COMPLEXITY OF DEEP COMPUTATIONS VIA TOPOLOGY OF FUNCTION SPACES

EDUARDO DUEÑEZ[1] JOSÉ IOVINO[1] TONATIUH MATOS-WIEDERHOLD[2]
LUCIANO SALVETTI[2] FRANKLIN D. TALL[2]

[1]Department of Mathematics, University of Texas at San Antonio
[2]Department of Mathematics, University of Toronto

Abstract. We use topological methods to study the complexity of deep computations and limit computations. We use topology of function spaces, specifically, the classification of Rosenthal compacta, to identify new complexity classes. We use the language of model theory, specifically, the concept of *independence* from Shelah's classification theory, to translate between topology and computation. We use the theory of Rosenthal compacta to characterize approximability of deep computations, both deterministically and probabilistically.

Introduction

In this paper we study asymptotic behavior of computations, e.g., the depth of a neural network tending to infinity, or the time interval between layers of a time-series network tending toward zero. Recently, particular cases of this concept have attracted considerable attention in deep learning research (e.g., Neural Ordinary Differential Equations [CRBD], Physics-Informed Neural Networks [RPK19], and deep equilibrium models [BKK]). The formal framework introduced here provides a unified setting to study these limit phenomena from a foundational viewpoint.

Informed by model theory, to each computation in a given computation model, we associate a continuous real-valued function, called the *type* of the computation, that describes the logical properties of this computation with respect to the rest of the model. This allows us to view computations in any given computational model as elements of a space of real-valued functions, which is called the *space of types* of the model. The idea of embedding models of theories into their type spaces is central in model theory. In the context of this paper, the embedding of computations into spaces of types allows us to utilize the vast theory of topology of function spaces, known as C_p-*theory*, to obtain results about complexity of topological limits of computations. As we shall indicate next, recent classification results for spaces of functions provide an elegant and powerful machinery to classify computations according to their levels of "tameness" or "wildness", with the former corresponding roughly to polynomial approximability and the latter to exponential approximability. The viewpoint of spaces of types, which we have borrowed from

---

Date: January 20, 2026.

2000 *Mathematics Subject Classification.* 54H30, 68T27, 68T07, 03C98, 03D15, 05D10.

Key words and phrases. Deep computations, deep equilibrium models, deep learning, physics-informed networks, computational complexity, independence property, NIP, infinite Ramsey theory, Baire class 1 functions, Rosenthal compacta, Todorčević trichotomy, Bourgain-Fremlin-Talagrand.

model theory, thus becomes a "Rosetta stone" that allows us to interconnect various classification programs: In topology, the classification of Rosenthal compacta pioneered by Todorčević [Tod99]; in logic, the classification of theories developed by Shelah [She90]; and in statistical learning, the notions of PAC learning and VC dimension pioneered by Vapnik and Chervonenkis [VC74, VC71].

In a previous paper [ADIW24], we introduced the concept of limits of computations, which we called *ultracomputations* (given they arise as ultrafilter limits of standard computations) and *deep computations* (following usage in machine learning [BKK]). There is a technical difference between both designations, but in this paper, to simplify the nomenclature, we will ignore the difference and use only the term "deep computation".

In [ADIW24], we proved a new "tame vs wild" (i.e., polynomially approximable vs nonpolynimially approximable) dichotomy for the complexity of deep computations by invoking a classical result of Grothendieck from the 1950s [Gro52]. Under our model-theoretic Rosetta stone, the property of polynomial approximability of computations is identified with continuous extendibility in the sense of topology, and with the notions of *stability* and *type definability* in model theory.

Simplest among deep computations are those arising as pointwise limits of (continuous) computations proper. In topology, the *first Baire class*, or *Baire class 1* consists of functions (also called simply "*Baire-1*") arising as pointwise limits of sequences of continuous functions. Intuitively, the Baire-1 class consists of functions with "controlled" discontinuities, and lies just one level of topological complexity away from the Baire class 0 which (by definition) consists of continuous functions.

We prove a new "tame vs wild" Ramsey-theoretic dichotomy for complexity of general deep computations by invoking a famous paper by Bourgain, Fremlin and Talagrand from the late 70s [BFT78], and a new trichotomy for the class of "tame" deep computations by invoking an equally celebrated result of Todorčević, from the late 90s, for functions of the first Baire class [Tod99].

Todorčević's trichotomy regards *Rosenthal compacta*; these are special classes of topological spaces, defined as compact spaces that can be embedded (homeomorphically identified as a subset) within the space of Baire class 1 functions on some Polish (separable, completely metrizable) space, under the topology of pointwise convergence. Rosenthal compacta exhibit "topological tameness," meaning that they behave in relatively controlled ways; since the late 70's, they have played a crucial role in understanding the complexity of structures of functional analysis, especially Banach spaces. Todorčević's trichotomy has been utilized to settle longstanding problems in topological dynamics and topological entropy [GM22]. It is noteworthy that Todorčević's proof relies on sophisticated set-theoretic forcing and infinite Ramsey theory. At the time of writing this paper, decades after his original argument, no elementary proof has been found [Tod23, HT19].

Through our Rosetta stone, Rosenthal compacta in topology correspond to the important concept of "Non Independence Property" (known as "NIP") in model theory, identified by Shelah [She71, She90], and to the concept of Probably Approximately Correct learning (known as "PAC learnability") in statistical learning theory identified by Valiant [Val84].

Going beyond Todorčević's trichotomy, we invoke a more recent heptachotomy for Rosenthal compacta obtained by Argyros, Dodos and Kanellopoulos [ADK08]; they identify seven fundamental "prototypes" of separable Rosenthal compacta, and show that any non-metrizable separable Rosenthal compactum must contain a "canonical" embedding of one of these prototypes.

We believe that the results presented in this paper show practitioners of computation, or topology, or descriptive set theory, or model theory, how classification invariants used in their field translate into classification invariants of other fields. In the interest of accessibility, we do not assume the reader to have previous familiarity with advanced topology, model theory, or computing. The only technical prerequisites to read this paper are undergraduate-level topology and measure theory. The necessary topological background beyond undergraduate topology is covered in section 1.

In section 1, we present basic topological and combinatorial preliminaries, and in section 2, we introduce the structural/model-theoretic viewpoint (no previous exposure to model theory is needed). Section 3 is devoted to the classification of deep computations.

Throughout the paper, our results pertain to classical models of computation (particularly computations involving real-valued quantities that are known and manipulated to a finite degree of precision). The final section, Section 4, introduces a probabilistic viewpoint, the development of which we intend to pursue in future research, extending the present framework to encompass non-deterministic and quantum computations.

Contents

1. General topological preliminaries: From continuity to Baire class 1

In this section we present some preliminaries from general topology and function space theory. In the interest of completeness, we some proofs that may be safely skipped by readers familiar with these topics.

Recall that 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 closed sets. A space is metrizable if its topology agrees with the topology induced by some metric therein. Two such metrics inducing the same topology may induce quite different properties in the category of metric spaces. For example, the interval $(0,1)$ with the usual metric (as a subset) of the reals is not complete; however, $(0,1)$ is homeomorphic to the real line, which is complete with respect to the usual metric thereon. In a metrizable space, every open set is F_σ; equivalently, every closed set is G_δ.

A *Polish space* is a separable and completely metrizable topological space, i.e., admitting some complete metric inducing its topology. Although other (possibly incomplete) metrics may induce the same topology, being Polish is a purely topological property. One of the most important Polish spaces is the real line $\mathbb{R}$; the others include the Cantor space $2^{\mathbb{N}}$ and the Baire space $\mathbb{N}^{\mathbb{N}}$. The class of Polish spaces is closed under countable topological products; in particular, the Cantor space $2^{\mathbb{N}}$ (the set of all infinite binary sequences, endowed with the product topology), the Baire space $\mathbb{N}^{\mathbb{N}}$ (the set of all infinite sequences of naturals, also with the product topology), and the space $\mathbb{R}^{\mathbb{N}}$ of sequences of real numbers are all Polish. Recall that the product topology on these spaces is the *topology of pointwise convergence*: a sequence converges in the space if and only if it converges at each coordinate index.

Fact 1.1. *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.*

For a proof, see [Eng89, 4.3.24].

Given two topological spaces X and Y we denote by $C_p(X,Y)$ the set of all continuous functions $f : X \to Y$ endowed with the topology of pointwise convergence. The space $C_p(X,\mathbb{R})$ of continuous real functions on X is denoted simply $C_p(X)$. A natural question is, how do topological properties of X translate into $C_p(X)$ and vice versa? This general question, and the study of these spaces in general, is the concern of C_p-theory, an active field of research in general topology which was pioneered by A. V. Arhangel'skiĭ and his students in the 1970's and 1980's [Ark92]. This field has found many applications in model theory and functional analysis. For a recent survey, see [Tka11].

A *Baire class 1* function between topological spaces is a function that can be expressed as the pointwise limit of a sequence of continuous functions. If X and Y are topological spaces, the space of Baire class 1 functions $f : X \to Y$ endowed with the topology of pointwise convergence is denoted $B_1(X,Y)$ (as above, $B_1(X,\mathbb{R})$ is denoted $B_1(X)$). Clearly, $C_p(X,Y) \subseteq B_1(X,Y)$. The Baire hierarchy of functions was introduced by René-Louis Baire in his 1899 doctoral thesis, *Sur les fonctions de* *variables réelles*. His work moved away from the 19th-century preoccupation with "pathological" functions toward a constructive classification based on pointwise limits.

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

A topological space X is *Baire* if every countable intersection of dense open sets is dense. The Baire Category Theorem states that every Hausdorff compact or completely metrizable space (hence, every Polish space) is Baire.

The following fact was established by Baire in his 1899 thesis. A proof can be found in [Tod97, Section 10].

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

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

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

(3) *For every 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

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

A subset L of a topological space X is *relatively compact* in X if the closure of L in X is compact. Relatively compact subsets of $B_1(X)$ (for X Polish) are of interest analysis and topological dynamics. We begin with the following well-known result. Recall that a set $A \subseteq \mathbb{R}^X$ of real-valued functions is *pointwise bounded* if for every $x \in X$ there is $M_x \geq 0$ (a *pointwise bound at x*) such that $|f(x)| \leq M_x$ for all $f \in A$. We include a proof for the reader's convenience:

Lemma 1.3. *Let X be a Polish space and $A \subseteq B_1(X)$ be pointwise bounded. The following are equivalent:*

(i) *A is relatively compact in $B_1(X)$.*
(ii) *A is relatively countably compact in $B_1(X)$, i.e., every countable subset of A has an accumulation point in $B_1(X)$.*
(iii) *$\overline{A} \subseteq B_1(X)$, where $\overline{A}$ denotes the closure in $\mathbb{R}^X$.*

Proof. (i)$\Rightarrow$(ii) Relatively compact subsets of any space are countably compact therein.

(ii)$\Rightarrow$(iii) Consider any $f \in \overline{A}$ and any countable subset $\{x_i\}_{i\in\mathbb{N}} \subseteq X$. We claim that there is a sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ such that $\lim_{n\to\infty} f_n(x_i) = f(x_i)$ for all $i \in \mathbb{N}$. Since A carries the relative product topology, for each $n \in \mathbb{N}$ there exists $f_n \in A$ such that $|f_n(x_i) - f(x_i)| < \frac{1}{n}$ for all $i \leq n$; the sequence $\{f_n\}$ is as claimed. Seeking a contradiction, assume that A is relatively countably compact in $B_1(X)$, but there exists some $f \in \overline{A} \setminus B_1(X)$. By Fact 1.2, there are countable $D_0, D_1 \subseteq X$ with $\overline{D_0} = \overline{D_1}$, and $a < b$ such that $D_0 \subseteq f^{-1}(-\infty, a]$ and $D_1 \subseteq f^{-1}[b, \infty)$. Per the claim above, let $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ satisfy $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in D_0 \cup D_1$ (the latter being a countable set). By relative countable compactness of A, there is an accumulation point $g \in B_1(X)$ of $\{f_n\}_{n\in\mathbb{N}}$; clearly, f and g agree on $D_0 \cup D_1$.

Thus g takes values $g(x_i) = f(x_i) \leq a$ as well as values $g(x_j) = f(x_j) \geq b > a$ on any open subset of the closed set $\overline{D_0} = \overline{D_1}$, contradicting the implication (1)⇒(3) in Fact 1.2.

