

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 study complexity of deep computations. We use topology of function spaces, specifically, the classification Rosenthal compacta, to identify new complexity classes. We use the language of model theory, specifically, the concept of the independence from Shelah’s classification theory, to translate between topology and computation.

0. INTRODUCTION

In this paper we study limit behavior of real-valued computations as the value of certain parameters of the computation model tend towards infinity, or towards zero, or towards some other fixed value, e.g., the depth of a neural network tending to infinity, or the time interval between layers of the network tending toward zero. Recently, particular cases of this situation have attracted considerable attention in deep learning research (e.g., Neural Ordinary Differential Equations [CRBD], Physics-Informed Neural Networks [RPK19], deep equilibrium models [BKK], among others). In this paper, we combine ideas of topology, measure theory, and model theory to study these limit phenomena from a unified viewpoint.

Informed by model theory, to each computation in a given computation model, we associate a continuous real-valued function, called the *type* of the computation, that describes the logical properties of this computation with respect to the rest of the model. This allows us to view computations in any given computational model as elements of a space of real-valued functions, which is called the *space of types* of the model. The idea of embedding models of theories into their type spaces is central in model theory. The embedding of computations into spaces of types allows us to utilize the vast theory of topology of function spaces, known as C_p -theory, to obtain results about complexity of topological limits of computations. As we shall indicate next, recent classification results for spaces of functions provide an elegant and powerful machinery to classify computations according to their levels of “tameness” or “wildness”, with the former corresponding roughly to polynomial approximability and the latter to exponential approximability. The viewpoint of spaces of types, which we have borrowed from model theory, thus becomes a “Rosetta stone” that allows us to interconnect various classification programs: In topology, the classification of Rosenthal compacta pioneered by Todorčević [Tod99]; in logic, the classification of theories developed by Shelah [She90]; and in statistical learning, the notion PAC learning and VC dimension pioneered by Vapkins and Chervonenkis [VC74, VC71].

³⁶ In a previous paper [ADIW24], we introduced the concept of limits of computations, which we called *ultracomputations* (given they arise as ultrafilter limits of

³⁸ standard computations) and *deep computations* (following usage in machine learning [BKK]). There is a technical difference between both designations, but in this
³⁹ paper, to simplify the nomenclature, we will ignore the difference and use only the
⁴⁰ term “deep computation”.

⁴¹ In [ADIW24], we proved a new “tame vs wild” (i.e., polynomial vs exponential)
⁴² dichotomy for complexity of deep computations by invoking a classical result of
⁴³ Grothendieck from late 50s [Gro52]. Under our model-theoretic Rosetta stone,
⁴⁴ polynomial approximability in the sense of computation becomes identified with the
⁴⁵ notion of continuous extendability in the sense of topology, and with the notions of
⁴⁶ *stability* and *type definability* in the sense of model theory.
⁴⁷

⁴⁸ In this paper, we follow a more general approach, i.e., we view deep computations
⁴⁹ as pointwise limits of continuous functions. In topology functions that arise as the
⁵⁰ pointwise limit of a sequence of continuous are called *functions of the first Baire*
⁵¹ class, or *Baire class 1* functions, or *Baire-1* for short; Baire class 1 form a step above
⁵² simple continuity in the hierarchy of functions studied in real analysis (Baire class
⁵³ 0 functions being continuous functions). Intuitively, Baire-1 functions represent
⁵⁴ functions with “controlled” discontinuities, so they are crucial in topology and set
⁵⁵ theory.

⁵⁶ We prove a new “tame vs wild” Ramsey-theoretic dichotomy for complexity of
⁵⁷ general deep computations by invoking a famous paper by Bourgain, Fremlin and
⁵⁸ Talagrand from the late 70s [BFT78], and a new trichotomy for the class of “tame”
⁵⁹ deep computations by invoking an equally celebrated result of Todorčević, from the
⁶⁰ late 90s, for functions of the first Baire class [Tod99].

