

COMPLEXITY OF DEEP COMPUTATIONS VIA TOPOLOGY OF FUNCTION SPACES

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

¹Department of Mathematics, University of Texas at San Antonio

²Department of Mathematics, University of Toronto

ABSTRACT. We use topological methods and a suitable framework to study the complexity of limit computations (such as so-called “deep computations” implied by recurrent neural networks). By using topology of function spaces—specifically, the classification of Rosenthal compacta—we identify new computational complexity classes. The language of model theory, particularly, the concept of *independence* from Shelah’s classification theory, provides a medium to translate between topology and computation. The theory of Rosenthal compacta allows us to characterize approximability of deep computations, both deterministically and probabilistically.

INTRODUCTION

In this paper we study asymptotic behavior of computations as parameters of the computation tend towards a limit, 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.

The main result of the paper is the identification of a hierarchy of four distinct complexity classes for limit computations. This is obtained by invoking a major result about topology of spaces of functions, namely, Todorčević’s trichotomy for Rosenthal compacta [Tod99] and transferring it into the realm of computation. The medium that we use to translate between topology and computation is model theory, specifically, Shelah’s classification theory [She90].

The translation between topology and computation proceeds as follows. To every 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 association 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 borrowed from model

Date: March 1, 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.

theory, where types spaces play a central role. The embedding of computations into spaces of types then 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 model-theoretic viewpoint of spaces of types 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], the authors introduced the concept of limits of computations, which they called *ultracomputations* (given that 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 primarily the term “deep computation”.

In [ADIW24], a new “tame vs wild” (i.e., polynomially approximable vs non-polynomially approximable) dichotomy for the complexity of deep computations was proved 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 the sense of model theory.

In this paper, we refine the dichotomy of [ADIW24], and obtain new complexity classes for deep computation. The “wild” part of the dichotomy from [ADIW24] is further subdivided between “tame” and “wild” classes; three new levels of tameness are identified within this new subdivision, and seven distinct minimal prototypes of tameness are detected.

The finer classification results of this paper are obtained through more contemporary topological and model-theoretic tools than in the previous paper [ADIW24]. From the topological side, instead of Grothendieck’s classical criterion, we use the modern theory of *Rosenthal compacta*, and from the model theory side, instead of stability, we use *neostability*. The main tool imported from neostability is Shelah’s *Independence Property*, or rather, its negation, which is commonly referred to as “NIP”. More on this below.

Deep computations arise as topological limits of standard (continuous) computations. 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 above the Baire class 0, which (by definition) consists of continuous functions.

We apply a result of Bourgain, Fremlin and Talagrand from the late 1970s on compact subsets of the first Baire class to obtain a new “tame vs wild” dichotomy for deep computations [BFT78]; then, we apply a trichotomy of Todorčević, from the late 90s, for *Rosenthal compacta* to identify new tame classes [Tod99].

A Rosenthal compactum is a compact topological space that is homeomorphic to a subspace of the space of Baire class 1 functions on some Polish (i.e., separable, completely metrizable) space equipped with 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 for Rosenthal compacta 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, there are no elementary proofs of the trichotomy; in fact, Todorčević’s methods have been analyzed and exploited further [Tod23, HT19].

Through our Rosetta stone, Rosenthal compacta in topology correspond to the important concept of *No-Independence Property* (known as “NIP”) in model theory, identified by Shelah [She71, She90], and to the concepts *VC-dimension* and *Probably Approximately Correct learning* (known as “PAC learnability”) in statistical learning theory [VC71, VC74, Val84].

Beyond Todorčević’s trichotomy, we invoke a more recent heptachotomy for separable Rosenthal compacta obtained by Argyros, Dodos and Kanellopoulos [ADK08]. These authors 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.

From the point of view of approximation theory, it is noteworthy that in the context of Grothendieck [Gro52] and Bourgain-Fremlin-Talagrand [BFT78], compactness and sequential compactness coincide; thus, the results of these papers allow us to ensure that deep computations, which in principle are abstract ultrafilter limits (over a possibly uncountable set), can always be approximated by concrete sequences of standard computations.

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 that the reader has 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. 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 extend in future research.

CONTENTS

1. Topological preliminaries:	
From continuity to Baire class 1	4
1.1. From Rosenthal's dichotomy to the Bourgain-Fremlin-Talagrand dichotomy to Shelah's NIP	6
1.2. NIP as a universal dividing line between polynomial and exponential complexity	8
1.3. Rosenthal compacta	9
1.4. The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with \mathcal{P} countable	9
Historical remarks	11
2. Compositional Computation Structures:	
A structural approach to arbitrary-precision computations	11
2.1. Computation States Structures	11
2.2. Compositional Computation Structures	13
2.3. Computability of deep computations and the Extendibility Axiom	14
2.4. Newton's method as a CCS	15
2.5. Finite-precision threshold classifiers	18
2.6. Finite precision prefix test	20
3. Classifying deep computations	21
3.1. NIP, Rosenthal compacta, and deep computations	21
3.2. The Todorčević trichotomy, and levels of NIP and PAC learnability	22
3.3. The Argyros-Dodos-Kanellopoulos heptachotomy, and approximability of deep computation by minimal classes	25
4. Randomized versions of NIP and Monte Carlo computability of deep computations	26
4.1. NIP and Monte Carlo computability of deep computations	26
4.2. Talagrand stability and Monte Carlo computability of deep computations	28
5. Conclusion	31
References	31

1. TOPOLOGICAL PRELIMINARIES: FROM CONTINUITY TO BAIRE CLASS 1

In this section we recall notions and results from general topology and function space theory. For simplicity, we assume that all spaces are Hausdorff unless explicitly stated otherwise.

A subset of a topological space is F_σ if it is a countable union of closed sets, and G_δ if it is a countable intersection of open sets. In a metric space, every open set is F_σ ; equivalently, every closed set is G_δ .

A *Polish space* is a separable and completely metrizable topological space, i.e., one admitting a complete metric that induces its topology. The class of Polish spaces is closed under countable topological products. Classical examples of Polish spaces (besides the real line) include 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.

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

If X topological space and A is a subset of X , the A is said to be *nowhere dense* if its closure has an empty interior, and of the *first category* if A is a countable union of closed nowhere dense sets; otherwise, A is said to be of the *second category*. Sets $A \subseteq X$ of the first category are also called *meager*, and their complements $X \setminus A$ are called *comeager*. Thus, the set of meager subsets of X is the σ -ideal generated by the nowhere dense subsets of X . A topological space X is *Baire* if every comeager subset of X is dense. Equivalently, X is Baire if any countable intersection of open dense sets remains dense. The Baire Category Theorem states that compact Hausdorff spaces and all completely metrizable spaces (hence, all Polish spaces) are Baire.¹

Given two topological spaces X and Y we denote by $C_p(X, Y)$ the set of all continuous functions $f : X \rightarrow Y$ endowed with the topology of pointwise convergence, i.e., with the relative (subspace) product topology $C_p(X, Y) \subseteq Y^X$ —where Y carries its topology, but X serves merely as a (discrete) index set stripped of any topology. The space $C_p(X, \mathbb{R})$ of continuous real-valued functions on X is denoted simply $C_p(X)$.

A function $f \in Y^X$ between topological spaces X, Y is said to be of *Baire class 1*, or a *Baire-1* function, if there is a sequence $(f_n)_{n \in \mathbb{N}} \subseteq C_p(X, Y)$ such that $f = \lim_{n \rightarrow \infty} f_n$ (i.e., $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ for all $x \in X$). An elementary argument shows that Baire-1 functions are continuous everywhere except on a set of the first category. If X and Y are topological spaces, the space of Baire-1 functions $f : X \rightarrow Y$ endowed with the topology of pointwise convergence is denoted $B_1(X, Y)$; thus, $C_p(X, Y) \subseteq B_1(X, Y) \subseteq Y^X$ are topological inclusions. We abbreviate $B_1(X, \mathbb{R})$ as $B_1(X)$. Relatively compact subsets of $C_p(X)$ and of $B_1(X)$ (particularly for X Polish) are of interest in analysis and topological dynamics.

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

The following fact was proved by Baire in his groundbreaking thesis. For a proof, see [Tod97, Section 10].

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

- (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 $U \subseteq \mathbb{R}$ 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

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

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

1.1. From Rosenthal's dichotomy to the Bourgain-Fremlin-Talagrand dichotomy to Shelah's NIP. In metrizable spaces, the closure of a subset consists of limits of convergent sequences therein. In general spaces, including many C_p -spaces, such a characterization of closures fails in remarkable ways.