(iii)⇒(i) For each $x \in X$, let $M_x \geq 0$ be a pointwise bound for A. Since $\overline{A}$ is a closed subset of the compact space $\prod_{x\in X}[-M_x, M_x] \subseteq \mathbb{R}^X$, it follows that $\overline{A}$ is compact. By (iii), it is also the closure of A in $B_1(X)$. Thus, A is relatively compact in $B_1(X)$. □

1.1. **From Rosenthal's dichotomy to the Bourgain-Fremlin-Talagrand dichotomy to Shelah's NIP.** In metrizable spaces, points of the closure of some subset can always be approximated by points in the set proper, via a convergent sequence. For more complicated spaces, such as C_p-spaces, this fails in remarkable ways. The n-th coordinate map $p_n : 2^{\mathbb{N}} \to \{0,1\}$ on the Cantor space $X = 2^{\mathbb{N}}$ ($= \{0,1\}^{\mathbb{N}}$) is continuous for each $n \in \mathbb{N}$, and one can show (e.g., [Tod97, Chapter 1.1]) that $\{p_n\}_{n\in\mathbb{N}}$ has *no* convergent subsequences, in $\mathbb{R}^X$. In a sense, this example exhibits the worst failure of sequential convergence possible. The closure of $\{p_n\}$ in $\{0,1\}^X$ (or in $\mathbb{R}^X$ for that matter) is homeomorphic to the *Stone-Čech compactification* of the discrete space of natural numbers, usually denoted $\beta\mathbb{N}$, which is an important object of study in general topology.

The following theorem, proved by Haskell Rosenthal in 1974, is fundamental in functional analysis and captures a sharp division in the behavior of sequences in a Banach spaces.

Theorem 1.4 (Rosenthal's Dichotomy, [Ros74])**.** *If X is Polish and $\{f_n\} \subseteq C_p(X)$ is pointwise bounded, then $\{f_n\}_{n\in\mathbb{N}}$ has a convergent subsequence, or a subsequence whose closure in $\mathbb{R}^X$ is homeomorphic to $\beta\mathbb{N}$.*

Rosenthal's Dichotomy states that a pointwise bounded set of continuous functions contains either a convergent subsequence, or a subsequence whose closure is essentially the same as the example mentioned in the previous paragraphs (i.e., "wildest" possible). The genesis of this theorem was Rosenthal's "ℓ_1-Theorem", which states that a Banach space includes an isomorphic copy of ℓ_1 (the space of absolutely summable sequences), or else every bounded sequence therein is weakly Cauchy. The ℓ_1-Theorem connects diverse areas: Banach space geometry, Ramsey theory, set theory, and topology of function spaces.

As we move from $C_p(X)$ to the larger space $B_1(X)$, a dichotomy paralleling the ℓ_1-Theorem holds: Either every point of the closure of a set of functions is a Baire class 1 function, or there is a sequence in the set behaving in the wildest possible way. This result is usually not phrased as a dichotomy, but rather as an equivalence as in Theorem 1.5 below.

First, we introduce some useful notation. For any set $A \subseteq \mathbb{R}^X$ and any real a, define

$$X^A_{\leq a} \coloneqq \bigcap_{f\in A} f^{-1}(-\infty, a] = \{x \in X : f(x) \leq a \text{ for all } f \in A\},$$

$$X^A_{\geq a} \coloneqq \bigcap_{f\in A} f^{-1}[a, +\infty) = \{x \in X : f(x) \geq a \text{ for all } f \in A\}.$$

(In case $A = \emptyset$, we define $X^{\emptyset}_{\geq a} = X = X^{\emptyset}_{\leq a}$.) For any sequence $\{f_n\} \subseteq \mathbb{R}^X$ and $I \subseteq \mathbf{N}$, define $I^{\complement} \coloneqq \mathbb{N} \setminus I$ and $f_I \coloneqq \{f_i : i \in I\}$.

Theorem 1.5 ("The BFT Dichotomy". Bourgain-Fremlin-Talagrand [BFT78])**.** *Let X be a Polish space and $A \subseteq C_p(X)$ be pointwise bounded. The following are equivalent:*

(i) 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 $X^{f_I}_{\leq a} \cap X^{f_{I^\complement}}_{\geq b} = \emptyset$.

(As stated above, the BFT Dichotomy is a particular case of the equivalence (ii)⇔(v) in [BFT78, Corollary 4G].)

The sets $X^{f_I}_{\leq a}$ and $X^{f_{I^\complement}}_{\geq b}$ appearing in condition Theorem 1.5*(ii)* are defined, respectively, in terms of $|I|$-many inequalities of the form $f_i(x) \leq a$, and $|I^\complement|$-many of the form $f_j(x) \geq b$. Thus, at least one of $X^{f_I}_{\leq a}$ and $X^{f_{I^\complement}}_{\geq b}$ is defined by the satisfaction of infinitely (countably) many inequalities. For our purposes, it is more natural to consider only finitely many inequalities at a time, which motivates the definitions below.

Definition 1.6. We say that a function collection $A \subseteq \mathbb{R}^X$ *has the finitary No-Independence Property (NIP)* if, for all sequences $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ and reals $a < b$, there exist finite disjoint sets $E, F \subseteq \mathbb{N}$ such that $X^{f_E}_{\leq a} \cap X^{f_F}_{\geq b} = \emptyset$. We say that such E, F *witness finitary NIP for A, $\{f_n\}$ and a, b.*

A set $A \subseteq \mathbb{R}^X$ *has the finitary Independence Property (IP)* if it does not have finitary NIP, i.e., if there exists a sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ and reals $a < b$ such that for every pair of finite disjoint sets $E, F \subseteq \mathbb{N}$, we have $X^{f_E}_{\leq a} \cap X^{f_F}_{\geq b} \neq \emptyset$.

If the word "finite" is omitted in the above definitions, we obtain the definitions of *countable* NIP (weaker than finitary NIP) and *countable* IP (stronger than finitary IP), respectively.

If we insist on witnesses $E, F \subseteq \mathbb{N}$ such that $F = E^\complement$, we call the respective properties "BFT-NIP" (even weaker than countable NIP) and "BFT-IP" (even stronger than countable IP). Thus, Theorem 1.5 becomes that statement, for pointwise bounded function collections $A \subseteq C_p(X)$, that A is relatively compact in $B_1(X)$ if and only if A has BFT-NIP.

Unless otherwise unspecified, IP/NIP shall mean *finitary* IP/NIP henceforth.

Proposition 1.7. *If X is compact and $A \subseteq C_p(X)$, then A has BFT-NIP if and only if it has finitary NIP.*

(No pointwise boundedness is assumed of A.)

Proof. Trivially (as per the preceding discussion), finitary NIP implies BFT-NIP. Reciprocally, assume that X is compact and has finitary IP. Fix $A \subseteq C_p(X)$, a sequence $\{f_n\} \subseteq A$ and reals $r < s$. For any $I, J \subseteq \mathbb{N}$ (eventually disjoint in applications), write $X_{I,J}$ for $X^{f_I}_{\leq r} \cap X^{f_J}_{\geq s}$. For $I \subseteq I' \subseteq \mathbb{N}$ and $J \subseteq J' \subseteq \mathbb{N}$, we have $X_{I,J} \supseteq X_{I',J'}$; moreover, $X_{I,J} = \bigcap_{E\subseteq I, F\subseteq J} X_{E,F}$, where the index variables $E \subseteq I$, $F \subseteq J$ range over *finite* subsets of I, J, respectively. Clearly, $E, F \subseteq \mathbb{N}$ witness finitary NIP for $\{f_n\}$ if and only if $X^{E,F} = \emptyset$.

Fix $I \subseteq \mathbb{N}$. Since $\{f_n\} \subseteq A \subseteq C_p(X)$ is a sequence of continuous functions, and X is compact, the nested family $\{X_{E,F} : E \subseteq I, F \subseteq I^\complement\}$ consists of closed, thus compact, sets. Since A has finitary IP by hypothesis, the nested family consists of nonempty sets, hence its intersection $X_{I,I^{\complement}} \neq \emptyset$ by compactness. This holds for arbitrary $\{f_n\} \subseteq A$ and $r < s$, so A has BFT-IP. □

Theorem 1.8. *Let X be a metrizable compact (hence Polish) space. For every pointwise bounded $A \subseteq C_p(X)$, the following properties are all equivalent:*

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

(The equivalences *(ii)*⇔*(iii)*⇔*(iv)* hold for arbitrary compact X.)

Proof. Corollary of Theorem 1.5 and Proposition 1.7. □

Theorem 1.8 may be stated as the following dichotomy (under the assumptions): either A is relatively compact in $B_1(X)$, or A has IP (in either sense).

The Independence Property was first isolated by Saharon Shelah in model theory as a dividing line between theories whose models are "tame" (corresponding to NIP) and theories whose models are "wild" (corresponding to IP). See [She71, Definition 4.1], [She90]. We will discuss this dividing line in more detail in the next section.

1.2. **NIP as a universal dividing line between polynomial and exponential complexity.** The particular case of the BFT dichotomy (Theorem 1.5) when A consists of $\{0,1\}$-valued (i.e., $\{\text{Yes}, \text{No}\}$-valued) strings was discovered independently, around 1971-1972 in many foundational contexts related to polynomial ("tame") vs exponential ("wild") complexity: In model theory, by Saharon Shelah [She71], [She90], in combinatorics, by Norbert Sauer [Sau72], and Shelah [She72], [She90], and in statistical learning, by Vladimir Vapnik and Alexey Chervonenkis [VC71], [VC74].

In model theory: Shelah's classification theory is a foundational program in mathematical logic devised to categorize first-order theories based on the complexity and structure of their models. A theory T is considered classifiable in Shelah's sense if the number of non-isomorphic models of T of a given cardinality can be described by a bounded number of numerical invariants. In contrast, a theory T is unclassifiable if the number of models of T of a given cardinality is the maximum possible number. A key fact is that the number of models of T is directly impacted by the number of *types* over sets of parameters in models of T; a controlled number of types is a characteristic of a classifiable theory.

In Shelah's classification program [She90], theories without the independence property (called NIP theories, or dependent theories) have a well-behaved, "tame" structure; the number of types over a set of parameters of size κ of such a theory is of polynomially or similar "slow" growth on κ.

In contrast, theories with the Independence Property (called IP theories) are considered "intractable" or "wild". A theory with the Independence Property produces the maximum possible number of types over a set of parameters; for a set of parameters of cardinality κ, the theory will have $2^{2^{\kappa}}$-many distinct types.

In combinatorics: Sauer [Sau72] and Shelah [She72] proved the following independently: Let $\mathscr{F}$ be a family of subsets of some set S. Either: for every $n \in \mathbb{N}$ there is a set $A \subseteq S$ with $|A| = n$ such that $|\{S_i \cap A : i \in \mathbb{N}\}| = 2^n$

($\mathscr{F}$ has "exponential complexity"); or: there exists $N \in \mathbb{N}$ such that for every $A \subseteq S$ with $|A| \geq N$, one has

$$|\{S_i \cap A : i \in \mathbb{N}\}| \leq \sum_{i=0}^{N-1} \binom{|A|}{i}.$$

($\mathscr{F}$ has "polynomial complexity"). Clearly, any family $\mathscr{F}$ of subsets of a *finite* set S has polynomial complexity. The "polynomial" name is justified: indeed, for fixed $N > 0$, as a function of the size $|A| = m > 0$, we have

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

(More precisely, the order of magnitude is $O(m^{N-1})$: polynomial in m for N fixed.)

In machine learning: Readers familiar with statistical learning may recognize the Sauer-Shelah lemma as the dichotomy discovered and proved slightly earlier (1971) by Vapnik and Chervonenkis [VC71, VC74] to address uniform convergence in statistics. The least integer N given by the preceding paragraph, when it exists, is called the *VC-dimension* of $\mathscr{F}$; it is a core concept in machine learning. If such an integer N does not exist, we say that the VC-dimension of $\mathscr{F}$ is infinite. The lemma provides upper bounds on the number of data points (sample size) needed to learn a concept class of known VC dimension d up to a given admissible error in the statistical sense. The Fundamental Theorem of Statistical Learning states that a hypothesis class is PAC-learnable (PAC stands for "Probably Approximately Correct") if and only if its VC dimension is finite.

1.3. **Rosenthal compacta.** The universal classification implied by Theorem 1.5, as attested by the examples outlined in the preceding section, led to the following definition (by Gilles Godefroy [God80]):

Definition 1.9. A Rosenthal compactum is any topological space realized as a compact subset of the space $B_1(X) = B_1(X, \mathbb{R})$ (equipped with the topology of pointwise convergence) of all real functions of the first Baire class on some Polish space X.

A Rosenthal compactum K is necessarily Hausdorff since it is a topological subspace of the Hausdorff product space $\mathbb{R}^X$.

Rosenthal compacta possess significant topological and dynamical tameness properties, and play an important role in functional analysis, measure theory, dynamical systems, descriptive set theory, and model theory. In this paper, we use them to study deep computations. For this, we shall first focus on countable languages, which is the theme of the next subsection.

1.4. **The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with $\mathcal{P}$ countable.** Fix an arbitrary (at most) countable set $\mathcal{P}$ whose elements $P \in \mathcal{P}$ will be called *predicate symbols* or *formal predicates*. Our present goal is to characterize relatively compact subsets of $B_1(X, \mathbb{R}^{\mathcal{P}})$, where X is always assumed to be a perfectly normal space (often a Polish space).