⁶¹ Todorčević’s trichotomy regards *Rosenthal compacta*; these are special classes of
⁶² topological spaces, defined as compact spaces that can be embedded (homeomor-
⁶³ phically identified as a subset) within the space of Baire class 1 functions on some
⁶⁴ Polish (separable, complete metric) space, under the pointwise convergence topol-
⁶⁵ ogy. Rosenthal compacta exhibit “topological tameness,” meaning that they behave
⁶⁶ in relatively controlled ways, and since the late 70’s, they have played a crucial role
⁶⁷ for understanding complexity of structures of functional analysis, especially, Banach
⁶⁸ spaces. Todorčević’s trichotomy has been utilized to settle longstanding problems
⁶⁹ in topological dynamics and topological entropy [GM22].

⁷⁰ Through our Rosetta stone, Rosenthal compacta in topology correspond to the
⁷¹ important concept of “No Independence Property” (known as “NIP”) in model
⁷² theory, identified by Shelah [She71, She90], and to the concept of Probably Ap-
⁷³ proximately Correct learning (known as “PAC learnability”) in statistical learning
⁷⁴ theory identified by Valiant [Val84].

⁷⁵ Going beyond Todorčević’s trichotomy, we invoke a more recent heptachotomy
⁷⁶ for Rosenthal compacta obtained by Argyros, Dodos and Kanellopoulos [ADK08].
⁷⁷ Argyros, Dodos and Kanellopoulos identified seven fundamental “prototypes” of
⁷⁸ separable Rosenthal compacta, and proved that any non-metrizable separable Rosen-
⁷⁹ thal compactum must contain a “canonical” embedding of one of these prototypes.
⁸⁰ They showed that if a separable Rosenthal compactum is not hereditarily separable,
⁸¹ it must contain an uncountable discrete subspace of the size of the continuum.

⁸² We believe that the results presented in this paper show practitioners of com-
⁸³ putation, or topology, or descriptive set theory, or model theory, how classification
⁸⁴ invariants used in their field translate into classification invariants of other fields.
⁸⁵ However, in the interest of accessibility, we do not assume previous familiarity with

86 high-level topology or model theory, or computing. The only technical prerequisite
 87 of the paper is undergraduate-level topology and measure theory. The necessary
 88 topological background beyond undergraduate topology is covered in section 1.

89 Throughout the paper, we focus on classical computation; however, by refining
 90 the model-theoretic tools, the results presented here can be extended to quantum
 91 computation and open quantum systems. This extension will be addressed in a
 92 forthcoming paper.

93 **CONTENTS**

94 0. Introduction	1
95 1. General topological preliminaries: From continuity to Baire class 1	3
96 1.1. From Rosenthal’s dichotomy to the Bourgain-Fremlin-Talagrand	
97 dichotomy to Shelah’s NIP	5
98 1.2. NIP as universal dividing line between polynomial and exponential	
99 complexity	6
100 1.3. Rosenthal compacta	7
101 1.4. The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with \mathcal{P} countable.	7
102 2. Compositional computation structures.	9
103 3. Classifying deep computations	11
104 3.1. NIP, Rosenthal compacta, and deep computations	11
105 3.2. The Todorčević trichotomy and levels of PAC learnability	12
106 3.3. The Argyros-Dodos-Kanellopoulos heptachotomy, and approximability	
107 of deep computation by minimal classes	14
108 3.4. Bourgain-Fremlin-Talagrand, NIP, and essential computability of deep	
109 computations	15
110 3.5. Talagrand stability, NIP, and essential computability of deep	
111 computations	17
112 References	19

113 **1. GENERAL TOPOLOGICAL PRELIMINARIES: FROM CONTINUITY TO BAIRE**
 114 **CLASS 1**

115 In this section we give preliminaries from general topology and function space
 116 theory. We include some of the proofs for completeness, but the reader familiar
 117 with these topics may skip them.