The n -th coordinate map $p_n : 2^{\mathbb{N}} \rightarrow \{0, 1\}$ on the Cantor space $X = 2^{\mathbb{N}}$ ($= \{0, 1\}^{\mathbb{N}}$) is continuous for each $n \in \mathbb{N}$; however, $(p_n)_{n \in \mathbb{N}}$ has no convergent subsequences in \mathbb{R}^X ([Tod97, Chapter 1.1]). 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 $\beta\mathbb{N}$ of the (discrete) space of natural numbers. The space $\beta\mathbb{N}$ has cardinality 2^{2^ω} (largest among all separable topological spaces) and, effectively, is the set of ultrafilters on \mathbb{N} .²

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 space.

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

The genesis of the dichotomy 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 in 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 in the above manner, but rather as an equivalence (see Theorem 1.4 below).

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

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

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

$$\begin{aligned} X_{\leq a}^A &:= \bigcap_{f \in A} X_{\leq a}^f = \{x \in X : f(x) \leq a \text{ for all } f \in A\}, \\ X_{\geq a}^A &:= \bigcap_{f \in A} X_{\geq a}^f = \{x \in X : f(x) \geq a \text{ for all } f \in A\}. \end{aligned}$$

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

Theorem 1.4 (“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 $B_1(X)$.*

²Our earlier proof [ADIW24] of the existence of deep equilibria in compositional computation structures relies on a classical result on the existence of idempotent elements of $\beta\mathbb{N}$.

(ii) For every $(f_n)_{n \in \mathbb{N}} \subseteq A$ and every $a < b$ there is $I \subseteq \mathbb{N}$ such that

$$X_{\leq a}^{f_I} \cap X_{\geq b}^{f_I^c} = \emptyset.$$

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

The sets $X_{\leq a}^{f_I}$ and $X_{\geq b}^{f_I^c}$ appearing in condition Theorem 1.4(ii) are defined, respectively, in terms of $|I|$ -many inequalities of the form $f_i(x) \leq a$, and $|I^c|$ -many of the form $f_j(x) \geq b$. Thus, at least one of $X_{\leq a}^{f_I}$ and $X_{\geq b}^{f_I^c}$ 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 1.5. We say that a collection of functions $A \subseteq \mathbb{R}^X$ 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 $X_{\leq a}^{f_E} \cap X_{\geq b}^{f_F} = \emptyset$. We say that such E, F witness *finitary NIP* for (f_n) and a, b .

A set $A \subseteq \mathbb{R}^X$ has the *finitary Independence Property (IP)* if it does not have finitary NIP, 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 $X_{\leq a}^{f_E} \cap X_{\geq b}^{f_F} \neq \emptyset$.

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 $F = E^c$, we call the respective properties ‘‘BFT-NIP’’ and ‘‘BFT-IP’’. Thus, Theorem 1.4 becomes (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 1.6. 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 $A \subseteq C_p(X)$ has finitary IP. Fix a sequence $(f_n) \subseteq A$ and reals $a < b$ such that $X_{\leq a}^{f_E} \cap X_{\geq b}^{f_F} \neq \emptyset$ for every finite disjoint $E, F \subseteq \mathbb{N}$. Fix $I \subseteq \mathbb{N}$. Since $(f_n) \subseteq A \subseteq C_p(X)$ is a sequence of continuous functions, the collection $\mathcal{F}_I = \{X_{\leq a}^{f_n} : n \in I\} \cup \{X_{\geq b}^{f_m} : m \in I^c\}$ consists of closed (non-empty) sets. Since all finite $E \subseteq I$ and $F \subseteq I^c$ witness finitary IP for (f_n) and a, b , we see that the collection \mathcal{F}_I has the finite intersection property. By compactness, $X_{\leq a}^{f_I} \cap X_{\geq b}^{f_I^c} = \bigcap \mathcal{F}_I \neq \emptyset$. \square

Theorem 1.7. Let X be a compact metrizable (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 equivalence $(ii) \Leftrightarrow (iv)$ holds for arbitrary compact X . The equivalence $(ii) \Leftrightarrow (iii)$, for arbitrary X .)

Proof. Corollary of Theorem 1.4 and Proposition 1.6. \square

Theorem 1.7 may be stated as the following dichotomy (under the assumptions) for any compact metrizable X : 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 1.4) when A consists of $\{0, 1\}$ -valued (i.e., {Yes, 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^κ} -many distinct types.

In combinatorics: Sauer [Sau72] and Shelah [She72] proved the following independently: Let \mathcal{F} be a collection 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 $|\{S_i \cap A : i \in \mathbb{N}\}| = 2^n$ (\mathcal{F} is a collection having “exponential complexity”); or: there exists $N \in \mathbb{N}$ such that for every $A \subseteq S$ with $|A| \geq N$, one has

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

(\mathcal{F} is a collection having “polynomial complexity”). Clearly, any collection \mathcal{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

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

(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 \mathcal{F} ; it is a core concept in machine learning. If such an integer N does not exist, we say that the VC-dimension of \mathcal{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 1.4, as attested by the examples outlined in the preceding section, led to the following definition (by Gilles Godefroy [God80]):

Definition 1.8. 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.

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 (typically 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

$$\begin{aligned}\mathbb{R}^{X \times \mathcal{P}} &\rightarrow (\mathbb{R}^{\mathcal{P}})^X : \varphi \mapsto \hat{\varphi} \\ \mathbb{R}^{X \times \mathcal{P}} &\rightarrow (\mathbb{R}^X)^{\mathcal{P}} : \varphi \mapsto \varphi^*,\end{aligned}$$

where

$$\hat{\varphi}(x) := \varphi(x, \cdot) \in \mathbb{R}^{\mathcal{P}}, \quad \varphi^*(P) := \varphi(\cdot, P) \in \mathbb{R}^X.$$

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

$$\lambda_x : \mathbb{R}^{X \times \mathcal{P}} \rightarrow \mathbb{R}^{\mathcal{P}} : \varphi \mapsto \lambda_x(\varphi) := \hat{\varphi}(x);$$

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

$$\rho_P : \mathbb{R}^{X \times \mathcal{P}} \rightarrow \mathbb{R}^X : \varphi \mapsto \rho_P(\varphi) := \varphi^*(P).$$

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

$$\lambda_x : \mathbb{R}^X \rightarrow \mathbb{R} : f \mapsto f(x), \quad \rho_P : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R} : f \mapsto f(P).$$

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

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

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

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 $X \rightarrow \mathbb{R}^{\mathcal{P}}$ is precisely a \mathcal{P} -indexed family of continuous functions $X \rightarrow \mathbb{R}$, and these correspond to continuous functions $X \times \mathcal{P} \rightarrow \mathbb{R}$. 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. \square

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 1.4.

Theorem 1.10. *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 1.4 and 1.6. \square

Lastly, a simple but useful lemma to the effect that a set of continuous functions has NIP over a subset of the domain if and only if it has NIP over the closure of the subset.

Lemma 1.11. *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) $A|_{\bar{L}}$ satisfies the NIP.

Proof. The implication (i) \Rightarrow (ii) is trivial. To see the opposite implication, suppose that (ii) does not hold, and choose $\{f_n\}_{n \in \mathbb{N}} \subseteq A$ and $a < b$ such that for all finite disjoint $E, F \subseteq \mathbb{N}$,

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

But then, if a', b' are such that $a < a' < b' < b$, for any finite disjoint $E, F \subseteq \mathbb{N}$ we can choose

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

which, by the definition of closure, gives

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

contradicting (i). \square

Historical remarks. The Baire hierarchy of functions was introduced by René-Louis Baire in his 1899 doctoral thesis, *Sur les fonctions de variables réelles*. The field of C_p -theory was pioneered by A. V. Arhangel'skiĭ and his students in the 1970s and 1980s [Ark92], and it has found many applications in model theory and functional analysis. For a recent survey, see [Tka11].

2. COMPOSITIONAL COMPUTATION STRUCTURES: A STRUCTURAL APPROACH TO ARBITRARY-PRECISION COMPUTATIONS

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

2.1. Computation States Structures. 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*, also called *features*. For each $P \in \mathcal{P}$, we call the value $P(v)$ the P -th (*observable*) *feature* of v .

A *transition* of a computation states structure (L, \mathcal{P}) is a map $f : L \rightarrow L$.