The set $\mathcal{P}$ shall be considered discrete whenever regarded as a topological space. Since $C_p(X,\mathbb{R}^{\mathcal{P}}) \subseteq B_1(X,\mathbb{R}^{\mathcal{P}}) \subseteq (\mathbb{R}^{\mathcal{P}})^X$, the "ambient" space $(\mathbb{R}^{\mathcal{P}})^X$ is quite relevant. The product $X \times \mathcal{P}$ will be regarded as either a pointset, or as a topological product depending on context. We have natural homeomorphic identifications

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

given by

$$\mathbb{R}^{X\times\mathcal{P}} \to (\mathbb{R}^{\mathcal{P}})^X : \varphi \mapsto \hat{\ }\varphi$$
$$\mathbb{R}^{X\times\mathcal{P}} \to (\mathbb{R}^X)^{\mathcal{P}} : \varphi \mapsto \varphi\hat{\ },$$

where

$$\hat{\ }\varphi(x) \coloneqq \varphi(x,\cdot) \in \mathbb{R}^{\mathcal{P}}, \qquad\qquad \varphi\hat{\ }(P) \coloneqq \varphi(\cdot,P) \in \mathbb{R}^X.$$

Such identifications view X and $\mathcal{P}$ as mere pointsets (the topology of X in particular plays no role).

For $x \in X$, define the "left projection" map

$$\lambda_x : \mathbb{R}^{X\times\mathcal{P}} \to \mathbb{R}^{\mathcal{P}} : \varphi \mapsto \lambda_x(\varphi) \coloneqq \varphi(x,\cdot);$$

for $P \in \mathcal{P}$, the "right projection" map

$$\rho_P : \mathbb{R}^{X\times\mathcal{P}} \to \mathbb{R}^X : \varphi \mapsto \varphi(\cdot,P).$$

For fixed $x \in X$ and $P \in \mathcal{P}$, we also have canonical projection maps

$$\pi_x : \mathbb{R}^X \to \mathbb{R} : f \mapsto f(x), \qquad\qquad \pi_P : \mathbb{R}^{\mathcal{P}} \to \mathbb{R} : f \mapsto f(P).$$

When clear from context, rather than using the specific symbols ("λ" for left, "ρ" for right) to denote projections, we may use the generic symbol "π"; thus, π_x may mean λ_x, and π_P may mean ρ_P.

The Proposition below reduces the study of $\mathbb{R}^{\mathcal{P}}$-valued continuous or Baire-1 functions on X to the special case of real-valued ones.

Proposition 1.10. *The identification $(\mathbb{R}^{\mathcal{P}})^X \cong \mathbb{R}^{X\times\mathcal{P}} \cong (\mathbb{R}^X)^{\mathcal{P}}$ induces identifications*

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

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

Proof. The identification of C_p-spaces follows trivially from the definition of topological product and the fact that $\mathcal{P}$ is discrete: a continuous map $X \to \mathbb{R}^{\mathcal{P}}$ is precisely a $\mathcal{P}$-tuple of continuous functions $X \to \mathbb{R}$, and these correspond to continuous functions $X\times\mathcal{P} \to \mathbb{R}$. The identification of Baire-1 spaces follows immediately, since it is defined in terms of the purely topological notion of limit (in the ambient space) of sequences of continuous functions. □

Given $A \subseteq Y^X$ and $K \subseteq X$ we write $A|_K := \{f|_K : f \in A\}$, i.e., the set of all restrictions of functions in A to K. The following Theorem is a slightly more general version of Theorem 1.5.

Theorem 1.11. *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 $\pi_P \circ A$ ($\subseteq C_p(X)$) is pointwise bounded for every $P \in \mathcal{P}$. The following are equivalent for every compact $K \subseteq X$:*

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

Proof. Compact subsets $K \subseteq X$ are closed, hence also Polish. Therefore, the asserted equivalence follows from Theorems 1.5 and 1.7. $\square$

Lastly, a simple but useful lemma that helps understand when we restrict a set of functions to a specific subspace of the domain space, we may always assume that the subspace is closed, as replacing the subspace by its closure has no effect on NIP.

Lemma 1.12. *Assume that X is Hausdorff and that $A \subseteq C_p(X)$. The following are equivalent for every $L \subseteq X$:*

(i) A_L satisfies the NIP.
(ii) $A|_{\overline{L}}$ satisfies the NIP.

Proof. It suffices to show that (i)$\Rightarrow$(ii). Suppose that (ii) does not hold, i.e., that there are $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ and $a < b$ such that for all finite disjoint $E, F \subseteq \mathbb{N}$:

$$\overline{L} \cap \bigcap_{n\in E} f_n^{-1}(-\infty, a] \cap \bigcap_{n\in F} f_n^{-1}[b, \infty) \neq \emptyset.$$

Pick $a' < b'$ such that $a < a' < b' < b$. Then, for any finite disjoint $E, F \subseteq \mathbb{N}$ we can choose

$$x \in \overline{L} \cap \bigcap_{n\in E} f_n^{-1}(-\infty, a') \cap \bigcap_{n\in F} f_n^{-1}(b', \infty)$$

By definition of closure:

$$L \cap \bigcap_{n\in E} f_n^{-1}(-\infty, a'] \cap \bigcap_{n\in F} f_n^{-1}[b', \infty) \neq \emptyset.$$

This contradicts (i). $\square$

2. Compositional Computation Structures: A structural approach to floating-point computations

In this section, we connect function spaces with floating-point computations. We start by summarizing some basic concepts from [ADIW24].

A *computation states structure* is a pair $(L, \mathcal{P})$, where L is a set whose elements we call *states* and $\mathcal{P}$ is a collection of real-valued functions on L that we call *predicates*. For a state $v \in L$, the *type* of a state v is the indexed family

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

For each $P \in \mathcal{P}$, we call the value $P(v)$ the P-th *feature* of v. A *transition* of a computation states structure $(L, \mathcal{P})$ is a map $f : L \to L$.

Intuitively, L is the set of states of a computation, and the predicates $P \in \mathcal{P}$ are primitives that are given and accepted as computable. Each state $v \in L$ is uniquely characterized by its type $\mathrm{tp}(v)$, so we may identify L with a subset of $\mathbb{R}^{\mathcal{P}}$. Important space states are $L = \mathbb{R}^{\mathbb{N}}$ and $L = \mathbb{R}^n$ for some positive integer n, endowed with predicate $P_i(v) = v_i$, one each for the i-th coordinate of v. We regard the space of types as a topological space, endowed with the topology of pointwise convergence induced by the product topology of $\mathbb{R}^{\mathcal{P}}$. Via the identification $v \mapsto \mathrm{tp}(v)$, the states space L is correspondingly topologized; in particular, for each $P \in \mathcal{P}$, the projection map $v \mapsto P(v)$ is continuous.

Definition 2.1. Given a computation states structure $(L, \mathcal{P})$, any element of $\mathbb{R}^{\mathcal{P}}$ in the image of L under the map $v \mapsto \mathrm{tp}(v)$ will be called a *realized (state) type*. The topological closure of the set of realized types in $\mathbb{R}^{\mathcal{P}}$ (endowed with the pointwise convergence topology) will be called the *space of types* of $(L, \mathcal{P})$, denoted $\mathcal{L}$; elements $\xi \in \mathcal{L}$ are called *state types*. Elements of $\mathcal{L} \setminus L$ will be called *unrealized types*.

Intuitively, state types capture a notion of "ultrastate".

The traditional setting of discrete first-order logic may be regarded as the special case in which predicates are binary (such predicates may take only the values 0 and 1, say). In such a traditional setting, the space of types of a structure is necessarily compact; for some purposes, it serves as a sort of compactification of the structure itself. Compactness in such setting is critical. However, predicates of a computational states structure take values in the non-compact space $\mathbb{R}$, so the corresponding type space $\mathcal{L}$ need not be compact. To bypass this obstacle, we follow the idea, introduced in [ADIW24], of covering $\mathcal{L}$ by "thin" compact subspaces called *shards*.

Definition 2.2. A *sizer* is a tuple $r_\bullet = (r_P)_{P \in \mathcal{P}}$ of positive real numbers, indexed by $\mathcal{P}$. Given a sizer $r_\bullet$, let $\mathbb{R}^{[r_\bullet]} = \prod_{P \in \mathcal{P}} [-r_P, r_P]$ (a compact space), and let the $r_\bullet$*-shard* of a states space L be

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

For a sizer $r_\bullet$, the $r_\bullet$*-type shard* is defined as $\mathcal{L}[r_\bullet] = \overline{L[r_\bullet]}$ (a closed, hence compact subset of $\mathbb{R}^{[r_\bullet]}$).

Let also $\mathcal{L}_{\text{sh}}$ be the union of all type-shards (as the sizer $r_\bullet$ varies).

In general, $\mathcal{L}_{\text{sh}} \subseteq \mathcal{L}$ may be a proper subset. Equality holds in the important special case when $\mathcal{P}$ is countable (see *** below).

2.1. **Compositional Computation Structures.**

Definition 2.3. A *Compositional Computation Structure* (CCS) is a triple $(L, \mathcal{P}, \Gamma)$, where

- $(L, \mathcal{P})$ is a computation states structure, and
- $\Gamma \subseteq L^L$ is a semigroup under composition.

Elements of the semigroup Γ are called the *computations* of the structure $(L, \mathcal{P}, \Gamma)$. We assume that the identity map id on L is an element of Γ (which is thus not merely a semigroup but a monoid of transformations of L).

We topologize Γ as a subset of the topological product L^L, where the "exponent" L serves merely as an index set, but the "base" L is topologized by type; consequently, one may identify Γ with a subset of the topological product $(\mathbb{R}^{\mathcal{P}})^L$. More specifically, Γ is identified with a subset of $\mathcal{L}^L$, which is a closed subspace of $(\mathbb{R}^{\mathcal{P}})^L$. Therefore, we have an inclusion $\overline{\Gamma} \subseteq \mathcal{L}^L$. Elements $\xi \in \overline{\Gamma}$ are called (real-valued) *deep computations* or *ultracomputations*.

The reason why we require Γ to be a semigroup is because in many practical applications we want to perform an iterative process of computations (e.g., see subsection 2.3). In these scenarios we need the set of computations to be closed under composition. This leads to other concepts that are not addressed in this paper but are rather discussed in [ADIW24], Section 5. However, in other applications we do not need to work on a set of computations that is closed under composition (e.g., see subsection 2.4). Given a set $\Delta \subseteq L^L$ of computations (not necessarily a semigroup), we can always take the semigroup Γ generated by Δ, i.e., the smallest semigroup containing Δ.

A collection R of sizers is *exhaustive* if $L = \bigcup_{r_\bullet \in R} L[r_\bullet]$ (shards $L[r_\bullet]$ exhaust L). A transformation $\gamma \in \Gamma$ is *R-confined* if γ restricts to a map $\gamma|_{L[r_\bullet]} : L[r_\bullet] \to L[r_\bullet]$ (into $L[r_\bullet]$ itself) for every $r_\bullet \in R$. A subset $\Delta \subseteq \Gamma$ is *R-confined* if each $\gamma \in \Delta$ is.

Proposition 2.4. *If $\Delta \subseteq \Gamma$ is confined by an exhaustive sizer collection, then $\overline{\Delta}$ is a compact subset of $\mathcal{L}^L_{\mathrm{sh}}$.*

Proof. Assume that R confines Δ. For each $v \in L$, let $r_\bullet^{(v)} \in R$ be a sizer such that $v \in L[r_\bullet^{(v)}]$. An arbitrary $\gamma \in \Delta$ restricts to a map $\gamma \restriction L[r_\bullet^{(v)}] : L[r_\bullet^{(v)}] \to L[r_\bullet^{(v)}]$, so $\Gamma \subseteq K := \prod_{v \in L} \mathcal{L}[r_\bullet^{(v)}]$. The space K is a product of compact spaces, hence compact, so $\overline{\Gamma}$ is a closed, hence compact subset thereof, and a subset of $\mathcal{L}^L_{\mathrm{sh}} \supseteq K$ *a fortiori.* $\square$

For a CCS $(L, \mathcal{P}, \Gamma)$, we regard the elements of Γ as "standard" finitary computations, and the elements of $\overline{\Gamma}$, i.e., deep computations, as possibly infinitary limits of standard computations. The main goal of this paper is to study the computability, definability and computational complexity of deep computations. Since ultracomputations are defined through a combination of topological concepts (namely, topological closure) and structural and model-theoretic concepts (namely, models and types), we will import technology from both topology and model theory.

2.2. **Computability and definability of deep computations and the Extendibility Axiom.** Let $f : L \to \mathcal{L}$ be a function that maps each input state type $(P(v))_{P \in \mathcal{P}} \in \mathbb{R}^{\mathcal{P}}$ to an output state type $(P \circ f(v))_{P \in \mathcal{P}} \in \mathbb{R}^{\mathcal{P}}$.