118 Recall that a subset of a topological space is F_σ if it is a countable union of
 119 closed sets, and G_δ if it is a countable intersection of closed sets. Note that in a
 120 metrizable space, every open set is F_σ ; equivalently, every closed set is G_δ .

121 A *Polish space* is a separable and completely metrizable topological space. The
 122 most important examples are the reals \mathbb{R} , the Cantor space $2^{\mathbb{N}}$ (the set of all infinite
 123 binary sequences, endowed with the product topology), and the Baire space $\mathbb{N}^{\mathbb{N}}$ (the
 124 set of all infinite sequences of naturals, also with the product topology). Countable
 125 products of Polish spaces are Polish; this includes spaces like $\mathbb{R}^{\mathbb{N}}$, the space of
 126 sequences of real numbers.

127 In this paper, we shall discuss subspaces, and so there is a pertinent subtlety of
 128 the definitions worth mentioning: *completely metrizable space* is not the same as
 129 *complete metric space*; for an illustrative example, the interval $(0, 1)$ with the metric

130 inherited from the reals not complete, but it is Polish since that is homeomorphic
 131 to the real line. Being Polish is a topological property.

132 The following result is a cornerstone of descriptive set theory, closely tied to the
 133 work of Wacław Sierpiński and Kazimierz Kuratowski, with proofs often built upon
 134 their foundations and formalized later, notably, involving Stefan Mazurkiewicz's
 135 work on complete metric spaces.

136 **Fact 1.1.** *A subset A of a Polish space X is itself Polish in the subspace topology
 137 if and only if it is a G_δ set. In particular, closed subsets and open subsets of Polish
 138 spaces are also Polish spaces.*

139 Given two topological spaces X and Y we denote by $C_p(X, Y)$ the set of all
 140 continuous functions $f : X \rightarrow Y$ endowed with the topology of pointwise conver-
 141 gence. When $Y = \mathbb{R}$, we denote this collection simply as $C_p(X)$. A natural ques-
 142 tion is, how do topological properties of X translate into $C_p(X)$ and vice versa?
 143 These questions, and in general the study of these spaces, are the concern of C_p -
 144 theory, an active field of research in general topology which was pioneered by A. V.
 145 Arhangel'skiĭ and his students in the 1970's and 1980's. This field has found many
 146 applications in model theory and functional analysis. Recent surveys on the topics
 147 include [HT23] and [Tka11].

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

157 A topological space X is *perfectly normal* if it is normal and every closed subset
 158 of X is a G_δ (equivalently, every open subset of X is a G_δ). Note that every
 159 metrizable space is perfectly normal.

160 The following fact was established by Baire in thesis. A proof can be found in
 161 Section 10 of [Tod97].

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

- 164 • *f is a Baire class 1 function, that is, $f \in B_1(X)$.*
- 165 • *$f^{-1}[U]$ is an F_σ subset of X whenever $U \subseteq \mathbb{R}$ is open.*
- 166 • *f is a pointwise limit of continuous functions.*
- 167 • *For every closed $F \subseteq X$, the restriction $f|_F$ has a point of continuity.*

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

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

170 A subset L of a topological space X is *relatively compact* in X if the closure
 171 of L in X is compact. Relatively compact subsets of $B_1(X)$ (for X Polish) have
 172 been objects of interest for researchers in Analysis and Topological Dynamics. We

begin with the following well-known result. Recall that a set $A \subseteq \mathbb{R}^X$ of real-valued functions is *pointwise bounded* if for every $x \in X$ there is $M_x > 0$ such that $|f(x)| < M_x$ for all $f \in A$. We include a proof for the reader's convenience:

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

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

Proof. Since A is pointwise bounded, for each $x \in X$, fix $M_x > 0$ such that $|f(x)| \leq M_x$ for every $f \in A$.

(i) \Rightarrow (ii) holds in general.