The *type* of a state $v \in L$ is the indexed family³

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

of its features.

³As in §1.9 above, we regard each predicate on L as named by a specific formal label P (a *predicate symbol*) which is given the semantic meaning of a *bona fide* real function $P(\cdot) : v \mapsto P(v)$ (the predicate proper) in the computation states structure. Such formal symbol P becomes *de facto* the name of an observable feature having the real value $P(v)$ at any state $v \in L$. Whenever \mathcal{P} is used as an index set (e.g., as in the topological product $\mathbb{R}^{\mathcal{P}}$ below), it is regarded as a set of formal symbols. This syntactic vs. semantic abuse of nomenclature is commonplace in mathematics, but the model-theoretic framework requires contextual awareness of the distinction. To disambiguate when necessary, we shall write $P(\cdot)$, implying the semantic meaning of a real function on L , in contrast to P , meant as a purely syntactic symbol. Nevertheless, by a standard abuse of nomenclature in appropriate contexts, we may also treat P as a function on states (i.e., take P to mean $P(\cdot)$ proper).

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. (I.e., each symbol $P \in \mathcal{P}$ denotes a function $P(\cdot)$ taken as a computational primitive on L .)

We assume that each state $v \in L$ is uniquely characterized by its type $\text{tp}(v)$, so we may identify L with a topological subspace of $\mathbb{R}^{\mathcal{P}}$ and topologized as such (i.e., L is identified with its forward image $\text{tp}[L] \subseteq \mathbb{R}^{\mathcal{P}}$ under the type map). As a topological space, L is thus endowed with the topology of pointwise convergence induced by the product topology of $\mathbb{R}^{\mathcal{P}}$. In particular, for each $P \in \mathcal{P}$, the projection map $v \mapsto P(v)$ is continuous.

Important state spaces include $L = \mathbb{R}^{\mathbb{N}}$ (resp., $L = \mathbb{R}^n$ for some positive integer n), endowed with one predicate $P_i(v) = v_i$ (giving the i -th coordinate of v) for each $i \in \mathbb{N}$ (resp., for each $0 \leq i < n$). State subspaces $L \subseteq \mathbb{R}^{\mathbb{N}}$ and $L \subseteq \mathbb{R}^n$ are of primary interest in this paper. (Any states structure having a finite or countable family of predicates is effectively such a subspace.)

Definition 2.1. Given a computation states structure (L, \mathcal{P}) , any type $\text{tp}(v) \in \mathbb{R}^{\mathcal{P}}$ of a state $v \in L$ 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 (state) types* of (L, \mathcal{P}) , denoted \mathcal{L} ; elements $\mathbf{v} \in \mathcal{L}$ are called *state types*. Elements $\mathbf{v} \in \mathcal{L} \setminus L$ are *unrealized* state types.

Intuitively, state types capture a notion of “limit state”. For each $P \in \mathcal{P}$, the P -th coordinate map (natural projection) $\mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R} : \mathbf{v} = (\mathbf{v}_Q)_{Q \in \mathcal{P}} \mapsto \mathbf{v}_P$, gives a continuous extension to \mathcal{L} of the predicate $P(\cdot)$ on L ; abusing notation, the extension will still be denoted $P(\cdot)$ (or just P).

Remark 2.2. Just as L is identified with a subset of \mathcal{L} , but going in the opposite direction, one may regard unrealized state types as “hyper-states” (when such type is obtained as a sequential limit—or ultralimit—of realized state types, unrealized state types may be called “deep states”). A comment for readers versed in model theory: at least for the purpose of realizing state types (with parameters in the original states space L), the type space $(\mathcal{L}, \mathcal{P})$ serves as a “poor man’s” ultraproduct of (L, \mathcal{P}) wherein, by the very definition of \mathcal{L} , each type \mathbf{v} is realized.

2.1.1. Bounds on precision and magnitude. Shards. In order for computation state structures to serve their purpose as models of real-world computing, we must keep in mind that computations involving real-valued quantities can only ever carried out approximately, to a preset degree of precision (i.e., to within acceptable error bounds). More subtly, yet equally important in practice, one must also bound the magnitude of real quantities involved in all computations: one cannot possibly store—let alone compute with—arbitrarily large numbers! Thus, although the collection \mathcal{P} of predicates may be infinite (possibly, uncountable even), “true” computations must imply choices, at the outset, of bounds for errors and for absolute magnitudes of quantities involved. Heuristically, different parts of a computation may require suitably matching (different) error bounds, as well as dealing with quantities of very different sizes. Focusing for now on the second of these issues, our results shall generally apply to sets of computations for which arbitrary but fixed bounds for each predicate are set in advance. This leads to the concept of *shard* first introduced in [ADIW24], as per the following definition.

Definition 2.3. Fix a computation states structure $\underline{L} = (L, \mathcal{P})$, with $L \subseteq \mathbb{R}^{\mathcal{P}}$ implicitly identified with its type space. A *sizer* is a family $r_{\bullet} = (r_P)_{P \in \mathcal{P}}$ 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 the 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}_{sh} (the space of *shard-supported types*) be the union of all type-shards as the sizer r_\bullet varies.

Evidently, $\mathcal{L}_{\text{sh}} \subseteq \mathcal{L}$, although the inclusion may be proper. However, the equality $\mathcal{L}_{\text{sh}} = \mathcal{L}$ holds in the important special case when \mathcal{P} is (at most) countable (see [ADIW24]).

2.2. Compositional Computation Structures.

Definition 2.4. A *Compositional Computation Structure* (CCS) is a triple (L, \mathcal{P}, Γ) , 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}, Γ) . The state space L is implicitly identified with a subset of the type space $\mathcal{L} \subseteq \mathbb{R}^\mathcal{P}$ of (L, \mathcal{P}) . We also 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 (as usual) the “exponent” L serves merely as an index set, but the “base” $L \subseteq \mathbb{R}^\mathcal{P}$ has the relative product topology (of pointwise convergence). In this manner, the semigroup Γ is identified with a subset of the topological product $\mathcal{L}^L \subseteq (\mathbb{R}^\mathcal{P})^L$, which leads to an inclusion $\overline{\Gamma} \subseteq \mathcal{L}^L$. Elements $\xi \in \overline{\Gamma}$ are called (real-valued) *deep computations* or *ultracomputations*.

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 $\gamma|_{L[r_\bullet]} : L[r_\bullet] \rightarrow L[r_\bullet]$ (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.

Unlike its subspace L^L (a topological semigroup), the space \mathcal{L}^L does not have a natural semigroup structure, so ultracomputations are not generally composable. Intuitively, ultracomputations are “final” or “ultimate”.

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

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 $\gamma|_{L[r_\bullet^{(v)}]} : L[r_\bullet^{(v)}] \rightarrow L[r_\bullet^{(v)}]$, 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 $\mathcal{L}_{\text{sh}}^L \supseteq K$ *a fortiori*. \square

For a CCS (L, \mathcal{P}, Γ) , 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 deep computations 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.

Remarks 2.6. (For readers versed in model theory.)⁴

- (1) compositional computation structures as defined above are not real-valued structures in a literal sense, because there is no intrinsic structural manner to ensure that Γ is realized as a literal subset of the product space L^L . However, in a semantically transparent manner, any CSS as above may alternatively be seen (per its original definition in [ADIW24]) as a real-valued structure $(L, \mathcal{P}, \Gamma, \circ, \text{ev})$, where
 - (L, \mathcal{P}) is a computation states structure,
 - (Γ, \circ) is a semigroup,
 - $\text{ev} : \Gamma \times L \rightarrow L : (\gamma, v) \mapsto \text{ev}(\gamma, v)$ is a semigroup action of Γ on L (i.e., $\text{ev}(\gamma \circ \delta, v) = \text{ev}(\gamma, \text{ev}(\delta, v))$ for $\gamma, \delta \in \Gamma$ and $v \in L$).
Both definitions are reconciled upon identifying $\gamma \in \Gamma$ with the map $(\gamma, v) \mapsto \text{ev}(\gamma, v)$.
- (2) In the preceding sense of real-valued structures, the class of compositional computation structures (for a fixed predicate symbol collection \mathcal{P}) is first-order elementary. The limited notion of state type we have introduced (essentially, quantifier-free state types with parameters of sort L only) may be enriched to state types involving more formulas (e.g., with parameters in Γ , or including quantified formulas), and extended to “transform-types” (i.e., types of sort Γ). In particular, given a structure $\underline{C} = (L, \mathcal{P}, \Gamma, \circ, \text{ev})$, all types are realized in some ultrapower $\underline{C}^{(U)}$ of C (constructed in the usual manner detailed in the references above).⁵ This approach to spaces of types was introduced in Banach space theory by Krivine [Kri74, KM81, Kri72].