(1) We will say that f is *definable* if for each $Q \in \mathcal{P}$, the output feature $Q \circ f : L \to \mathbb{R}$ is a definable predicate in the following sense: There is an *approximating function* $\varphi_{Q,\,K,\varepsilon} : L \to \mathcal{L}$ that can be built recursively out of a finite number of the (primitively computable) predicates in $\mathcal{P}$ and by a finite number of iterations of the finitary lattice operations $\wedge$ (=min) and $\vee$ (=max), the operations of $\mathbb{R}^{\mathbb{R}}$ as a vector algebra (that is, vector addition and multiplication and scalar multiplication) and the operators sup and inf applied on individual variables from L, and such that

$$|\varphi_{Q,\,K,\varepsilon}(v) - Q \circ f(v)| \le \epsilon, \qquad \text{for all } v \in K.$$

Remark: What we have defined above is a model-theoretic concept; it is a special case of the concept of *first-order definability* for real-valued predicates in the model the theory of real-valued structures first introduced in [Iov94] for model theory of functional analysis and now standard in model theory (see [Kei03]). The $\wedge$ (=min) and $\vee$ (=max) operations correspond to the positive Boolean logical connectives "and" and "or", and the sup and inf operators correspond to the first-order quantifiers, $\forall$ and $\exists$.

(2) We will say that f is *computable* if it is definable in the sense defined above under (1), but without the use of the sup/inf operators; in other words, if for every choice of Q, K, ε, the approximation function $\varphi_{Q,K,\varepsilon} : L \to \mathcal{L}$ can be constructed without any use of sup or inf operators. This is quantifier-free definability (i.e., definability as given by the preceding paragraph, but without use of quantifiers), which, from a logic viewpoint, corresponds to computability (the presence of the quantifiers $\exists$ and $\forall$ are the reason behind the undecidability of first-order logic).

It is shown in [ADIW24] that:

(1) For a definable $f : L \to \mathcal{L}$, the approximating functions $\varphi_{Q,K,\varepsilon}$ may be taken to be *polynomials* of the input features, and
(2) Definable transforms $f : L \to \mathcal{L}$ are precisely those that extend to continuous $\tilde{f} : \mathcal{L} \to \mathcal{L}$.

To summarize, a function $f : L \to \mathcal{L}$ is computable if and only if it is definable if and only if it is polynomially approximable if and only it can be extended to a continuous $\tilde{f} : \mathcal{L} \to \mathcal{L}$. This motivates the following definition.

Definition 2.5. We say that a CCS $(L, \mathcal{P}, \Gamma)$ satisfies the *Extendibility Axiom* if for all $\gamma \in \Gamma$, there is $\tilde{\gamma} : \mathcal{L}_{\text{sh}} \to \mathcal{L}_{\text{sh}}$ such that for every sizer $r_\bullet$ there is a sizer $s_\bullet$ such that $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \to \mathcal{L}[s_\bullet]$ is continuous. We refer to $\tilde{\gamma}$ as a *free* extension of γ.

In practice, we should expect that finitary computations (i.e., elements in Γ) are describing continuous functions. This is precisely what the Extendibility Axiom is saying. For example, suppose we want to approximate e^π. We can't input the exact value of π in a computer, but we can work with finite approximations. The fact that e^x is continuous is telling us that good enough approximations of π yield good enough approximations of e^π. We want to emphasize that, on the other hand, deep computations may become discontinuous even if they are the pointwise limit of (continuous) computations. A simple illustrative example of this phenomenon is example 2.6.

For the rest of the paper, fix for each $\gamma \in \Gamma$ a free extension $\tilde{\gamma}$ of γ. For any $\Delta \subseteq \Gamma$, let $\tilde{\Delta}$ denote $\{\tilde{\gamma} \colon \gamma \in \Delta\}$.

For a more detailed discussion of the Extendibility Axiom, we refer the reader to [ADIW24].

2.3. Newton's method as a CCS.

Let $p(z)$ be a non-constant polynomial with complex coefficients. Newton's method is an iterative method that is used to approximate a root of $p(z)$. Define the *Newton map* as:

$$N_p(z) = z - \frac{p(z)}{p'(z)}$$

The method consists of taking an initial guess $z_0 \in \mathbb{C}$ and iterating the rational map N_p to obtain a sequence given by

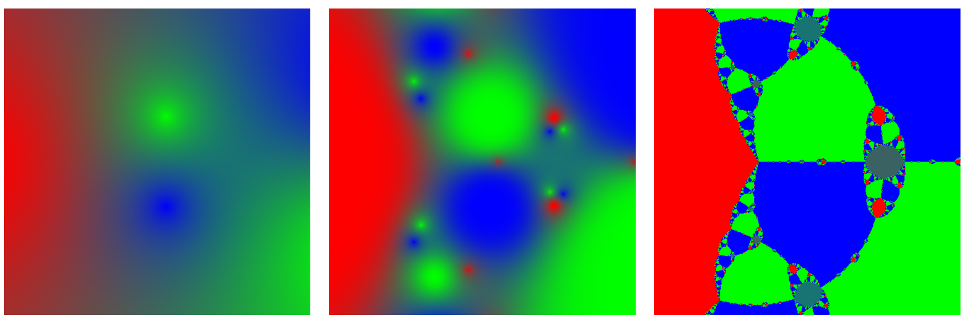
$$z_{n+1} = N_p(z_n) = z_n - \frac{p(z_n)}{p'(z_n)}$$

For each root $r \in \mathbb{C}$ of $p(z)$, there exists $\varepsilon > 0$ such that for any initial guess z_0 in the ε-ball centered at r, Newton's iteration converges to r (provided $p'(r) \neq 0$) and the convergence is quadratic in that case, meaning the error at each step is roughly squared, causing the number of correct digits to double, leading to fast convergence.

Denote by $N_p^n := N_p \circ N_p \circ \cdots \circ N_p$ (nth iteration of N_p). Given a root r of $p(z)$, the set $B_r = \{z \in \mathbb{C} : \lim_{n\to\infty} N_p^n(z) = r\}$ is an open set called the *basin* of r. However, Newton's method can fail to converge to any root for some choices of z_0. For example, consider the polynomial $p(z) = z^3 - 2z + 2$. The Newton map is given by

$$N_p(z) = z - \frac{z^3 - 2z + 2}{3z^2 - 2} = \frac{2z^3 - 2}{3z^2 - 2}$$

Notice that taking $z_0 = 0$ as an initial guess will yield the sequence $0, 1, 0, 1, 0, 1, \ldots$ that oscillates between 0 and 1 but none of them are roots of $p(z)$. Another more chaotic way Newton's method can fail to converge is when the sequence of iterations has no convergent subsequence. The set of such points is known as the *Julia set* associated to N_p and it is typically a fractal. This can be visualizaed by adding a dash of color: let us give each complex number z_0 a color (R, G, B) where $R, G, B \in [0, 1]$ (so that $(1, 0, 0)$ is red, $(0, 1, 0)$ is green, $(0, 0, 1)$ is blue and $(0.5, 0, 0.5)$ is a light purple, for example). The values of R, G and B are determined by looking at the image of said number at each stage of the iteration, $N_p^n(z_0)$, and computing the current distance to each of the roots of $p(z)$; so $R = 1/d_r$ where d_r is the positive distance to the root which is colored red, and so on. In this way, the roots themselves are colored red, green, and blue, and every other point gets a mix of the three colors. As the number of iterations increases, each point gets a sharper color, as the sequence of images $\{N_p^n(z_0)\}_{n=1}^{\infty}$ converges to one of the three roots. Each stage, the complex plane looks as if out of focus because the coloring function is continuous. As the reader can see in Figure 1, the points at the boundary of each color class form the famous Newton fractal (of which, interestingly, Newton was unaware).



(A) After 1 iteration (B) After 3 iterations (C) After 100 iterations

FIGURE 1. Newton's method approximating $p(z) = z^3 - 2z + 2$. Notice the regions of divergence.

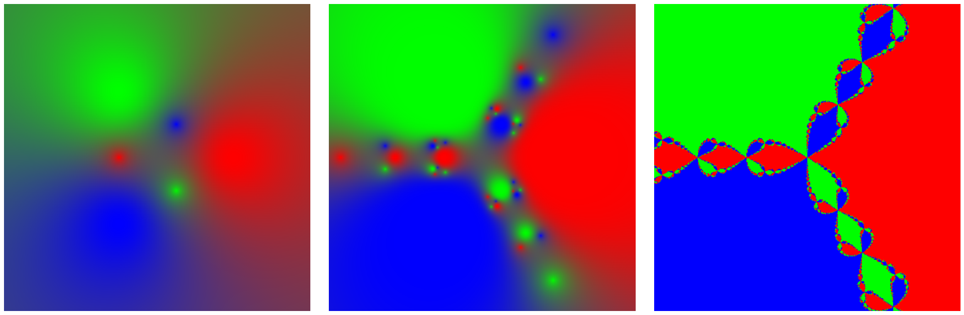
Another example of a Newton's fractal is for $p(z) = z^3 - 1$. The roots of $p(z)$ are the 3rd roots of unity and the Newton map is given by:

$$N_p(z) = z - \frac{z^3 - 1}{3z^2} = \frac{2z^3 + 1}{3z^2}$$

In this case, there are three basins of attraction (one for each root) and the complement of their union is the Julia set, i.e., the common boundary.

We can frame Newton's method as a CCS (satisfying the Extendibility Axiom) as follows. Given a non-constant polynomial $p(z)$, let $L \subseteq \mathbb{C}$ be the complement of the Julia set (this is called the *Fatou set*) associated to N_p. This is the set where Newton's method is tame. By [Bla84, Corollary 4.6], L is open and dense in $\mathbb{C}$. Let $\mathcal{P} = \{P_1, P_2, P_3\}$ where the map $z \mapsto (P_1(z), P_2(z), P_3(z))$ is the stereographic projection into the Riemann Sphere S^2, i.e.,

$$P_1(z) = \frac{2\mathrm{Re}(z)}{|z|^2 + 1},$$



(A) After 1 iteration (B) After 3 iterations (C) After 100 iterations

FIGURE 2. Newton's method approximating $p(z) = z^3 - 1$.

$$P_2(z) = \frac{2\mathrm{Im}(z)}{|z|^2 + 1},$$

$$P_3(z) = \frac{|z|^2 - 1}{|z|^2 + 1}.$$

Since L is dense in the extended complex plane, then $\mathcal{L} = S^2$. Computations are the iterations of Newton's method, i.e., $\Delta = \Gamma = \{N_p^n : n \in \mathbb{N}\}$. Then, $(L, \mathcal{P}, \Gamma)$ is a CCS. Since all computations N_p^n are rational maps, they can be continuously extended to the extended complex plane $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, i.e., to $\mathcal{L}$. In particular, $(L, \mathcal{P}, \Gamma)$ satisfy the Extendibility Axiom. The set of deep computations $\overline{\Delta}$ might behave different for various polynomials. Let us look at various examples:

Example 2.6. Computation of square roots. Let a be a positive real number and $p(x) = x^2 - a$. Let $L = \mathbb{R}\backslash\{0\}$. Let $\mathcal{P} = (P_1, P_2)$ where $x \mapsto (P_1(x), P_2(x))$ is the stereographic projection into $S^1 \subseteq \mathbb{R}^2$, i.e.,

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

$$P_2(x) = \frac{x^2 - 1}{x^2 + 1}.$$

The Newton's method map $N_p : L \to L$ is given by

$$N_p(x) = \frac{x^2 + a}{2x}.$$

Note that $\mathcal{L} = S^1$ and that each iterate N_p^n can be continuously extended to the extended real line $\mathbb{R} \cup \{\infty\}$, i.e., $\mathcal{L}$. For example,

$$\tilde{N_p}(x) = \begin{cases} \frac{x^2+a}{2x}, & \text{if } x \in L; \\ \infty, & \text{if } x = 0, \infty. \end{cases}$$

For every initial guess $x \in L$, the limit $f(x) = \lim_{n\to\infty} N_p^n(x)$ converges pointwise to one of the roots. Moreover,

$$f(x) = \begin{cases} \sqrt{a}, & \text{if } x > 0; \\ -\sqrt{a}, & \text{if } x < 0. \end{cases}$$

Notice that f can be extended to $\mathcal{L}$ by

$$\tilde{f}(x) = \begin{cases} \sqrt{a}, & \text{if } x > 0; \\ -\sqrt{a}, & \text{if } x < 0; \\ \infty, & \text{if } x = 0, \infty. \end{cases}$$

However, $\tilde{f} : \mathcal{L} \to \mathcal{L}$ is not continuous. The set $\overline{\tilde{\Delta}}$ of deep computations is $\tilde{\Delta} \cup \{\tilde{f}\} \subseteq B_1(\mathcal{L}, \mathcal{L})$.

Example 2.7. Newton's method for $p(z) = z^3 - 2z + 2$. Let r_1, r_2 and r_3 be the three roots of $p(z)$. Let B_1, B_2 and B_3 be their respective basins. Let B be the basin of the attractive cycle $0, 1, 0, 1, \ldots$. Let $L = B \cup \bigcup_{i=1}^{3} B_i$ and J be its complement, i.e., the Julia set. Notice that N_p^n does not converge pointwise. However, the subsequences N_p^{2n} and N_p^{2n+1} are pointwise convergent to functions f_1 and f_2 respectively. f_1 and f_2 are two distinct deep computations. Note that for $z \in J$, no subsequence of $\tilde{N_p^n}(z)$ converges to a complex number. However, since $\mathcal{L} = S^2$ is compact there is a subsequence of $\tilde{N_p^n}(z)$ that converges to ∞. We can extend $f_i : L \to \mathcal{L}$ to $\tilde{f}_i : \mathcal{L} \to \mathcal{L}$ by:

$$\tilde{f}_i(z) = \begin{cases} f_i(z), & \text{if } z \in L; \\ \infty, & \text{if } z \in J. \end{cases}.$$

Again, note that $\tilde{f}_i$ for $i = 1, 2$ are not continuous and that $\tilde{f}_i \in \overline{\tilde{\Delta}}$.