(ii) \Rightarrow (iii) Assume that A is relatively countably compact in $B_1(X)$ and that $f \in \overline{A} \setminus B_1(X)$. By Fact 1.2, there are countable $D_0, D_1 \subseteq X$ with $\overline{D_0} = \overline{D_1}$, and $a < b$ such that $D_0 \subseteq f^{-1}(-\infty, a]$ and $D_1 \subseteq f^{-1}[b, \infty)$. We claim that there is a sequence $\{f_n\}_{n \in \mathbb{N}} \subseteq A$ such that $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for all $x \in D_0 \cup D_1$. Indeed, use the countability to enumerate $D_0 \cup D_1$ as $\{x_n\}_{n \in \mathbb{N}}$. Then for each positive n find $f_n \in A$ with $|f_n(x_i) - f(x_i)| < \frac{1}{n}$ for all $i \leq n$. The claim follows.

By relative countable compactness of A , there is an accumulation point $g \in B_1(X)$ of $\{f_n\}_{n \in \mathbb{N}}$. It is straightforward to show that since f and g agree on $D_0 \cup D_1$, g does not have a point of continuity on the closed set $\overline{D_0} = \overline{D_1}$, which contradicts Fact 1.2.

(iii) \Rightarrow (i) Suppose that $\overline{A} \subseteq B_1(X)$. Then $\overline{A} \cap B_1(X) = \overline{A}$ is a closed subset of $\prod_{x \in X} [-M_x, M_x]$; Tychonoff's theorem states that the product of compact spaces is always compact, and since closed subsets of compact spaces are compact, \overline{A} must be compact, as desired. \square

1.1. From Rosenthal's dichotomy to the Bourgain-Fremlin-Talagrand dichotomy to Shelah's NIP. In metrizable spaces, points of closure of some subset can always be approximated by points inside the set, via a convergent sequence. For more complicated spaces, such as $C_p(X)$, this fails in remarkable ways. To see an example, consider the Cantor space $X = 2^\mathbb{N}$, and for each $n \in \mathbb{N}$ define $p_n : X \rightarrow \{0, 1\}$ by $p_n(x) = x(n)$ for each $x \in X$. Then p_n is continuous for each n , but one can show (see Chapter 1.1 of [Tod97] for details) that the only continuous functions in the closure of $\{p_n\}_{n \in \mathbb{N}}$ are the functions p_n themselves; moreover, none of the subsequences of $\{p_n\}_{n \in \mathbb{N}}$ converge. In some sense, this example is the worst possible scenario for convergence. The topological space obtained from this closure is well-known: it is the *Stone-Čech compactification* of the discrete space of natural numbers, or $\beta\mathbb{N}$ for short, and it is an important object of study in general topology.

The following theorem, established by Haskell Rosenthal in 1974, is fundamental in functional analysis, and describes a sharp division in the behavior of sequences within a Banach space:

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

218 In other words, a pointwise bounded set of continuous functions either contains
 219 a convergent subsequence, or a subsequence whose closure is essentially the same as
 220 the example mentioned in the previous paragraphs (the “wildest” possible scenario).
 221 Note that in the preceding example, the functions are trivially pointwise bounded
 222 in \mathbb{R}^X as the functions can only take values 0 and 1.

223 The genesis of Theorem 1.4 was Rosenthal’s ℓ_1 theorem, which states that the
 224 only reason why Banach space can fail to have an isomorphic copy of ℓ_1 (the space
 225 of absolutely summable sequences) is the presence of a bounded sequence with no
 226 weakly Cauchy subsequence. The theorem is famous for connecting diverse areas
 227 of mathematics, namely, Banach space geometry, Ramsey theory, set theory, and
 228 topology of function spaces.