2.3. Computability of deep computations and the Extendibility Axiom. Let (L, \mathcal{P}) be a computation structure with *countable* predicate collection \mathcal{P} .

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

It is shown in [ADIW24] that computable transforms $f : L \rightarrow \mathcal{L}$ are precisely those that extend to continuous maps $\tilde{f} : \mathcal{L} \rightarrow \mathcal{L}$.

To summarize, for a function $f : L \rightarrow \mathcal{L}$, the following conditions are equivalent:

- f is computable
- f has uniformly polynomially approximable features on shards
- f extends to a continuous map $\mathcal{L} \rightarrow \mathcal{L}$.

These equivalences motivate the following definition:

A computation $\gamma \in \Gamma$ is *extendable* if it has some extension $\tilde{\gamma} : \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ such that, for every sizer r_\bullet , there is a sizer s_\bullet such that $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$ is continuous.

⁴Refer to [Kei23] where a general framework for real-valued structures is introduced; see also [DnI17, section 6] for a basic tutorial.

⁵By contrast (cf., Remark 2.2), the ultrapower construction is not necessary to realize state types in a computation states structures.

Because $\tilde{\gamma}$ extends γ continuously on each compact type-shard, we call it the *K-extension* of γ ; such extension is evidently unique (if it exists).

Abusing notation, for any $\Delta \subseteq \Gamma$ consisting of extendable computations, the set $\{\tilde{\gamma}: \gamma \in \Delta\}$ will be denoted $\tilde{\Delta}$.

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

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

For the rest of the paper, the Extendibility Axiom is imposed on all CCSs.

2.4. Newton's method as a CCS. We formalize the Newton-Raphson root-finding method as a CCS [Gal00], [BFB16, Section 2.3].

Let $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the extended complex plane (the Alexandroff —one-point— compactification of \mathbb{C}). We regard the unit sphere $S^2 \subseteq \mathbb{R}^3$ as the Riemann sphere and endowed with the stereographic projection map $S^2 \rightarrow \hat{\mathbb{C}}$ —a topological homeomorphism whose inverse $\hat{\mathbb{C}} \rightarrow S^2$ is given by $z \mapsto (P_1(z), P_2(z), P_3(z))$, where $(P_1(\infty), P_2(\infty), P_3(\infty)) = (0, 0, 1)$ and, for $z \neq \infty$,

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

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

2.4.1. Newton maps. Let $p = p(z)$ be a non-constant rational map with complex coefficients. The *Newton method step* (or *Newton map*) for p is the map

$$(2.7) \quad N_p : z \mapsto z - \frac{p(z)}{p'(z)}.$$

Any rational map extends continuously to an analytic (meromorphic) function on the extended complex plane, so N_p naturally defines a map $\hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$; therefore, successive iterations of the Newton-Raphson method (i.e., compositional powers $N_p^{(n)}$) may be regarded as computations of a CCS whose states space is $L = \hat{\mathbb{C}}$. The natural choice of predicates is $\mathcal{P} = (P_1, P_2, P_3)$, the inverse stereographic projection coordinates. (The real- and imaginary part maps $z \mapsto \Re z, z \mapsto \Im z$ are not defined at $z = \infty$.)

2.4.2. Newton CCSs. Let $p = p(z)$ be a nonconstant rational map $\hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$. We say that the CCS $(\hat{\mathbb{C}}, \mathcal{P}, \Gamma_p)$ is *Newton's method for $p(z)$* (or the *Newton CCS for $p(z)$*) if:

- $\mathcal{P} := (P_1, P_2, P_3)$ are the coordinates of inverse stereographic projection (as above);
- $\Gamma_p := \{N_p^{(n)} : n \in \mathbb{N}\}$ consists of iterates of the Newton map N_p in equation (2.7).

The states space $\hat{\mathbb{C}}$ is compact and the Newton map N_p is continuous thereon, so the Extendibility Axiom is satisfied by Newton CCSs.

Given an initial estimate z_0 , one hopes for the sequence of Newton steps applied to z_0 (namely, $N_p^{(\bullet)}(z_0) := (z_n)_{n \in \mathbb{N}}$, where $z_n := N_p^{(n)}(z_0)$) to converge to a root r of p .

A real cubic without rational roots. It is well known—if perhaps surprising at first—that Newton’s method fails to converge to any root⁶ in many cases. Consider, as a first example, the polynomial $\rho(z) = z^3 - 2z + 2$ with Newton map

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

Successive Newton steps obtained from the initial guess $z_0 = 0$ yield the divergent sequence $N_\rho^{(\bullet)}(0) = (0, 1, 0, 1, \dots)$ having sequential limits 0, 1—neither a root of ρ . Visualizing Newton iterates. Assign to each complex point $z \in \mathbb{C}$ an “RGB color” $(R, G, B) \in [0, 1]^3$ whose entries give the corresponding (Red/Green/Blue) color intensity; thus, $(1, 0, 0)$ is red, $(0, 1, 0)$ is green, and $(0, 0, 1)$ is blue—while, for example, $(0.5, 0, 0.5)$ is a light purple. Each of the three distinct roots of ρ will be given one of these colors (the real root r_0 , red; the roots $r_{\pm 1}$ on the upper- and r_{-1} on the lower half plane, respectively green and blue). The RGB color of all other $z \in \mathbb{C}$ is a convex combination whose relative weights are $1/d_i$ where $d_i = |z - r_i|$ for $i = 0, \pm 1$; this will be called the “standard” color of z : the closer z is to root r_i , the closer the colors of these points are.⁷

Recall that the *Fatou set* \mathcal{F}_q of a rational map $q = q(z)$ on $\hat{\mathbb{C}}$ consists of all points that have a compact neighborhood on which some subsequence of the sequence $q^{(\bullet)}$ of iterates of q converges uniformly; the *Julia set* $\mathcal{J}_q := \hat{\mathbb{C}} \setminus \mathcal{F}_q$ is its complement: it is (necessarily closed and) typically a fractal. The Julia set of the Newton map N_q is called the *Newton fractal* of q .

For any $n \in \mathbb{N}$, define a coloring of \mathbb{C} in the natural way: color each $z_0 \in \mathbb{C}$ using the standard color of $z_n := N_\rho^{(n)}(z_0)$. At any stage, particularly early ones, the coloring looks “fuzzy” (out of focus) because the coloring function is continuous; however, as n increases, the coloring becomes “sharper” (the coloring function has large regions of near-constancy, in the transition areas between which color varies rapidly—i.e., is “nearly discontinuous”). For one thing, for many z_0 in the Fatou set, the (entire) sequence (z_n) converges to one of the three roots: the RGB colors of (z_n) tend towards (a single one of) red, green or blue correspondingly. The Fatou set also has an open region consisting of points z_0 for which $(z_n)_{n \in \mathbb{N}}$ has a subsequence converging to 0 or 1 (uniformly on some compact neighborhood of z_0). The four open regions are immediately recognizable in Figure ??.

The deep computations of

The Newton fractal of the 3-cyclotomic polynomial. Consider next the 3-cyclotomic polynomial $\kappa(z) = z^3 - 1$. The roots of $\kappa(z)$ are the cube roots of unity $r_0 = 1$ and $r_{\pm 1} = \exp(\pm 2\pi i/3)$; the Newton map is

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

In Figure 2(C), one sees a Fatou set consisting of solid-color (disconnected) open sets, one in each of the colors red (for $r_0 = 1$), green and blue (for $r_{\pm 1}$, respectively), where the sequence of all iterates $z_n := N_\kappa^{(n)}(z_0)$ starting with any point z_0

⁶Even when $z = \infty$ happens to be a root of a rational map $p(z)$.

⁷While ∞ is a fixed point of N_ρ , it is a repeller: if $|z|$ is large, then $N_\rho(z) = \frac{2}{3}z + O(1)$ has roughly two-thirds its magnitude. Therefore, ∞ is not the limit of any subsequence of $N_\rho^{(\bullet)}(z_0)$ with z_0 in the Fatou set of N_ρ ; in particular, ∞ is a point of the Newton fractal of $\rho(z)$.