2.4. **Finite precision threshold classifiers as a CCS.** Let $L = 2^{\mathbb{N}}$, i.e., the set consisting of all infinite binary sequences with the topology of pointwise convergence. Let $\mathcal{P} = \{P_n : n \in \mathbb{N}\}$ is the collection of projections, i.e., $P_n(x) = x(n)$ for each n and each $x \in L$. In other words, P_n is the predicate that reads the nth digit of a sequence. Notice that $L \subseteq \mathbb{R}^{\mathcal{P}}$ is closed. Therefore, $\mathcal{L} = L$. We denote by 0^∞ the infinite binary sequence consisting of 0s, and by 1^∞ the infinite binary sequence consisting of 1s. The set of finite binary strings is denoted by $2^{<\mathbb{N}}$. Given a finite binary string w, we consider the transition $\phi_w : L \to L$ given by the rule

$$\phi_w(x) = \begin{cases} 1^\infty, & \text{if } x|_{|w|} \leq_{\text{lex}} w; \\ 0^\infty, & \text{otherwise}; \end{cases}$$

where $|w|$ is the length of the string w and $x|_{|w|}$ is the prefix of x of length $|w|$. That is, $\phi_w(x)$ is equal to the constant sequence of ones if $x|_{|w|}$ comes before or is equal to w in the lexicographic order of strings, and it is equal to the constant sequence of zeros otherwise. In words, ϕ_w checks if a number is less than or equal to the scalar value of threshold w (the string w is finite, hence the classifier has *finite precision*). Note that $P_n \circ \phi_w(x) = 1$ if and only if $x|_{|w|}$ comes before w.

Proposition 2.8. *$\phi_w : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ is continuous for all $w \in 2^{<\mathbb{N}}$.*

Proof. It suffices to prove that $P_n \circ \phi_w : 2^{\mathbb{N}} \to \{0, 1\}$ is continuous for all $n \in \mathbb{N}$. For simplicity, let us call $f := P_n \circ \phi_w$, i.e.,

$$f(x) = \begin{cases} 1, & \text{if } x|_{|w|} \leq_{\text{lex}} w; \\ 0, & \text{otherwise}. \end{cases}$$

We first prove that $f^{-1}(1) = \{x \in 2^{\mathbb{N}} : x|_{|w|} \leq_{\text{lex}} w\}$ is an open set. Fix $x_0 \in f^{-1}(1)$. Let $t := x_0|_{|w|}$ and consider the basic open set $[t] = \{x \in 2^{\mathbb{N}} : x|_{|t|} = t\}$.

Then, it is not difficult to check that $x_0 \in [t] \subseteq f^{-1}(1)$. The same reasoning shows that $f^{-1}(0)$ is open. $\square$

Let $\Delta = \{\phi_w : w \in 2^{<\mathbb{N}}\} \cup \{\mathbf{0}^\infty, \mathbf{1}^\infty\}$ where $\mathbf{0}^\infty, \mathbf{1}^\infty : L \to L$ are the constant maps identical to 0^∞ and 1^∞, respectively. Let Γ be the semigroup generated by Δ. The preceding proposition shows that Δ (and hence Γ) consists of continuous functions. In particular, the CCS $(L, \mathcal{P}, \Gamma)$ satisfies the Extendibility Axiom. In contrast with Newton's method, the algebraic structure of Δ is quite simple: composing two classifiers results in something similar to a Boolean logic gate. The topological structure is far more interesting. Intuitively, the crucial difference between Newton's method and threshold classifiers is that the complexity of the former comes from *depth*: the semigroup is generated by a single map but its iterates are highly complex. The complexity of threshold classification comes from *width*: the semigroup has infinitely many generators, but their compositions are simple.

Intuitively, the closure of Δ consists of the set of all possible threshold classifiers on the real line, but there are two sorts: the ones that classify strict inequalities and those that classify $\leq$. The members of Δ are finite-precision approximations of classifiers that check all bits of information. But here it gets interesting: what is the difference, in terms of floating-point arithmetic, between $x < 0.5$ and $x \leq 0.5$?

Suppose that f_a^+ represents the $\leq$ classifier for a target $a \in L$. We identify the scalar truth values with constant sequences, formally $f_a^+ : L \to \{0^\infty, 1^\infty\}$ is defined by $f_a^+(x) = 1^\infty$ if $x \leq_{\text{lex}} a$ and $f_a^+(x) = 0^\infty$ otherwise. Note that if a is the constant 1^∞, then $f_a^+ = \mathbf{1}^\infty$. Similarly, let f_a^- be the strict inequality $<$ classifier, i.e., $f_a^-(x) = 1^\infty$ if $x <_{\text{lex}} a$ and $f_a^-(x) = 0^\infty$ otherwise. Note that if a is the constant zero, then $f_a^- = \mathbf{0}^\infty$.

Proposition 2.9. *$f_a^+, f_a^- \in \overline{\Delta}$ for all $a \in 2^{\mathbb{N}}$.*

Proof. First, we show that $f_a^+ \in \overline{\Delta}$. If $a = 1^\infty$, then $f_a^+ = \mathbf{1}^\infty \in \Delta$. If a is not identically 1, we argue that the pointwise limit of the threshold classifiers on $w_n := a|_n^\frown 1$ (that is, the sequence obtained from appending a 1 to the first n bits of a) is precisely f_a^+. Specifically, for every $x \in L$, we intend to prove that $\lim_{n\to\infty} \phi_{w_n}(x) = f_a^+(x)$. Assume that $x >_{\text{lex}} a$. Let m be the least index at which the two sequences differ. Then $a(m) = 0 < 1 = x(m)$, and for all $n \geq m$, w_n agrees with a up to m. Crucially, $w_n(m) = 0 < 1 = x(m)$, which implies that $w_n <_{\text{lex}} x|_{n+1}$, and hence $\phi_{w_n}(x) = 0^\infty = f_a^+(x)$ for large enough n. If $x \leq_{\text{lex}} a$, then $x|_{n+1} \leq_{\text{lex}} w_n$ for all $n \in \mathbb{N}$. Hence, $\phi_{w_n}(x) = 1^\infty = f_a^+(x)$ for all $n \in \mathbb{N}$.

Now, we prove that $f_a^- \in \overline{\Delta}$. If a is the constant zero, then $f_a^- = \mathbf{0}^\infty \in \Delta$. Suppose that a is not constantly zero; then we have two cases.

(1) If a is eventually zero (a is often called a *dyadic rational*), that is $a = u^\frown 1^\frown 0^\infty$ (here $^\frown$ denotes concatenation) for some finite u. Let $w_n := u^\frown 0^\frown 1^n <_{\text{lex}} a$. We claim that $\lim_{n\to\infty} \phi_{w_n}(x) = f_a^-(x)$. Assume that $x <_{\text{lex}} a$. Then, $x|_{|w_n|} \leq_{\text{lex}} w_n$ for large enough n. Hence, $\phi_{w_n}(x) = 1^\infty = f_a^-(x)$ for large enough n. Now assume that $x \geq_{\text{lex}} a$. Then, $w_n <_{\text{lex}} a|_{|w_n|} \leq_{\text{lex}} x|_{|w_n|}$ so $\phi_{w_n}(x) = 0^\infty = f_a^-(x)$ for all $n \in \mathbb{N}$.

(2) If a is not eventually zero, enumerate the indices of all positive bits in a, $\{n \in \mathbb{N} : a(n) = 1\}$, strictly increasingly as $\{n_k : k \in \mathbb{N}\}$ (this is possible as the former set is infinite by assumption). Let $w_k := (a|_{n_k - 1})^\frown 0$; that is, w_k is the result of flipping the k-th positive bit in a. Once again, observe

that $w_k <_{\text{lex}} a$ for all k. The crucial step follows from the fact that for any $x <_{\text{lex}} a$, there is a large enough K such that $x <_{\text{lex}} w_k$ for all $k \geq K$.

$\square$

The preceding proposition shows that the topological structure of deep computations can be very complicated. Indeed, $\overline{P_n \circ \Delta}$ contains the *Split Cantor* space for all $n \in \mathbb{N}$. (see Examples 3.3(3)).

2.5. **Finite precision prefix test.** In this subsection we present another example of a CCS with a more complicated set of deep computations. Let $L = 2^{\mathbb{N}}$ and $\mathcal{P} = \{P_n : n \in \mathbb{N}\}$ where $P_n(x) = x(n)$ are the projection maps so clearly $L \subseteq \mathbb{R}^{\mathcal{P}}$ and $\mathcal{L} = L$ (same computation states structure as subsection 2.4). For each $w \in 2^{<\mathbb{N}}$, let $\psi_w : L \to L$ be the transition given by:

$$\psi_w(x) = \begin{cases} 1^\infty, & \text{if } x|_{|w|} = w; \\ 0^\infty, & \text{otherwise.} \end{cases}$$

In other words, ψ_w determines whether the first $|w|$ bits of a binary sequence is exactly w. Let $\Delta = \{\psi_w : w \in 2^{<\mathbb{N}}\}$ and Γ be the semigroup generated by Δ. Since the sets $\{x \in 2^{\mathbb{N}} : x|_{|w|} = w\}$ are open and closed in $2^{\mathbb{N}}$, then the transitions ψ_w are all continuous. In particular, $(L, \mathcal{P}, \Gamma)$ satisfies the Extendibility Axiom.

Let us analyze the set of deep computations of Δ. The idea of these finite precision prefix tests ψ_w is that they are approximating the equality relation on infinite binary sequences. For a given $a \in 2^{\mathbb{N}}$, let $\delta_a : L \to \{0^\infty, 1^\infty\}$ be the indicator function at a, i.e., $\delta_a(x) = 1^\infty$ if $x = a$ and $\delta_a(x) = 0^\infty$ otherwise.

Proposition 2.10. *$\delta_a \in \overline{\Delta}$ for all $a \in 2^{\mathbb{N}}$.*

Proof. Fix $a \in 2^{\mathbb{N}}$. Let $w_n := a|_n$ for each $n \in \mathbb{N}$. We claim that $\lim_{n\to\infty} \psi_{w_n}(x) = \delta_a(x)$ for all $x \in L$. If $x = a$, then $x|_{|w_n|} = w_n$ for all n and so $\psi_{w_n}(x) = 1^\infty = \delta_a(x)$ for all n. If $x \neq a$, then $x|_{|w_n|} \neq w_n$ for large enough n. Hence, $\psi_{w_n}(x) = 0^\infty = \delta_a(x)$ for large enough n. $\square$

These equality tests δ_a are not all the deep computations. The other deep computation we are missing is the constant map $\mathbf{0}^\infty$.

Proposition 2.11. *$\mathbf{0}^\infty \in \overline{\Delta}$.*

Proof. To show that $\mathbf{0}^\infty \in \overline{\Delta}$ consider, for each $n \in \mathbb{N}$, $w_n = 1^n {}^\frown 0$, i.e., the string consisting of n consecutive 1s followed by a 0. If $x = 1^\infty$, then $x|_{|w_n|} \neq w_n$ for all $n \in \mathbb{N}$. Hence, $\psi_{w_n}(x) = 0^\infty$ for all $n \in \mathbb{N}$. If $x \neq 1^\infty$, let N be the smallest such that $x(N) = 0$. Then, $x|_{|w_n|} \neq w_n$ for all $n > N$. Hence, $\psi_{w_n}(x) = 0^\infty$ for large enough n. $\square$

In fact, $\overline{\Delta} = \Delta \cup \{\delta_a : a \in 2^{\mathbb{N}}\} \cup \{\mathbf{0}^\infty\}$ and this space is known as the *Extended Alexandroff compactification of* $2^{\mathbb{N}}$ (see Example 3.3(2)). One key topological property about this space is that $\mathbf{0}^\infty$ is not a G_δ point, i.e., $\{\mathbf{0}^\infty\}$ is not a countable intersection of open sets. Moreover, $\mathbf{0}^\infty$ is the only non-G_δ point. It is well-known that in a Hausdorff, first countable space every point is G_δ. This shows that our space of deep computations is not first countable. This space also contains a discrete subspace of size continuum, namely $\{\delta_a : a \in 2^{\mathbb{N}}\}$.

3. Classifying deep computations

3.1. **NIP, Rosenthal compacta, and deep computations.** Under what conditions are deep computations Baire class 1, and thus well-behaved according to our framework, on type-shards? The following theorem says that, under the assumption that $\mathcal{P}$ is countable, the space of deep computations is a Rosenthal compactum (when restricted to shards) if and only if the set of computations satisfies the NIP feature by feature. Hence, we can import the theory of Rosenthal compacta into this framework of deep computations.