229 As we move from $C_p(X)$ to the larger space $B_1(X)$, we find a similar dichotomy.
 230 Either every point of closure of the set of functions will be a Baire class 1 function,
 231 or there is a sequence inside the set that behaves in the wildest possible way. The
 232 theorem is usually not phrased as a dichotomy but rather as an equivalence:

233 **Theorem 1.5** (The BFT Dichotomy. Bourgain-Fremlin-Talagrand, [BFT78, The-
 234 orem 4G]). *Let X be a Polish space and $A \subseteq C_p(X)$ be pointwise bounded. The
 235 following are equivalent:*

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

$$\bigcap_{n \in I} f_n^{-1}(-\infty, a] \cap \bigcap_{n \notin I} f_n^{-1}[b, \infty) = \emptyset.$$

237 **Definition 1.6.** We shall say that a set $A \subseteq \mathbb{R}^X$ satisfies the *Independence Prop-
 238 erty*, or IP for short, if it satisfies the following condition: There exists every
 239 $\{f_n\}_{n \in \mathbb{N}} \subseteq A$ and $a < b$ such that for every pair of disjoint sets $E, F \subseteq \mathbb{N}$, we
 240 have

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

241 If A satisfies the negation of this condition, we will say that A *satisfies NIP*, or
 242 that has the NIP.

Remark 1.7. Note that if X is compact and $A \subseteq C_p(X)$, then A satisfies the NIP if and only if for every $\{f_n\}_{n \in \mathbb{N}} \subseteq A$ and for every $a < b$ there is $I \subseteq \mathbb{N}$ such that

$$\bigcap_{n \in I} f_n^{-1}(-\infty, a] \cap \bigcap_{n \notin I} f_n^{-1}[b, \infty) = \emptyset.$$

243 To summarize, the particular case of Theorem 1.8 for X compact can be stated
 244 in the following way:

245 **Theorem 1.8.** *Let X be a compact Polish space. Then, for every pointwise bounded
 246 $A \subseteq C_p(X)$, one and exactly one of the following two conditions must hold:*

- 247 (i) $\overline{A} \subseteq B_1(X)$.
 248 (ii) A has NIP.

249 The Independence Property was first isolated by Saharon Shelah in model theory
 250 as a dividing line between theories whose models are “tame” (corresponding to
 251 NIP) theories of models are “wild” (corresponding to IP). See [She71, Definition
 252 4.1], [She90]. We will discuss this dividing line in more detail in the next section.

253 **1.2. NIP as universal dividing line between polynomial and exponential**
 254 **complexity.** The particular case of the BFT Dichotomy (Theorem 1.8) when A
 255 consists of $\{0, 1\}$ -valued (i.e., {Yes, No}-valued) strings was discovered indepen-
 256 dently, around 1971-1972 in many foundational contexts related to polynomial
 257 (“tame”) vs exponential (“wild”) complexity: In model theory, by Saharon She-
 258 lah [She71],[She90], in combinatorics, by Norbert Sauer [Sau72], and Shelah [She72,
 259 She90], and in statistical learning, by Vladimir Vapnik and Alexey Chervonenkis [VC71,
 260 VC74].

261 **In model theory:** Shelah’s classification theory is a foundational program
 262 in mathematical logic devised to categorize first-order theories based on
 263 the complexity and structure of their models. A theory T is considered
 264 classifiable in Shelah’s sense if the number of non-isomorphic models of T
 265 of a given cardinality can be described by a bounded number of numerical
 266 invariants. In contrast, a theory T is unclassifiable if the number of models
 267 of T of a given cardinality is the maximum possible number. The number
 268 of models of T is directly impacted by the number of “types” over of pa-
 269 rameters in models of T ; a controlled number of types is a characteristic of
 270 a classifiable theory.

271 In Shelah’s classification program [She90], theories without the indepen-
 272 dence property (called NIP theories, or dependent theories) have a well-
 273 behaved, “tame” structure; the number of types over a set of parameters
 274 of size κ of such a theory is of polynomially or similar “slow” growth on κ .
 275 In contrast, Theories with the Independence Property (called IP theories)
 276 are considered “intractable” or “wild”. A theory with the Independence
 277 Property produces the maximum possible number of types over a set of
 278 parameters; for a set of parameters of cardinality κ , the theory will have
 279 2^{2^κ} -many distinct types.