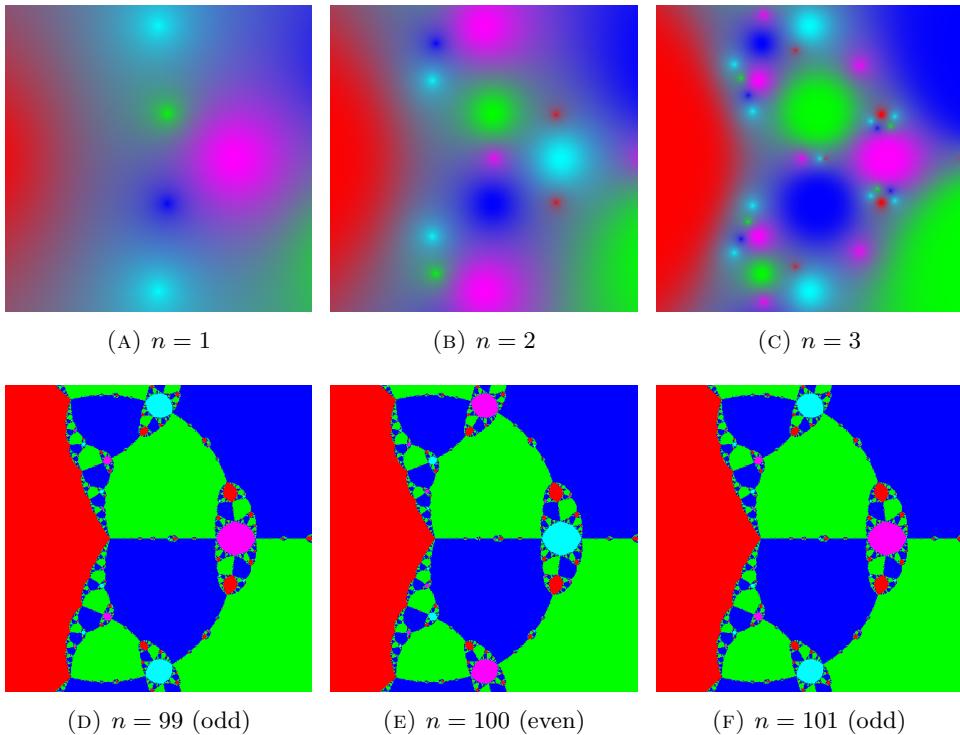


FIGURE 1. Iterates of the Newton map N_ρ of $\rho(z) = z^3 - 2z + 2$. Points in the basin of each root $r_0, r_{\pm 1}$ (respectively, red, green and blue) converge to the root. The basin of attraction of the 2-cycle $\{0, 1\}$ is the union of magenta and cyan regions; these two colors alternate like semaphore lights: points z in the cyan region are those whose subsequence $N_\rho^{(2\bullet)}(z)$ of even Newton iterates (resp., $N_\rho^{(2\bullet+1)}(z)$ of odd iterates) has limit 1 (resp., limit 0). For points z in the magenta region, the limits of the even and odd iterate subsequences are exchanged. The boundaries of colored regions constitute the Newton fractal of ρ .

converges to the corresponding root. (Thus, each solid-color region is the basin of attraction of the respective root).

The set of deep computations $\overline{\Gamma}$ might behave different for various polynomials. Let us look at various examples:

Example 2.8. Computation of square roots. Let a be a positive real number and $p(x) = x^2 - a$. Let $\mathbb{R} \rightarrow S^1 : x \mapsto (P_1(x), P_2(x))$ be the (inverse) stereographic projection of the real line onto $S^1 \subseteq \mathbb{R}^2$, i.e.,

$$P_1(x) = \frac{2x}{x^2 + 1}, \quad P_2(x) = \frac{x^2 - 1}{x^2 + 1},$$

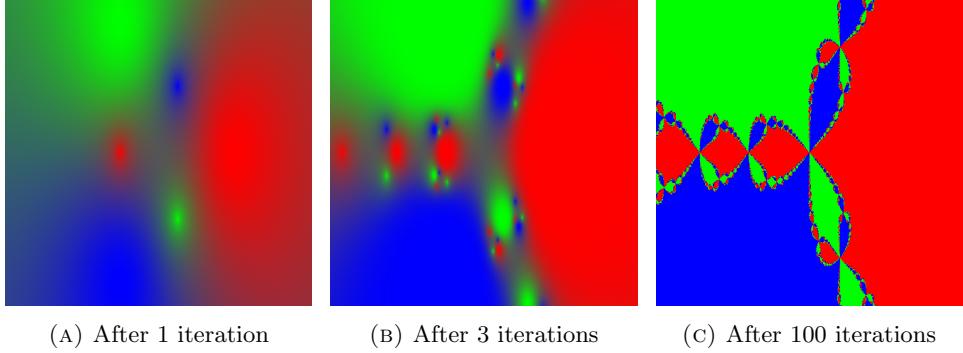


FIGURE 2. Visualization of the Newton iterates of $\kappa(z) = z^3 - 1$.

extended to a projective transformation of $\mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$ by $\infty \mapsto (P_1(\infty), P_2(\infty)) = (0, 1)$. Let $\mathcal{P} = (P_1, P_2)$ and $\Gamma = \{N_p^n : n \in \mathbb{N}\}$, where $N_p : \mathbb{RP}^1 \rightarrow \mathbb{RP}^1$ is given by

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

As before, (L, \mathcal{P}, Γ) is a CCS with $L := \mathbb{RP}^1 \setminus \{0, \infty\} = \mathbb{R} \setminus \{0\}$ whose type space is $\mathcal{L} = S^1 \cong \mathbb{RP}^1$. In fact (with the identification $\mathcal{L} \cong \mathbb{RP}^1$), the (full) sequence of iterates of the Newton map has a pointwise limit $f : \mathcal{L} \rightarrow \mathcal{L}$, namely

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

Thus, f is discontinuous at $x = 0, \infty$. The set of deep computations is $\{N_p^n\}_n \cup \{f\} \subseteq B_1(\mathcal{L}, \mathcal{L})$.

Example 2.9. Newton's method for $\rho(z) = z^3 - 2z + 2$. Let B_i be the basin of the root r_i ($i = 0, \pm 1$), and let B_{01} be the basin of the attractive cycle $0, 1, 0, 1, \dots$. It may be shown that the Fatou set of N_ρ is $F_\rho = B_{01} \cup B_0 \cup B_{+1} \cup B_{-1}$, each of these four sets being nonempty open. Since B_{01} is nonempty, N_ρ^n does not converge pointwise; however, the subsequences N_ρ^{2n} and N_ρ^{2n+1} are pointwise convergent to two distinct deep computations f_0 (constant 0 on B_{01}) and f_1 (constant 1 on B_{01}).⁸ The behavior of the iterates N_ρ on the Newton fractal of ρ (the Julia set $\mathbb{C} \setminus L_\rho$) is chaotic.

2.5. Finite-precision threshold classifiers. Consider the states space $L = 2^\mathbb{N}$ ($= \{0, 1\}^\mathbb{N}$) consisting of all countable binary sequences $x = (x_n)_{n \in \mathbb{N}} \subseteq \{0, 1\}$, endowed with the topology of pointwise convergence. Let $\mathcal{P} = \{P_n : n \in \mathbb{N}\}$ be the collection of projections (coordinates) $P_n : x \mapsto x_n$. Since $2^\mathbb{N} \subseteq \mathbb{R}^\mathbb{N}$ ($\cong \mathbb{R}^\mathcal{P}$) is closed, the type space of $(2^\mathbb{N}, \mathcal{P})$ is $\mathcal{L} = 2^\mathbb{N}$ ($= L$).

Since the index set \mathbb{N} and the value space $2 = \{0, 1\}$ are well-ordered, the states space $2^\mathbb{N}$ is lexicographically ordered, say by \preccurlyeq .

The set 2^n of finite binary strings w (“words”) of any fixed length $|w| := n$ will be implicitly identified with a subset of $2^\mathbb{N}$ by identifying w with its zero extension

⁸As remarked above, ∞ is a repelling fixed point of N_ρ , hence $f_0(\infty) = \infty = f_1(\infty)$.