Theorem 3.1. *Let $(L, \mathcal{P}, \Gamma)$ be a compositional computational structure (Definition 2.3) satisfying the Extendibility Axiom (Definition 2.5) with $\mathcal{P}$ countable. Let R be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be R-confined. The following are equivalent.*

- *(i)* $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ *for all* $r_\bullet \in R$.
- *(ii)* $\pi_P \circ \Delta|_{L[r_\bullet]}$ *satisfies the NIP for all* $P \in \mathcal{P}$ *and* $r_\bullet \in R$*; that is, for all* $P \in \mathcal{P}$, $r_\bullet \in R$, $a < b$, $\{\gamma_n\}_{n\in\mathbb{N}} \subseteq \Delta$ *there are finite disjoint* $E, F \subseteq \mathbb{N}$ *such that*

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

Moreover, if any (hence all) of the preceding conditions hold, then every deep computation $f \in \overline{\Delta}$ can be extended to a Baire-1 function on shards, i.e., there is $\tilde{f} : \mathcal{L}_{sh} \to \mathcal{L}_{sh}$ such that $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ for all $r_\bullet \in R$. In particular, on each shard every deep computation is the pointwise limit of a countable sequence of computations.

Proof. Since $\mathcal{P}$ is countable, $\mathcal{L}[r_\bullet] \subseteq \mathbb{R}^{\mathcal{P}}$ is Polish. Also, the Extendibility Axiom implies that $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is a pointwise bounded set of continuous functions for all $P \in \mathcal{P}$. Hence, Theorem 1.11 and Lemma 1.12 prove the equivalence of (i) and (ii). If (i) holds and $f \in \overline{\Delta}$, then write $f = \mathcal{U}\lim_i \gamma_i$ as an ultralimit. Define $\tilde{f} := \mathcal{U}\lim_i \tilde{\gamma}_i$. Hence, for all $r_\bullet \in R$ we have $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$. That every deep computation is a pointwise limit of a countable sequence of computations follows from the fact that for a Polish space X every compact subset of $B_1(X)$ is Fréchet-Urysohn (that is, a space where topological closures coincide with sequential closures, see Theorem 3F in [BFT78] or Theorem 4.1 in [Deb13]). ☐

3.2. **The Todorčević trichotomy and levels of PAC learnability.** In this subsection we study the case when the set of deep computations is a separable Rosenthal compactum. We are interested in the separable case for two reasons:

1. In practice, the set Δ of computations is countable. This implies that the set $\overline{\Delta}$ of deep computations is separable.
2. The non-separable case lacks some tools and nice examples, which makes their study more complicated. In the separable case we have two important results, which are introduced in this subsection (Todorčević's Trichotomy) and the next subsection (Argyros-Dodos-Kanellopoulos heptachotomy). By introducing Todorčević's Trichotomy into this framework, we obtain a classification of the complexity of deep computations.

Given a countable set Δ of computations satisfying the NIP on features and shards (condition (ii) of Theorem 3.1) we have that $\overline{\tilde{\Delta}_{\mathcal{L}[r_\bullet]}}$ (for a fixed sizer $r_\bullet$)

is a separable *Rosenthal compactum* (see Definition 1.9). Todorčević proved a remarkable trichotomy for Rosenthal compacta [Tod99], and later Argyros, Dodos, Kanellopoulos [ADK08] proved an heptachotomy that refined Todorčević's classification. In this section, inspired by the work of Glasner and Megrelishvili [GM22], we study ways in which this classification allows us to obtain different levels of PAC-learnability and NIP.

Recall that a topological space X is *hereditarily separable* if every subspace is separable, and that X is *first countable* if every point in X has a countable local basis. Every separable metrizable space is hereditarily separable, and R. Pol proved that every hereditarily separable Rosenthal compactum is first countable (see section 10 of [Deb13]). This suggests the following definition:

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

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

Observe that $\mathrm{NIP}_3 \Rightarrow \mathrm{NIP}_2 \Rightarrow \mathrm{NIP}_1 \Rightarrow \mathrm{NIP}$. In [Tod99], Todorčević isolates three canonical examples of Rosenthal compacta that witness the failure of the converse implications above.

We now present some separable and non-separable examples of Rosenthal compacta. These show that the previously discussed classes NIP_i are not equal.

Examples 3.3.

(1) *Alexandroff compactification of a discrete space of size continuum.* For each $a \in 2^{\mathbb{N}}$ consider the map $\delta_a : 2^{\mathbb{N}} \to \mathbb{R}$ given by $\delta_a(x) = 1$ if $x = a$ and $\delta_a(x) = 0$ otherwise. Let $A(2^{\mathbb{N}}) = \{\delta_a : a \in 2^{\mathbb{N}}\} \cup \{0\}$, where 0 is the zero map. Notice that $A(2^{\mathbb{N}})$ is a compact subset of $B_1(2^{\mathbb{N}})$, in fact $\{\delta_a : a \in 2^{\mathbb{N}}\}$ is a discrete subspace of $B_1(2^{\mathbb{N}})$ and its pointwise closure is precisely $A(2^{\mathbb{N}})$. Hence, this is a Rosenthal compactum which is not first countable. Notice that this space is also not separable.

(2) *Extended Alexandroff compactification.* For each finite binary sequence $s \in 2^{<\mathbb{N}}$, let $v_s : 2^{\mathbb{N}} \to \mathbb{R}$ be given by $v_s(x) = 1$ if x extends s and $v_s(x) = 0$ otherwise. Let $\hat{A}(2^{\mathbb{N}})$ be the pointwise closure of $\{v_s : s \in 2^{<\mathbb{N}}\}$, i.e., $\hat{A}(2^{\mathbb{N}}) = A(2^{\mathbb{N}}) \cup \{v_s : s \in 2^{<\mathbb{N}}\}$. Note that this space is a separable Rosenthal compactum which is not first countable. This is the example discussed in Section 2.5. It is an example of a CCS that is NIP but not NIP_1.

(3) *Split Cantor.* Let $<$ be the lexicographic order in the space of infinite binary sequences, i.e., $2^{\mathbb{N}}$. For each $a \in 2^{\mathbb{N}}$ let $f_a^- : 2^{\mathbb{N}} \to \mathbb{R}$ be given by $f_a^-(x) = 1$ if $x < a$ and $f_a^-(x) = 0$ otherwise. Let $f_a^+ : 2^{\mathbb{N}} \to \mathbb{R}$ be given by $f_a^+(x) = 1$ if $x \leq a$ and $f_a^+(x) = 0$ otherwise. The split Cantor is the space $S(2^{\mathbb{N}}) = \{f_a^- : a \in 2^{\mathbb{N}}\} \cup \{f_a^+ : a \in 2^{\mathbb{N}}\}$, which was obtained as the closure of the space discussed in Section 2.4, giving an example separating NIP_2 from NIP_3. This is a separable Rosenthal compactum. One example of a countable dense subset is the set of all f_a^+ and f_a^- where a is an infinite

binary sequence that is eventually constant. Moreover, it is hereditarily separable, but it is not metrizable.

(4) *Alexandroff Duplicate*. Let K be any compact metric space and consider the Polish space $X = C(K) \sqcup K$, i.e., the disjoint union of $C(K)$ (with its supremum norm topology) and K. For each $a \in K$ define $g_a^0, g_a^1 : X \to \mathbb{R}$ as follows:

$$g_a^0(x) = \begin{cases} x(a), & \text{if } x \in C(K); \\ 0, & \text{if } x \in K; \end{cases}$$

$$g_a^1(x) = \begin{cases} x(a), & \text{if } x \in C(K); \\ \delta_a(x), & \text{if } x \in K. \end{cases}$$

Let $D(K) = \{g_a^0 : a \in K\} \cup \{g_a^1 : a \in K\}$. Notice that $D(K)$ is a first countable Rosenthal compactum. It is not separable if K is uncountable. The interesting case will be when $K = 2^{\mathbb{N}}$.

(5) *Extended Alexandroff Duplicate of the split Cantor*. For each finite binary sequence $t \in 2^{<\mathbb{N}}$ let $a_t \in 2^{\mathbb{N}}$ be the sequence starting with t and ending with 0's and let $b_t \in 2^{\mathbb{N}}$ be the sequence starting with t and ending with 1's. Define $h_t : 2^{\mathbb{N}} \to \mathbb{R}$ by

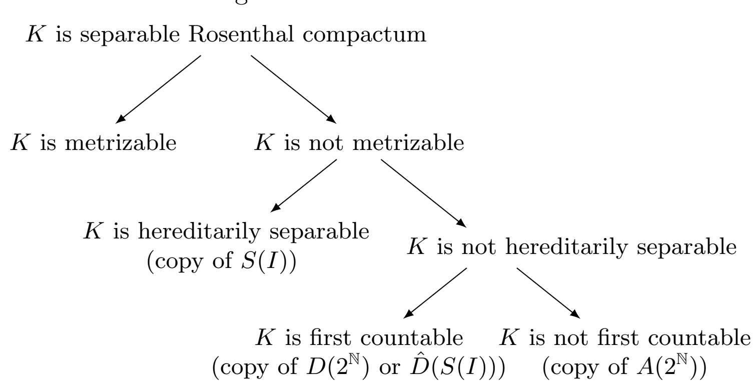
$$h_t(x) = \begin{cases} 0, & \text{if } x < a_t; \\ 1/2, & \text{if } a_t \leq x \leq b_t; \\ 1, & \text{if } b_t < x. \end{cases}$$

Let $\hat{D}(S(2^{\mathbb{N}}))$ be the pointwise closure of the set $\{h_t : t \in 2^{<\mathbb{N}}\}$. Hence, $\hat{D}(S(2^{\mathbb{N}}))$ is a separable first countable Rosenthal compactum which is not hereditarily separable. In fact, it contains an uncountable discrete subspace (see Theorem 5 in [Tod99]).

Theorem 3.4 (Todorčević's Trichotomy, [Tod99], Theorem 3 in [ADK08])**.** *Let K be a separable Rosenthal Compactum.*

(i) If K is hereditarily separable but non-metrizable, then $S(2^{\mathbb{N}})$ embeds into K.
(ii) If K is first countable but not hereditarily separable, then either $D(2^{\mathbb{N}})$ or $\hat{D}(S(2^{\mathbb{N}}))$ embeds into K.
(iii) If K is not first countable, then $A(2^{\mathbb{N}})$ embeds into K.

We thus have the following classification:



Todorčević's Trichotomy suggests that in order to distinguish the classes NIP_i, the examples in 3.3 are essential. The following examples show that the levels NIP_i $(i = 1, 2, 3)$ may be distinguished by the topological complexity of deep computations.

Examples 3.5.

(1) Let $(L, \mathcal{P}, \Gamma)$ be the computation of square root (example 2.6 with $\Delta = \Gamma$. We saw that $\overline{\tilde{\Delta}} = \tilde{\Delta} \cup \{\tilde{f}\} \subseteq B_1(\mathcal{L}, \mathcal{L})$. This corresponds to the Alexandroff compactification of a countable discrete set, which is metrizable. Hence, Δ is NIP_3 but it is not stable, in the sense that $\overline{\tilde{\Delta}} \not\subseteq C(\mathcal{L}, \mathcal{L})$.

(2) Let $(L, \mathcal{P}, \Gamma)$ be the finite precision threshold classifiers (Section 2.4) with $\Delta = \{\phi_w : w \in 2^{<\mathbb{N}}\} \cup \{\mathbf{0}^\infty, \mathbf{1}^\infty\}$. We saw that $\overline{\Delta}$ is homeomorphic to the Split Cantor space $S(2^{\mathbb{N}})$ (Example 3.3(3)), which is hereditarily separable but not metrizable. Hence, Δ is NIP_2 but not NIP_3.

(3) Let $(L, \mathcal{P}, \Gamma)$ be the finite precision prefix test (Section 2.5) with $\Delta = \{\psi_w : w \in 2^{<\mathbb{N}}\}$. We saw that $\overline{\tilde{\Delta}}$ is homeomorphic to the Extended Alexandroff compactification $\hat{A}(2^{\mathbb{N}})$ (Example 3.3(3)), which is separable but not first countable. Hence, Δ is NIP but not NIP_1.

The definitions provided here for NIP_i $(i = 1, 2, 3)$ are topological. This raises the following question:

Question 3.6. Is there a non-topological characterization for NIP_i, $i = 1, 2, 3$?

3.3. **The Argyros-Dodos-Kanellopoulos heptachotomy, and approximability of deep computation by minimal classes.** In the three separable cases given in 3.3, namely, $\hat{A}(2^{\mathbb{N}})$, $S(2^{\mathbb{N}})$ and $\hat{D}(S(2^{\mathbb{N}}))$, the countable dense subsets are indexed by the binary tree $2^{<\mathbb{N}}$. This choice of index is useful for two reasons:

(1) Our emphasis is computational. Elements of $2^{<\mathbb{N}}$ represent finite bitstrings, i.e., standard computations, while Rosenthal compacta represent deep computations, i.e., limits of finite computations. Mathematically, deep computations are pointwise limits of standard computations. However, computationally, we are interested in the manner (and the efficiency) in which the approximations can occur.

