Mathematics, vol. 2213, Springer, Cham, 2018, pp. 351–392. MR 3821724
- [GM22] ———, *Todorčević’ trichotomy and a hierarchy in the class of tame dynamical systems*, Transactions of the American Mathematical Society **375** (2022), no. 7, 4513–4548.
- [God80] Gilles Godefroy, *Compacts de Rosenthal*, Pacific J. Math. **91** (1980), no. 2, 293–306. MR 615679
- [Gro52] A. Grothendieck, *Critères de compacité dans les espaces fonctionnels généraux*, Amer. J. Math. **74** (1952), 168–186. MR 47313
- [HT19] Haim Horowitz and Stevo Todorčević, *Compact sets of baire class one functions and maximal almost disjoint families*, 2019.
- [Hua06] Wen Huang, *Tame systems and scrambled pairs under an abelian group action*, Ergodic Theory and Dynamical Systems **26** (2006), no. 5, 1549–1567.
- [Kec95] Alexander S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, vol. 156, Springer-Verlag, 1995.
- [Kei23] H. Jerome Keisler, *Model theory for real-valued structures*, Beyond First Order Model Theory, Volume II (José Iovino, ed.), CRC Press, Boca Raton, FL, 2023.
- [KL07] David Kerr and Hanfeng Li, *Independence in topological and C^* -dynamics*, Mathematische Annalen **338** (2007), no. 4, 869–926.
- [KM81] Jean-Louis Krivine and Bernard Maurey, *Espaces de Banach stables*, Israel J. Math. **39** (1981), no. 4, 273–295. MR 83a:46030
- [Köh95] Angela Köhler, *Enveloping semigroups for flows*, Proceedings of the Royal Irish Academy. Section A **95A** (1995), no. 2, 179–191.
- [Kri72] J.-L. Krivine, *Théorie des modèles et espaces L^p* , C. R. Acad. Sci. Paris Sér. A-B **275** (1972), A1207–A1210. MR 46 #7879
- [Kri74] ———, *Langages à valeurs réelles et applications*, Fund. Math. **81** (1974), 213–253, Collection of articles dedicated to Andrzej Mostowski on the occasion of his sixtieth birthday, III. MR 50 #1873
- [Leb04] Henri Lebesgue, *Une propriété caractéristique des fonctions de classe 1*, Bulletin de la Société Mathématique de France **32** (1904), 229–242.
- [Lot02] M. Lothaire, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.
- [MH40] Marston Morse and Gustav A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, American Journal of Mathematics **62** (1940), 1–42.
- [Mil06] John Milnor, *Dynamics in one complex variable*, third ed., Annals of Mathematics Studies, vol. 160, Princeton University Press, Princeton, NJ, 2006.
- [MT88] John Milnor and William Thurston, *On iterated maps of the interval*, Dynamical Systems (College Park, MD, 1986–87), Lecture Notes in Mathematics, vol. 1342, Springer, Berlin, 1988, pp. 465–563.
- [Pap02] E. Pap (ed.), *Handbook of measure theory. Vol. I, II*, North-Holland, Amsterdam, 2002. MR 1953489
- [Ros74] Haskell P. Rosenthal, *A characterization of Banach spaces containing l^1* , Proc. Nat. Acad. Sci. U.S.A. **71** (1974), 2411–2413. MR 358307
- [Ros77] ———, *Point-wise compact subsets of the first Baire class*, Amer. J. Math. **99** (1977), no. 2, 362–378.
- [RPK19] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys. **378** (2019), 686–707. MR 3881695
- [Sau72] N. Sauer, *On the density of families of sets*, J. Combinatorial Theory Ser. A **13** (1972), 145–147. MR 307902
- [She71] Saharon Shelah, *Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory*, Ann. Math. Logic **3** (1971), no. 3, 271–362. MR 317926

- [She72] ———, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific J. Math. **41** (1972), 247–261. MR 307903
- [She90] ———, *Classification theory and the number of nonisomorphic models*, second ed., Studies in Logic and the Foundations of Mathematics, vol. 92, North-Holland Publishing Co., Amsterdam, 1990.
- [Sim15] Pierre Simon, *A guide to nip theories*, Lecture Notes in Logic, vol. 44, Cambridge University Press, 2015.
- [SSBD14] Shai Shalev-Shwartz and Shai Ben-David, *Understanding machine learning: From theory to algorithms*, Cambridge University Press, Cambridge, 2014.
- [Tal84] Michel Talagrand, *Pettis integral and measure theory*, Memoirs of the American Mathematical Society, vol. 51, American Mathematical Society, Providence, RI, USA, 1984, Includes bibliography (pp. 220–224) and index.
- [Tal87] ———, *The Glivenko-Cantelli problem*, The Annals of Probability **15** (1987), no. 3, 837–870.
- [Tka11] Vladimir V. Tkachuk, *A C_p -theory problem book: Topological and function spaces*, Problem Books in Mathematics, Springer, 2011.
- [Tod97] Stevo Todorčević, *Topics in topology*, Lecture Notes in Mathematics, vol. 1652, Springer Berlin, Heidelberg, 1997.
- [Tod99] Stevo Todorčević, *Compact subsets of the first Baire class*, Journal of the American Mathematical Society **12** (1999), no. 4, 1179–1212.
- [Tod23] ———, *Dense metrizability*, Annals of Pure and Applied Logic **175** (2023), 103327.
- [Val84] Leslie G. Valiant, *A theory of the learnable*, Communications of the ACM **27** (1984), no. 11, 1134–1142.
- [VC68] Vladimir N. Vapnik and Alexey Ya. Chervonenkis, *The uniform convergence of frequencies of the appearance of events to their probabilities*, Dokl. Akad. Nauk SSSR **181** (1968), no. 4, 781–783, In Russian; English translation in Soviet Math. Dokl. **9** (1968), 915–918.
- [VC71] Vladimir N Vapnik and Alexey Ya Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory of Probability & Its Applications **16** (1971), no. 2, 264–280.
- [VC74] ———, *Theory of pattern recognition*, Nauka, Moscow, 1974, German Translation: Theorie der Zeichenerkennung, Akademie-Verlag, Berlin, 1979.
- [vH13] Ramon van Handel, *The universal Glivenko–Cantelli property*, Probability Theory and Related Fields **155** (2013), no. 3–4, 911–934, arXiv:1009.4434.