$\iota(w) = (x_i)_{i \in \mathbb{N}} \in 2^{\mathbb{N}}$, namely

$$x_i = \begin{cases} w_i, & i < n; \\ 0, & i \geq n. \end{cases}$$

The set of all finite binary strings is thus identified with the subset $2^{<\mathbb{N}} := \bigcup_{n \in \mathbb{N}} 2^n \subseteq 2^{\mathbb{N}}$. The “inclusion map” ι above has a natural left inverse: the restriction map

$$\begin{aligned} \cdot \upharpoonright_n : 2^{\mathbb{N}} &\rightarrow 2^n \\ x &\mapsto x \upharpoonright_n := (x_i)_{i < n}. \end{aligned}$$

Denote by 0_{\bullet} (resp., 1_{\bullet}) the constant sequence with value 0 (resp., value 1). For any fixed word $w \in 2^n$, consider the transition

$$\begin{aligned} \phi_w : 2^{\mathbb{N}} &\rightarrow 2^{\mathbb{N}} \\ x &\mapsto \begin{cases} 1_{\bullet}, & x \upharpoonright_n \preceq w; \\ 0_{\bullet}, & \text{otherwise.} \end{cases} \end{aligned}$$

Such ϕ_w are called *finite-precision classifiers*.

Since $\{u \in 2^n : u \preceq w\} \subseteq 2^n$ is clopen⁹ and $\cdot \upharpoonright_n : 2^{\mathbb{N}} \rightarrow 2^n$ is evidently continuous, it follows that both $\phi_w^{-1}(1_{\bullet}) = \{x \in 2^{\mathbb{N}} : x \upharpoonright_n \preceq w\}$ and $\phi_w^{-1}(0_{\bullet}) = 2^{\mathbb{N}} \setminus \phi_w^{-1}(1_{\bullet})$ are clopen, so ϕ_w is continuous.

Let $\Delta = \{\phi_w : w \in 2^{<\mathbb{N}}\} \cup \{\mathbf{0}, \mathbf{1}\}$, where $\mathbf{0}, \mathbf{1} : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}}$ are the constant maps $x \mapsto 0_{\bullet}$ and $x \mapsto 1_{\bullet}$, 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}, Γ) 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. [COMPLETELY FALSE!] The topological structure is far more interesting. 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 $\overline{\Delta}$ 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 exact (i.e., infinite-precision) 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 $f_a^+ : L \rightarrow \{0^\infty, 1^\infty\}$ 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 2.10. $f_a^+, f_a^- \in \overline{\Delta}$ for all $a \in 2^{\mathbb{N}}$.

⁹ 2^n is a finite and Hausdorff —hence discrete— topological space.

Proof. First, we show that $f_a^+ \in \overline{\Delta}$. If $a = 1^\infty$, then $f_a^+ = 1^\infty \in \Delta$. If a is not identically 1, we argue that the pointwise limit of the threshold classifiers on $w_n := a|_n 1$ (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 $\lim_{n \rightarrow \infty} \phi_{w_n}(x) = f_a^+(x)$. 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 $f_a^- = 0^\infty \in \Delta$. 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 $a = u \widehat{\cup} 1 \widehat{\cup} 0^\infty$ (here $\widehat{\cup}$ denotes concatenation) for some finite u . Let $w_n := u \widehat{\cup} 0 \widehat{\cup} 1^n <_{\text{lex}} a$. We claim that $\lim_{n \rightarrow \infty} \phi_{w_n}(x) = f_a^-(x)$. 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 $w_k := (a|_{n_k-1}) \widehat{\cup} 0$; 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 $k \geq K$.

□

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

2.6. 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 2.5). For each $w \in 2^{<\mathbb{N}}$, let $\psi_w : L \rightarrow L$ 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 $2^\mathbb{N}$, we see that the transitions ψ_w are all continuous. In particular, (L, \mathcal{P}, Γ) 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 $\delta_a : L \rightarrow \{0^\infty, 1^\infty\}$ be the indicator function at a , i.e., $\delta_a(x) = 1^\infty$ if $x = a$ and $\delta_a(x) = 0^\infty$ otherwise.

Proposition 2.11. $\delta_a \in \overline{\Delta}$ for all $a \in 2^\mathbb{N}$.

Proof. Fix $a \in 2^\mathbb{N}$, and let $w_n := a|_n$ for each $n \in \mathbb{N}$. We claim that $\lim_{n \rightarrow \infty} \psi_{w_n}(x) = \delta_a(x)$ 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 2.12. $\mathbf{0}^\infty \in \overline{\Delta}$.

Proof. To show that $\mathbf{0}^\infty \in \overline{\Delta}$, for each $n \in \mathbb{N}$, consider, $w_n = 1^n \cap 0$, 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. This is the main section of the paper. We use topology of function spaces to obtain a new hierarchy of complexity classes for deep computations. The main topological tool is Todorčević's trichotomy for Rosenthal compacta [Tod99]. Our inspiration is the work of Glasner and Megrelishvili on tame semigropus in topological dynamics [GM22].

Theorem 3.1 below is the cornerstone of the section; it establishes a powerful link between deep computations and Rosenthal compacta. It extends the results of [ADIW24] from the setting of continuous deep computations to the more general (and realistic) setting of deep computations of the first Baire class. Informally, Theorem 3.1 states that if \mathcal{P} and $\Delta \subseteq \Gamma$ are countable, then the space of deep computations generated by Δ is a Rosenthal compactum (when restricted to shards) if and only if Δ satisfies the NIP feature by feature. Furthermore, Theorem 3.1 shows that every deep computation generated by Δ is a sequential limit (rather than an abstract ultrafilter limit) of computations in Δ .

Theorem 3.1. *Let (L, \mathcal{P}, Γ) be a compositional computation structure (Definition 2.4) satisfying the Extendibility Axiom (see Definition 2.3) with \mathcal{P} countable. Let R be an exhaustive collection of sizers, and suppose that $\Delta \subseteq \Gamma$ is R -confined. Then the following conditions 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

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

Furthermore, 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: there is $\tilde{f} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$ such that $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ for all $r_\bullet \in R$. 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. 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 1.10 and Lemma 1.11 prove the equivalence of (i) and (ii). If (i) holds and $f \in \overline{\Delta}$, then we can write as an ultralimit of computations in $\underline{\Delta}$, say $f = \mathcal{U}\lim_i \gamma_i$. Define $\tilde{f} := \mathcal{U}\lim_i \tilde{\gamma}_i$. Then, for all $r_\bullet \in R$ we have $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$. The assertion 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 any point in the closure of a set A is the limit of a sequence of points in A ; see Theorem 3F in [BFT78] or Theorem 4.1 in [Deb13]). \square

3.2. The Todorčević trichotomy, and levels of NIP and PAC learnability. 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}, Γ) 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 $\text{NIP}_3 \Rightarrow \text{NIP}_2 \Rightarrow \text{NIP}_1 \Rightarrow \text{NIP}$. We now present examples of separable and non-separable examples Rosenthal compacta that show that implications are strict. These examples are due to Todorčević [Tod99].

Examples 3.3.

- (1) *Alexandroff compactification of a discrete space of size continuum.* For each $a \in 2^{\mathbb{N}}$ consider the map $\delta_a : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ given by $\delta_a(x) = 1$ if $x = a$ and $\delta_a(x) = 0$ otherwise. Let $A(2^{\mathbb{N}}) = \{\delta_a : a \in 2^{\mathbb{N}}\} \cup \{\mathbf{0}\}$, where $\mathbf{0}$ 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 an uncountable discrete subspace of $B_1(2^{\mathbb{N}})$, and its pointwise closure is precisely $A(2^{\mathbb{N}})$. Hence, this is a Rosenthal compactum that is not hereditarily separable (and therefore not first countable). In particular, this space is not separable. However, as the next example shows, it can be made separable space by “softening” the neighborhoods of each point through the addition of a countable set.
- (2) *Extended Alexandroff compactification.* For each finite binary sequence $s \in 2^{<\mathbb{N}}$, let $v_s : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ 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 that is not first countable. This is the example

discussed in Section 2.6. 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 $2^{\mathbb{N}}$ of infinite binary sequences. For each $a \in 2^{\mathbb{N}}$ let $f_a^- : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be given by $f_a^-(x) = 1$ if $x < a$ and $f_a^-(x) = 0$ otherwise. Let $f_a^+ : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be given by $f_a^+(x) = 1$ if $x \leq a$ and $f_a^+(x) = 0$ otherwise. The Split Cantor is the space $S(2^{\mathbb{N}}) = \{f_a^- : a \in 2^{\mathbb{N}}\} \cup \{f_a^+ : a \in 2^{\mathbb{N}}\}$. This space was obtained as the closure of the space discussed in Section 2.5, thus giving an example that separates NIP_2 from NIP_3 .