(2) Infinite branches of the binary tree $2^{<\mathbb{N}}$ correspond to the Cantor space $2^{\mathbb{N}}$, the canonical perfect set (in the sense that any Polish space with no isolated points contains a copy of $2^{\mathbb{N}}$). The use of infinite dimensional Ramsey theory for trees (pioneered by the work of James D. Halpern, Hans Läuchli in [HL66] and also Keith Milliken in [Mil81], and Alain Louveau, Saharon Shelah, Boban Velickovic in [LSV93]) and perfect sets (Fred Galvin and Andreas Blass in [Bla81], Arnold W. Miller in [Mil89], and Stevo Todorčević in [Tod99]) allowed S.A. Argyros, P. Dodos and V. Kanellopoulos in [ADK08] to obtain an improved version of Theorem 3.4. It is no surprise that Ramsey Theory becomes relevant in the study of Rosenthal compacta as it was a key ingredient in Rosenthal's ℓ_1 Theorem. For this reason, the main results in [ADK08] (which we cite in this paper) are better explained by indexing Rosenthal compacta with the binary tree.

Definition 3.7. Let X be a Polish space.

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

One of the main results in [ADK08] is that, up to equivalence, there are seven minimal families of Rosenthal compacta and that for every relatively compact $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$ there is a dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ such that $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$ is equivalent to one of the minimal families. We shall describe the seven minimal families next. We follow the same notation as in [ADK08]. For any node $t \in 2^{<\mathbb{N}}$, let us denote by $t^\frown 0^\infty$ ($t^\frown 1^\infty$) the infinite binary sequence starting with t and continuing will all 0's (respectively, all 1's). Fix a regular dyadic subtree $R = \{s_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$ (i.e., a dyadic subtree such that every level of R is contained in a level of $2^{<\mathbb{N}}$) with the property that for all $s, s' \in R$, $s^\frown 0^\infty \neq s'^\frown 0^\infty$ and $s^\frown 1^\infty \neq s'^\frown 1^\infty$. Given $t \in 2^{<\mathbb{N}}$, let v_t be the characteristic function of the set $\{x \in 2^{\mathbb{N}} : x \text{ extends } t\}$. Let $<$ be the lexicographic order in $2^{\mathbb{N}}$. Given $a \in 2^{\mathbb{N}}$, let $f_a^+ : 2^{\mathbb{N}} \to \{0,1\}$ be the characteristic function of $\{x \in 2^{\mathbb{N}} : a \leq x\}$ and let $f_a^- : 2^{\mathbb{N}} \to \{0,1\}$ be the characteristic function of $\{x \in 2^{\mathbb{N}} : a < x\}$. Given two maps $f, g : 2^{\mathbb{N}} \to \mathbb{R}$ we denote by $(f,g) : 2^{\mathbb{N}} \sqcup 2^{\mathbb{N}} \to \mathbb{R}$ the function which is f on the first copy of $2^{\mathbb{N}}$ and g on the second copy of $2^{\mathbb{N}}$.

(1) $D_1 = \{\frac{1}{|t|+1} v_t : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_1} = A(2^{\mathbb{N}})$.
(2) $D_2 = \{s_t^\frown 0^\infty : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_2} = 2^{\leq N}$.
(3) $D_3 = \{f^+_{s_t^\frown 0^\infty} : t \in 2^{<\mathbb{N}}\}$. This is a discrete in $\overline{D_3} = S(2^{\mathbb{N}})$.
(4) $D_4 = \{f^-_{s_t^\frown 1^\infty} : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_4} = S(2^{\mathbb{N}})$.
(5) $D_5 = \{v_t : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_5} = \hat{A}(2^{\mathbb{N}})$.
(6) $D_6 = \{(v_{s_t}, s_t^\frown 0^\infty) : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_6} = \hat{D}(2^{\mathbb{N}})$.
(7) $D_7 = \{(v_{s_t}, x^+_{s_t^\frown 0^\infty}) : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_7} = \hat{D}(S(2^{\mathbb{N}}))$

Theorem 3.8 (Heptachotomy of minimal families, Theorem 2 in [ADK08])**.** *Let X be Polish. For every relatively compact $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$, there exists $i = 1, 2, \ldots, 7$ and a regular dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$ such that $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$ is equivalent to D_i. Moreover, all D_i are minimal and mutually non-equivalent.*

The implications of this result for deep computations is the following: for any countable set of computations Δ satisfying the NIP (for some CCS $(L, \mathcal{P}, \Gamma)$), we can always find a countable discrete set of deep computations that approximate all the other deep computations. For example: in the finite precision prefix test example (subsection 2.5), the prefix test computations (family D_5) approximate all other deep computations. However, note that this discrete set D_i may not consist of continuous functions (i.e., they will not be computable in general). For example, functions in D_3 are not continuous.

4. Randomized versions of NIP and Monte Carlo computability of deep computations

In this section, we replace deterministic computability by probabilistic ('Monte Carlo') computability. We do not assume that $\mathcal{P}$ is countable. The main results of the section are Theorem 4.8 (connecting NIP and Monte Carlo computability) and 4.14 (connecting Talagrand stability and Monte Carlo computability).

Fundamental in this section is a measure-theoretic version of Theorem 1.11, namely, Theorem 4.5. For the proof of Theorem 1.11, we assumed countability of $\mathcal{P}$ — this ensured that $\mathbb{R}^{\mathcal{P}}$ a Polish space. In this section, the countability assumption is not needed.

4.1. **NIP and Monte Carlo computability of deep computations.** The *raison d'être* of the Baire class-1 functions is to work with a class of functions that are obtained as accumulation points of continuous functions (deep computations) but are not too far from being continuous. By Fact 1.2, for perfectly normal X, a function f is in $B_1(X, Y)$ if and only if $f^{-1}[U]$ is an F_σ subset of X for every open $U \subseteq Y$. Recall that a function $f : X \to Y$ is continuous if and only if $f^{-1}[U]$ is open in X for every open $U \subseteq Y$. In other words, we can view the class $B_1(X, Y)$ as a weaker notion of continuity. In this section we will study a different (and larger) class of functions: the space of universally measurable functions.

Definition 4.1. Given a Hausdorff space X and a measurable space (Y, Σ), we say that $f : X \to Y$ 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 every Radon measure μ on X is finite. Then, the measure given by $\nu(A) := \mu(A)/\mu(X)$ is a probability measure on X with the same null sets as μ. Hence, Radon measures on compact spaces are equivalent to (Radon) probability measures. We summarize this fact in the next remark:

Remark 4.2. If X is compact, then a function $f : X \to \mathbb{R}$ 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 $U \subseteq \mathbb{R}$.

Intuitively, a function is universally measurable if it is "measurable no matter which reasonable way you try to measure things on its domain". The concept of universal measurability emerged from work of Kallianpur and Sazonov, in the late 1950's and 1960s — with later developments by Blackwell, Darst and others — building on earlier ideas of Gnedenko and Kolmogorov from the 1950s. See [Pap02, Chapters 1 and 2].

Notation 4.3. Following [BFT78], the collection of all universally measurable real-valued functions on X will be denoted by $M_r(X)$. Given a fixed Radon measure μ on X, the collection of all μ-measurable real-valued functions on X will be denoted by $\mathscr{M}^0(X, \mu)$.

In the context of deep computations, we are interested in transition maps of a state space $L \subseteq \mathbb{R}^{\mathcal{P}}$ into itself. There are two natural σ-algebras one can consider in the product space $\mathbb{R}^{\mathcal{P}}$: the Borel σ-algebra, i.e., the σ-algebra generated by open sets in $\mathbb{R}^{\mathcal{P}}$, and the cylinder σ-algebra (i.e., the σ-algebra generated by the sub-basic open sets in $\mathbb{R}^{\mathcal{P}}$, that is, the sets $\pi_P^{-1}(U)$ with $U \subseteq \mathbb{R}$ open and $P \in$

$\mathcal{P}$). Note that when $\mathcal{P}$ is countable, both σ-algebras coincide, but in general the cylinder σ-algebra is strictly smaller. We will use the cylinder σ-algebra to define universally measurable maps $f : \mathbb{R}^{\mathcal{P}} \to \mathbb{R}^{\mathcal{P}}$. The reason for this choice is the following characterization:

Proposition 4.4. *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). The following are equivalent for $f : X \to Y$.:*

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

Proof. (i)$\Rightarrow$(ii) is clear since the projection maps π_i are measurable and the composition of measurable functions is measurable. To prove (ii)$\Rightarrow$(i), suppose that $C = \prod_{i\in I} C_i$ is a measurable cylinder and let J be the finite subset of I such that $C_i \neq Y_i$ for $i \in J$. Then, $C = \bigcap_{i\in J} \pi_i^{-1}(C_i)$, so $f^{-1}(C) = \bigcap_{i\in J}(\pi_i \circ f)^{-1}(C_i)$ is universally measurable by assumption. $\square$

The preceding proposition says that a transition map is universally measurable if and only if it is universally measurable on all its features; in other words, we can check measurability of a transition just by checking measurability feature by feature. This is the same as in the Baire class-1 case (compare with Proposition 1.10).

The main result in section 3 is that, as long as we work with countably many features, PAC-learning (or NIP) corresponds to relative compactness in the space of Baire class-1 functions. The following result (which does not assume countability of the number of features) gives an analogous characterization of the NIP in terms of universal measurability:

Theorem 4.5 (Bourgain-Fremlin-Talagrand, Theorem 2F in [BFT78])**.** *Let X be a Hausdorff space and $A \subseteq C(X)$ be pointwise bounded. The following are equivalent:*

(i) $\overline{A} \subseteq M_r(X)$.
(ii) For every compact $K \subseteq X$, $A|_K$ satisfies the NIP.
(iii) For every Radon measure μ on X, A is relatively countably compact in $\mathscr{M}^0(X, \mu)$, i.e., every countable subset of A has an accumulation point in $\mathscr{M}^0(X, \mu)$.

This result allows us to formalize the concept of a deep computation being *Monte Carlo computable*, which we define below (Definition 4.6). To motivate the definition, let us first recall two facts:

(1) Littlewoood's second principle states that every Lebesgue measurable function is "nearly continuous". The formal statement of this, which is Luzin's theorem, is that if (X, Σ, μ) a Radon measure space and Y is a second-countable topological space (e.g., $Y = \mathbb{R}^{\mathcal{P}}$ with $\mathcal{P}$ countable) equipped with the Borel σ-algebra, then any given $f : X \to Y$ is measurable if and only if for every $E \in \Sigma$ and every $\varepsilon > 0$ there exists a closed $F \subseteq E$ such that the restriction $f|_F$ is continuous and $\mu(E\backslash F) < \varepsilon$.
(2) Computability of deep computations is characterized in terms of continuous extendibility of computations. This is at the core of [ADIW24].

These two facts motivate the following definition:

Definition 4.6. Let $(L, \mathcal{P}, \Gamma)$ be a CCS. We say that a transition $f : L \to L$ is *universally Monte Carlo computable* if and only if there exists $\tilde{f} : \mathcal{L}_{\mathrm{sh}} \to \mathcal{L}_{\mathrm{sh}}$ extending f such that for every sizer $r_\bullet$ there is a sizer $s_\bullet$ such that the restriction

$\tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \to \mathcal{L}[s_\bullet]$ is universally measurable, i.e., $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \to [-s_P, s_P]$ is μ-measurable for every Radon probability measure μ on $\mathcal{L}[r_\bullet]$ and $P \in \mathcal{P}$.

Remark 4.7. Condition (2) of Theorem 4.5 shows that to study measure-theoretic versions of NIP, we need only consider compact subsets of X. Now, every Radon measure on a compact space is finite; hence, by proper normalization, it can be treated as a probability measure. Therefore, in the context of Monte Carlo measurability, we focus on Radon probability measures rather than general Radon measures.

Theorem 4.8. *Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom. Let R be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be R-confined. If $\pi_P \circ \Delta|_{L[r_\bullet]}$ satisfies the NIP for all $P \in \mathcal{P}$ and all $r_\bullet \in R$, then every deep computation in Δ is universally Monte Carlo computable.*

Proof. Fix $P \in \mathcal{P}$ and $r_\bullet \in R$. By the Extendibility Axiom, $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is a set of pointwise bounded continuous functions on the compact set $\mathcal{L}[r_\bullet]$. Since $\pi_P \circ \tilde{\Delta}|_{L[r_\bullet]} = \pi_P \circ \Delta|_{L[r_\bullet]}$ has the NIP, so does $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ by Lemma 1.12. By Theorem 4.5, we have $\overline{\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq M_r(\mathcal{L}[r_\bullet])$ for all $r_\bullet \in R$ and $P \in \mathcal{P}$. Let $f \in \overline{\Delta}$ be a deep computation. Write $f = \mathcal{U}\lim_i \gamma_i$ as an ultralimit of computations in Δ. Define $\tilde{f} := \mathcal{U}\lim_i \tilde{\gamma}_i$. Then, $\tilde{f} : \mathcal{L}_{\text{sh}} \to \mathcal{L}_{\text{sh}}$ extends f. Since Δ is R-confined we have that $f : L[r_\bullet] \to L[r_\bullet]$ and $\tilde{f} : \mathcal{L}[r_\bullet] \to \mathcal{L}[r_\bullet]$ for all $r_\bullet \in R$. Lastly, note that for all $r_\bullet \in R$ and $P \in \mathcal{P}$ we have that $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_\bullet]} \in \overline{\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq M_r(\mathcal{L}[r_\bullet])$. $\square$

Question 4.9. Under the same assumptions of the preceding theorem, suppose that every deep computation of Δ is universally Monte Carlo computable. Must $\pi_P \circ \Delta|_{L[r_\bullet]}$ have the NIP for all $P \in \mathcal{P}$ and all $r_\bullet \in R$?