This is a well known separable Rosenthal compactum; an 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. It is homeomorphic to the space $2^{\mathbb{N}} \times \{0, 1\}$ with the lexicographic order topology via the identification $(a, 1) \mapsto f_a^+$ and $(a, 0) \mapsto f_a^-$.

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

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

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

Let $D(K) = \{g_a^0 : a \in K\} \cup \{g_a^1 : a \in K\}$. Observe that all points g_a^1 are isolated and that open neighborhoods of $g_{a_0}^0$ are of the form $\{g_a^i : a \in U, i \in \{0, 1\}\} \setminus \{g_a^1 : a \in F\}$ where $U \subseteq K$ is an open neighborhood of a_0 and $F \subseteq K$ is a finite set.

Another way to present this space is as the space $K \times \{0, 1\}$ with the identifications $(a, 0) \mapsto g_a^0$ and $(a, 1) \mapsto g_a^1$ and the basic open neighborhoods are given as above. We can also embed $D(K)$ into the product $A(K) \times K$ by identifying $(a, 0) \mapsto (\mathbf{0}, a)$ and $(a, 1) \mapsto (\delta_a, a)$.

Notice that $D(K)$ is a first countable Rosenthal compactum. It is not separable if K is uncountable; thus, we typically study the interesting case when $K = 2^{\mathbb{N}}$.

As with the Alexandroff compactification $A(2^{\mathbb{N}})$, we can make the space $D(2^{\mathbb{N}})$ separable by “softening” the neighborhoods of each point through the addition of a countable set. For example, the closure of the set $\{(v_s, s^\frown 0^\infty) : s \in 2^{<\mathbb{N}}\} \subseteq \hat{A}(2^{\mathbb{N}}) \times 2^{\mathbb{N}}$ is $\{(\mathbf{0}, a) : a \in 2^{\mathbb{N}}\} \cup \{(\delta_a, a) : a \in 2^{\mathbb{N}}\} \cup \{(v_s, s^\frown 0^\infty) : s \in 2^{<\mathbb{N}}\}$, where $\{(\mathbf{0}, a) : a \in 2^{\mathbb{N}}\} \cup \{(\delta_a, a) : a \in 2^{\mathbb{N}}\}$ is homeomorphic to $D(2^{\mathbb{N}})$.

- (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 $h_t : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ by

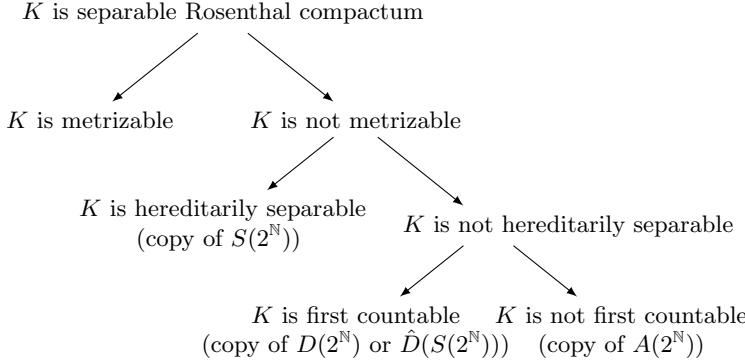
$$h_t(x) = \begin{cases} 0, & \text{if } x < a_t; \\ 1/2, & \text{if } a_t \leq x \leq b_t; \\ 1, & \text{if } b_t < x. \end{cases}$$

Let $\hat{D}(S(2^{\mathbb{N}}))$ be the pointwise closure of the set $\{h_t : t \in 2^{<\mathbb{N}}\}$. The identification $h_t \mapsto (v_t, f_{t^-0^\infty}^+)$ lifts to a homeomorphism between $\hat{D}(S(2^{\mathbb{N}}))$ and the subspace of $A(2^{\mathbb{N}}) \times S(2^{\mathbb{N}})$ consisting of $(\mathbf{0}, f_a^+)$, $(\mathbf{0}, f_a^-)$, (δ_a, f_a^+) and $(v_t, f_{t^-0^\infty}^+)$ for $a \in 2^{\mathbb{N}}$ and $t \in 2^{<\mathbb{N}}$ (see 4.3.7 in [ADK08]). Hence, $\hat{D}(S(2^{\mathbb{N}}))$ is a separable first countable Rosenthal compactum that is not hereditarily separable. In fact, it contains an uncountable discrete subspace.

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 for separable Rosenthal compacta:



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}, Γ) be the computation of square root (example 2.8 with $\Delta = \Gamma$). We saw that $\overline{\Delta} = \tilde{\Delta} \cup \{\tilde{f}\} \subseteq B_1(\mathcal{L}, \mathcal{L})$. 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 $\overline{\Delta} \not\subseteq C(\mathcal{L}, \mathcal{L})$.
- (2) Let (L, \mathcal{P}, Γ) be the finite precision threshold classifiers (Section 2.5) 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}, Γ) be the CCS given by $L = 2^{\mathbb{N}}$, $\mathcal{P} = \{P_n : n \in \mathbb{N}\}$ and Γ is the semigroup generated by $\Delta = \{\gamma_t : t \in 2^{<\mathbb{N}}\}$, where $P_n : 2^{\mathbb{N}} \rightarrow \{0, 1\}$ is the

projection map $P_n(x) = x(n)$ and $\gamma_t : L \rightarrow L$ is given by

$$\gamma_t(x) = \begin{cases} 0^\infty, & \text{if } x <_{\text{lex}} t^\frown 0^\infty; \\ (01)^\infty, & \text{if } t^\frown 0^\infty \leq_{\text{lex}} x \leq_{\text{lex}} t^\frown 1^\infty; \\ 1^\infty, & \text{if } t^\frown 1^\infty <_{\text{lex}} x. \end{cases}$$

where $(01)^\infty$ denotes the sequence of alternating bits: $010101\dots$. As in the other examples, it is not difficult to see that (L, \mathcal{P}, Γ) satisfies the Extendibility Axiom. For example, the condition $t^\frown 0^\infty \leq_{\text{lex}} x \leq_{\text{lex}} t^\frown 1^\infty$ is equivalent to x extending t . Observe that the set of deep computations is homeomorphic to $\hat{D}(S(2^\mathbb{N}))$ (see Example 3.3-(5)). Hence, Δ is NIP₁ but not NIP₂.

- (4) Let (L, \mathcal{P}, Γ) be the finite precision prefix test (Section 2.6) with $\Delta = \{\psi_w : w \in 2^{<\mathbb{N}}\}$. We saw that $\bar{\Delta}$ 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 NIP₁.

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 because our emphasis is computational. Real numbers can be represented as infinite binary sequences, i.e., infinite branches of the binary tree $2^\mathbb{N}$, and we approximate real numbers with elements in $2^{<\mathbb{N}}$, i.e., finite bitstrings.

Definition 3.7. Let X be a Polish space.

- (1) If I is countable and $\{f_i : i \in I\} \subseteq \mathbb{R}^X$, $\{g_i : i \in I\} \subseteq \mathbb{R}^X$ are two 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 $t \in 2^{<\mathbb{N}}$, we denote by $t^\frown 0^\infty$ ($t^\frown 1^\infty$) the infinite binary sequence starting with t and continuing with 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 $s, s' \in R$, we have $s^\frown 0^\infty \neq s'^\frown 0^\infty$ 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 $f_a^+ : 2^\mathbb{N} \rightarrow \{0, 1\}$ be the characteristic function of $\{x \in 2^\mathbb{N} : a \leq x\}$ and let

$f_a^- : 2^{\mathbb{N}} \rightarrow \{0, 1\}$ be the characteristic function of $\{x \in 2^{\mathbb{N}} : a > x\}$. Given two maps $f, g : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ we denote by $(f, g) : 2^{\mathbb{N}} \sqcup 2^{\mathbb{N}} \rightarrow \mathbb{R}$ 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) $D_2 = \{s_t \cap 0^\infty : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_2} = 2^{\leq N}$.
- (3) $D_3 = \{f_{s_t \cap 0^\infty}^+ : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_3} = S(2^{\mathbb{N}})$.
- (4) $D_4 = \{f_{s_t \cap 1^\infty}^- : t \in 2^{<\mathbb{N}}\}$. 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) $D_6 = \{(v_{s_t}, s_t \cap 0^\infty) : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_6} = \hat{D}(2^{\mathbb{N}})$.
- (7) $D_7 = \{(v_{s_t}, x_{s_t \cap 0^\infty}^+) : t \in 2^{<\mathbb{N}}\}$. 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 $i = 1, 2, \dots, 7$ 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, the families D_i are minimal and mutually non-equivalent.*

The implication 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}, Γ)), we can always find a countable discrete set of deep computations that approximates all the other deep computations. For example: in the finite precision prefix test example (subsection 2.6), 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 countability assumption on \mathcal{P} played a crucial role in the proof of Theorem 1.10, as it makes $\mathbb{R}^{\mathcal{P}}$ a Polish space.) The main results of the section are Theorem 4.8 (which connects NIP and Monte Carlo computability) and 4.14 (which connects Talagrand stability and Monte Carlo computability).

Fundamental in this section is a measure-theoretic version of Theorem 1.10, namely, Theorem 4.5. For the proof of Theorem 1.10, 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. 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 $U \subseteq Y$. This motivates the following definition.

Definition 4.1. Given a Hausdorff space X and a measurable space (Y, Σ) , we say that $f : X \rightarrow Y$ is *universally measurable* (with respect to Σ) if $f^{-1}(E)$ is μ -measurable for every Radon measure μ on X and every $E \in \Sigma$. When $Y = \mathbb{R}$, we will always take $\Sigma = \mathcal{B}(\mathbb{R})$, the Borel σ -algebra of \mathbb{R} .

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

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 $\mathcal{M}^0(X, \mu)$.

In the context of deep computations, we are interested in transition maps of a state space $L \subseteq \mathbb{R}^{\mathcal{P}}$ into itself. In the product space $\mathbb{R}^{\mathcal{P}}$, we have two natural σ -algebras: 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 set $\pi_i^{-1}(U)$ with $U \subseteq \mathbb{R}$ open). In general, the cylinder σ -algebra is strictly smaller, but when \mathcal{P} is countable, both σ -algebras coincide. We will use the cylinder σ -algebra to define universally measurable maps $f : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}^{\mathcal{P}}$. 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 $f : X \rightarrow Y$:*

- (i) $f : X \rightarrow Y$ is universally measurable (with respect to Σ_Y).
- (ii) $\pi_i \circ f : X \rightarrow Y_i$ 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 1.9).

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 $\mathcal{M}^0(X, \mu)$, i.e., every countable subset of A has a limit point in $\mathcal{M}^0(X, \mu)$.

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 $f : X \rightarrow Y$ 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 $\mu(E \setminus F) < \varepsilon$.
- (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}, Γ) be a CCS. We say that a transition $f : L \rightarrow L$ is *universally Monte Carlo computable* if and only if there exists $\tilde{f} : \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ extending f such that for every sizer r_{\bullet} there is a sizer s_{\bullet} such that the restriction $\tilde{f}|_{\mathcal{L}[r_{\bullet}]} : \mathcal{L}[r_{\bullet}] \rightarrow \mathcal{L}[s_{\bullet}]$ is universally measurable, i.e., $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_{\bullet}]} : \mathcal{L}[r_{\bullet}] \rightarrow [-s_P, s_P]$ is μ -measurable for every Radon probability measure μ on $\mathcal{L}[r_{\bullet}]$ and $P \in \mathcal{P}$.

Remark 4.7. Every Radon measure on a compact space (e.g., \mathcal{L}_{sh}) is finite; hence, by proper normalization, it can be treated as a probability measure. For this reason, in the context of Monte Carlo computability (4.6), we restrict our attention to Radon probability measures instead of general Radon measures.

Theorem 4.8. Let (L, \mathcal{P}, Γ) 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 1.11. By Theorem 4.5, we have $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_{\bullet}]} \subseteq M_r(\mathcal{L}[r_{\bullet}])$ 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, $\tilde{f} : \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ extends f . Since Δ is R -confined we have that $f : L[r_{\bullet}] \rightarrow L[r_{\bullet}]$ and $f : \mathcal{L}[r_{\bullet}] \rightarrow \mathcal{L}[r_{\bullet}]$ for all $r_{\bullet} \in R$. Lastly, note that for all $r_{\bullet} \in R$ and $P \in \mathcal{P}$ we have that $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_{\bullet}]} \in \pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_{\bullet}]} \subseteq M_r(\mathcal{L}[r_{\bullet}])$. \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:

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

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 $k \geq 1$ 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, i.e., limit points of a Talagrand stable set of computations are Monte Carlo computable; this is Theorem 4.14 below. We now prove that limit points 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 , Σ is the completion of the Borel σ -algebra by μ , and \mathcal{K} is the collection of compact subsets of X , all these conditions hold). Then, $f : X \rightarrow \mathbb{R}$ 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 $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$.*

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 $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$. Suppose that there exists $f \in \overline{A}$ such that $f \notin \mathcal{M}^0(X, \mu)$. 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. \square

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 $\mathcal{M}^0(X, \mu)$ for every Radon probability measure μ on X . Since A is Talagrand μ -stable for any such μ , we have $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$ by Lemma 4.11. In particular, A is relatively countably compact in $\mathcal{M}^0(X, \mu)$. \square

Corollary 4.14. *Let (L, \mathcal{P}, Γ) 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. \square

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., if A is a set of computations, then $A \subseteq C_p(X)$.

5. CONCLUSION

We have seen how the tools of model theory and the topology of function spaces can give us new insights into classifying the difficulty of computations. In future papers, we plan to extend this research in a variety of directions. A direction of particular interest to us is replacing classical computation by quantum computation. Topologically, we expect to extend some of our results to K -countably determined (Lindelöf Σ) spaces or K -analytic spaces, which have previously been used in functional analysis — see e.g., Kakol *et al* [KKLPS25]. Our restricting our considerations to continuous and Baire class 1 functions can be relaxed by applying more high-powered model-theoretic and topological techniques. This is work in progress. We have hinted at applications to deep equilibria; see [ADIW24]. In another direction, we have hinted at the role additional set-theoretic axioms may play. Since real world computations are finite and hence do not depend on other than the usual set-theoretic axioms, in principle we can assume these axioms and then dispense with them; however, this is delicate, since many of our techniques involve infinite sets so we have to check that our assertions are sufficiently simple so as to be *absolute*, i.e., unaffected by additional axioms. We hope that this paper, the preceding one [ADIW24], and the future ones will stimulate interactions among these three seemingly unrelated fields of research.

REFERENCES

- [ADIW24] Samson Alva, Eduardo Dueñez, Jose Iovino, and Claire Walton. Approximability of deep equilibria. *arXiv preprint arXiv:2409.06064*, 2024. To appear in Mathematical Structures in Computer Science.
- [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.
- [BFB16] Richard L. Burden, J. Douglas Faires, and Annette M. Burden. *Numerical Analysis*. Cengage Learning, Boston, MA, 10th edition, 2016.
- [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/>.
- [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.

- [DnI17] Eduardo Dueñez and José N. Iovino. Model theory and metric convergence I: Metastability and dominated convergence. In *Beyond First Order Model Theory, Volume I*, pages 131–187. CRC Press, Boca Raton, FL, 2017.
- [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.
- [Gal00] A. Galántai. The theory of newton’s method. *Journal of Computational and Applied Mathematics*, 124(1):25–44, 2000. Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.
- [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.
- [HT19] Haim Horowitz and Stevo Todorcevic. Compact sets of baire class one functions and maximal almost disjoint families, 2019.
- [Kei23] 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, 2023.
- [KKLPS25] Jerzy Kąkol, Wiesław Kubiś, Manuel López-Pellicer, and Damian Sobota. *Descriptive topology in selected topics of functional analysis—updated and expanded*, volume 24 of *Developments in Mathematics*. Springer, Cham, second edition, [2025] ©2025.
- [KM81] Jean-Louis Krivine and Bernard Maurey. Espaces de Banach stables. *Israel J. Math.*, 39(4):273–295, 1981.
- [Kri72] J.-L. Krivine. Théorie des modèles et espaces L^p . *C. R. Acad. Sci. Paris Sér. A-B*, 275:A1207–A1210, 1972.
- [Kri74] J.-L. Krivine. Langages à valeurs réelles et applications. *Fund. Math.*, 81:213–253, 1974. Collection of articles dedicated to Andrzej Mostowski on the occasion of his sixtieth birthday, III.
- [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 Todorcević. *Topics in Topology*, volume 1652 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 1997.
- [Tod99] Stevo Todorcević. 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.