4.2. **Talagrand stability and Monte Carlo computability of deep computations.** There is another notion closely related to NIP, introduced by Talagrand in [Tal84] while studying Pettis integration. Suppose that X is a compact Hausdorff space and $A \subseteq \mathbb{R}^X$. Let μ be a Radon probability measure on X. Given a μ-measurable set $E \subseteq X$, a positive integer k and real numbers $a < b$. we write:

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

We say that A is *Talagrand μ-stable* if and only if for every μ-measurable set $E \subseteq X$ of positive measure and for every $a < b$ there is a $k \ge 1$ such that

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

where μ^* denotes the outer measure (we need to work with outer measure since the sets $D_k(A, E, a, b)$ need not be μ^{2k}-measurable). The inequality certainly holds when A is a countable set of continuous (or μ-measurable) functions.

The main result of this section is that deep computations of a Talagrand stable set of computations are Monte Carlo computable; this is Theorem 4.14 below. We now prove that accumulation points (i.e., deep computations) of a Talagrand μ-stable set are μ-measurable. But first, let us state the following useful characterization of measurable functions (compare with Fact 1.2):

Fact 4.10 (Lemma 413G in [Fre03])**.** *Suppose that (X, Σ, μ) is a measure space and $\mathcal{K} \subseteq \Sigma$ is a collection of measurable sets satisfying the following conditions:*

(1) (X, Σ, μ) *is complete, i.e., for all* $E \in \Sigma$ *with* $\mu(E) = 0$ *and* $F \subseteq E$ *we have* $F \in \Sigma$.

(2) (X, Σ, μ) *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* $0 < \mu(F) < \infty$.

(3) (X, Σ, μ) *is saturated, i.e.,* $E \in \Sigma$ *if and only if* $E \cap F \in \Sigma$ *for all* $F \in \Sigma$ *with* $\mu(F) < \infty$.

(4) (X, Σ, μ) *is inner regular with respect to* $\mathcal{K}$*, i.e., for all* $E \in \Sigma$

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

(In particular, if X *is compact Hausdorff,* μ *is a Radon probability measure on* X *and* $\mathcal{K}$ *is the collection of compact subsets of* X*, all these conditions hold). Then,* $f : X \to \mathbb{R}$ *is measurable if and only if for every* $K \in \mathcal{K}$ *with* $0 < \mu(K) < \infty$ *and* $a < b$*, either* $\mu^*(P) < \mu(K)$ *or* $\mu^*(Q) < \mu(K)$ *where* $P = \{x \in K : f(x) \leq a\}$ *and* $Q = \{x \in K : f(x) \geq b\}$.

The following technical lemma will be instrumental for proving Proposition 4.13, which, in turn, will yield the main result of the subsection, namely Theorem 4.14.

Lemma 4.11. *If* A *is Talagrand* μ*-stable, then* $\overline{A}$ *is also Talagrand* μ*-stable and* $\overline{A} \subseteq \mathscr{M}^0(X, \mu)$.

Proof. First, we claim that a subset of a μ-stable set is μ-stable. To see this, suppose that $A \subseteq B$ and B is μ-stable. Fix any μ-measurable $E \subseteq X$ of positive measure and $a < b$. Let $k \geq 1$ such that

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

Since $A \subseteq B$, we have $D_k(A, E, a, b) \subseteq D_k(B, E, a, b)$; therefore,

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

We now show that $\overline{A}$ is μ-stable. Fix $E \subseteq X$ measurable with positive measure and $a < b$. Let $a' < b'$ be such that $a < a' < b' < b$. Since A is μ-stable, let $k \geq 1$ be such that

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

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 definition of pointwise convergence topology, there exists $g \in A$ such that $g(x_{2i}) < a'$ and $g(x_{2i+1}) > b'$ for all $i < k$. Hence, $x \in D_k(A, E, a', b')$. We have shown that $D_k(\overline{A}, E, a, b) \subseteq D_k(A, E, a', b')$; hence,

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

It suffices to show that $\overline{A} \subseteq \mathscr{M}^0(X, \mu)$. Suppose that there exists $f \in \overline{A}$ such that $f \notin \mathscr{M}^0(X, \mu)$. By fact 4.10, there exists a μ-measurable set E of positive measure and $a < b$ such that $\mu^*(P) = \mu^*(Q) = \mu(E)$ where $P = \{x \in E : f(x) \leq a\}$ and $Q = \{x \in E : f(x) \geq b\}$. Then, for any $k \geq 1$: $(P \times Q)^k \subseteq D_k(\{f\}, E, a, b)$, so $(\mu^{2k})^*(D_k(\{f\}, E, a, b)) = (\mu^*(P)\mu^*(Q))^k = (\mu(E))^{2k}$. Thus, $\{f\}$ is not μ-stable. However, we argued above that a subset of a μ-stable set must be μ-stable, so we have a contradiction. □

Definition 4.12. We say that A is *universally Talagrand stable* if A is Talagrand μ-stable for every Radon probability measure μ on X.

We first observe that universal Talagrand stability corresponds to a complexity class smaller than or equal to the NIP class:

Proposition 4.13. *Let X be a compact Hausdorff space and $A \subseteq C(X)$ be pointwise bounded. If A is universally Talagrand stable, then A satisfies the NIP.*

Proof. By Theorem 4.5, it suffices to show that A is relatively countably compact in $\mathscr{M}^0(X, \mu)$ for every Radon probability measure μ on X. Since A is Talagrand μ-stable for any such μ, we have $\overline{A} \subseteq \mathscr{M}^0(X, \mu)$ by Lemma 4.11. In particular, A is relatively countably compact in $\mathscr{M}^0(X, \mu)$. □

Corollary 4.14. *Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom. If $\pi_P \circ \Delta|_{L[r_\bullet]}$ is universally Talagrand stable for all $P \in \mathcal{P}$ and all sizers $r_\bullet$, then every deep computation is universally Monte Carlo computable.*

Proof. This is a direct consequence of Proposition 4.13 and Theorem 4.8. □

In the context of deep computations, we have identified two ways to obtain Monte Carlo computability, namely, NIP/PAC and Talagrand stability. It is natural to ask whether these two notions are equivalent. The following results show that, even in the simple case of countably many computations, this question is sensitive to the set-theoretic axioms. On the one hand, it is consistent (with respect to the standard ZFC axioms of set theory) that these two classes are the same:

Theorem 4.15 (Talagrand, Theorem 9-3-1(a) in [Tal84])**.** *Let X be a compact Hausdorff space and $A \subseteq M_r(X)$ be countable and pointwise bounded. Assume that $[0, 1]$ is not the union of $< \mathfrak{c}$ closed measure zero sets. If A satisfies the NIP, then A is universally Talagrand stable.*

(The assumption that $[0, 1]$ is not the union of $< \mathfrak{c}$ closed measure zero sets is a consequence of, for example, the Continuum Hypothesis.)

On the other hand, by fixing a particular well-known measure, namely the Lebesgue measure, we see that the other case is also consistent:

Theorem 4.16 (Fremlin, Shelah, [FS93])**.** *It is consistent with the usual axioms of set theory that there exists a countable pointwise bounded set of Lebesgue measurable functions with the NIP which is not Talagrand stable with respect to Lebesgue measure.*

Notice that the preceding two results apply to sets of measurable functions, a class of functions larger than the class of continuous functions. However, by the Extendibility Axiom, finitary computations are continuous, i.e., $A \subseteq C_p(X)$ (instead of having $A \subseteq M_r(X)$ as in Theorem 4.15). The question of whether we can remove the set-theoretic assumption in Theorem 4.15 when $A \subseteq C_p(X)$ remains open.

References

[ADIW24] Samson Alva, Eduardo Dueñez, Jose Iovino, and Claire Walton. Approximability of deep equilibria. *arXiv preprint arXiv:2409.06064*, 2024. Last revised 20 May 2025, version 3.

[ADK08] Spiros A. Argyros, Pandelis Dodos, and Vassilis Kanellopoulos. A classification of separable Rosenthal compacta and its applications. *Dissertationes Mathematicae*, 449:1–55, 2008.

[Ark92] A. V. Arkhangel'skii. *Topological function spaces*, volume 78 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1992. Translated from the Russian by R. A. M. Hoksbergen.

[BFT78] J. Bourgain, D. H. Fremlin, and M. Talagrand. Pointwise compact sets of Baire-measurable functions. *American Journal of Mathematics*, 100(4):845–886, 1978.

[BKK] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. Preprint. https://arxiv.org/abs/1909.01377, https://implicit-layers-tutorial.org/.

[Bla81] Andreas Blass. A partition theorem for perfect sets. *Proceedings of the American Mathematical Society*, 82(2):271–277, 1981.

[Bla84] Paul Blanchard. Complex analytic dynamics on the Riemann sphere. *Bulletin of the American Mathematical Society*, 11(1):85–141, 1984.

[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. In *Recent Progress in General Topology III*, pages 205–227. Springer, 2013.

[Eng89] Ryszard Engelking. *General topology*, volume 6 of *Sigma Series in Pure Mathematics*. Heldermann Verlag, Berlin, second edition, 1989. Translated from the Polish by the author.

[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(2):435–455, 1993.

[GM22] Eli Glasner and Michael Megrelishvili. Todorčević' trichotomy and a hierarchy in the class of tame dynamical systems. *Transactions of the American Mathematical Society*, 375(7):4513–4548, 2022.

[God80] Gilles Godefroy. Compacts de Rosenthal. *Pacific J. Math.*, 91(2):293–306, 1980.

[Gro52] A. Grothendieck. Critères de compacité dans les espaces fonctionnels généraux. *Amer. J. Math.*, 74:168–186, 1952.

[HL66] J. D. Halpern and H. Läuchli. A partition theorem. *Transactions of the American Mathematical Society*, 124:360–367, 1966.

[HT19] Haim Horowitz and Stevo Todorcevic. Compact sets of baire class one functions and maximal almost disjoint families, 2019.

[Iov94] José N. Iovino. *Stable Theories in Functional Analysis*. PhD thesis, University of Illinois at Urbana-Champaign, 1994.

[Kei03] H. Jerome Keisler. Model theory for real-valued structures. In José Iovino, editor, *Beyond First Order Model Theory, Volume II*. CRC Press, Boca Raton, FL, 2003.

[LSV93] Alain Louveau, Saharon Shelah, and Boban Veličković. Borel partitions of infinite subtrees of a perfect tree. *Annals of Pure and Applied Logic*, 63(3):271–281, 1993.

[Mil81] Keith Milliken. A partition theorem for the infinite subtrees of a tree. *Transactions of the American Mathematical Society*, 263:137–148, 1981.

[Mil89] Arnold W. Miller. Infinite combinatorics and definability. *Annals of Pure and Applied Logic*, 41(2):179–203, 1989.

[Pap02] E. Pap, editor. *Handbook of measure theory. Vol. I, II*. North-Holland, Amsterdam, 2002.

[Ros74] Haskell P. Rosenthal. A characterization of Banach spaces containing l^1. *Proc. Nat. Acad. Sci. U.S.A.*, 71:2411–2413, 1974.

[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.*, 378:686–707, 2019.

[Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972.

[She71] Saharon Shelah. Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory. *Ann. Math. Logic*, 3(3):271–362, 1971.

[She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[She90] Saharon Shelah. *Classification theory and the number of nonisomorphic models*, volume 92 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, second edition, 1990.

[Tal84] Michel Talagrand. *Pettis Integral and Measure Theory*, volume 51 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence, RI, USA, 1984. Includes bibliography (pp. 220–224) and index.

[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*, volume 1652 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 1997.

[Tod99] Stevo Todorčević. Compact subsets of the first Baire class. *Journal of the American Mathematical Society*, 12(4):1179–1212, 1999.

[Tod23] Stevo Todorčević. Dense metrizability. *Annals of Pure and Applied Logic*, 175:103327, 07 2023.

[Val84] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[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(2):264–280, 1971.

[VC74] Vladimir N Vapnik and Alexey Ya Chervonenkis. *Theory of Pattern Recognition*. Nauka, Moscow, 1974. German Translation: Theorie der Zeichenerkennung, Akademie-Verlag, Berlin, 1979.