280 **In combinatorics:** Sauer [Sau72] and Shelah [She72] proved the following:
 281 If $\mathcal{F} = \{S_0, S_1, \dots\}$ is a family of subsets of some infinite set S , then
 282 either for every $n \in \mathbb{N}$, there is either a set $A \subseteq S$ with $|A| = n$ such that
 283 $|\{S_i \cap A) : i \in \mathbb{N}\}| = 2^n$ (yielding exponential complexity), or there exists
 284 $N \in \mathbb{N}$ such that for every $A \subseteq S$ with $|A| \geq N$, one has

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

285 (yielding polynomial complexity). This answered a question of Erdős.
 286 **In machine learning:** Readers familiar with statistical learning may rec-
 287 ognize the Sauer-Shelah lemma as the dichotomy discovered and proved
 288 slightly earlier (1971) by Vapnik and Chervonenkis [VC71, VC74] to ad-
 289 dress the problem of uniform convergence in statistics. The least integer
 290 N given by the preceding paragraph, when it exists, is called the *VC-*
 291 *dimension* of \mathcal{F} . This is a core concept in machine learning. If such an
 292 integer N does not exist, we say that the VC-dimension of \mathcal{F} is infinite. The
 293 lemma provides upper bounds on the number of data points (sample size m)
 294 needed to learn a concept class with VC dimension $d \in \mathbb{N}$ by showing this
 295 number grows polynomially with m and d (namely, $\sum_{i=0}^d \binom{m}{i} \approx O(m^d)$),
 296 not exponentially. The Fundamental Theorem of Statistical Learning states

297 that a hypothesis class is PAC-learnable (PAC stands for “Probably Ap-
298 proximately Correct”) if and only if its VC dimension is finite.

299 **1.3. Rosenthal compacta.** The comprehensiveness of Theorem 1.8, attested by
300 the examples outlined in the preceding section, led to the following definition (iso-
301 lated by Gilles Godefroy [God80]):

302 **Definition 1.9.** A Rosenthal compactum is a compact Hausdorff topological space
303 K that can be topologically embedded as a compact subset into the space of all
304 functions of the first Baire class on some Polish space X , equipped with the topology
305 of pointwise convergence.

306 Rosenthal compacta are characterized by significant topological and dynamical
307 tameness properties. They play an important role in functional analysis, measure
308 theory, dynamical systems, descriptive set theory, and model theory. In this paper,
309 we introduce their applicability in deep computation. For this, we shall first focus
310 on countable languages, which is the theme of the next subsection.

311 **1.4. The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with \mathcal{P} countable.** Our goal now is to charac-
312 terize relatively compact subsets of $B_1(X, Y)$ for the particular case when $Y = \mathbb{R}^{\mathcal{P}}$
313 with \mathcal{P} countable. Given $P \in \mathcal{P}$ we denote the projection map onto the P -coordinate
314 by $\pi_P : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}$. From a high-level topological interpretation, the next lemma
315 states that, in this context, the spaces \mathbb{R} and $\mathbb{R}^{\mathcal{P}}$ are really not that different,
316 and that if we understand the Baire class 1 functions of one space, then we also
317 understand the functions of both.

318 **Lemma 1.10.** *Let X be a Polish space and \mathcal{P} be a countable set. Then, $f \in$
319 $B_1(X, \mathbb{R}^{\mathcal{P}})$ if and only if $\pi_P \circ f \in B_1(X)$ for all $P \in \mathcal{P}$.*

Proof. Only one implication needs a proof. Suppose that $\pi_P \circ f \in B_1(X)$ for all
 $P \in \mathcal{P}$. Let V be a basic open subset of $\mathbb{R}^{\mathcal{P}}$. That is, there exists a finite $\mathcal{P}' \subseteq \mathcal{P}$
such that $V = \bigcap_{P \in \mathcal{P}'} \pi_P^{-1}[U_P]$ where U_P is open in \mathbb{R} . Then,

$$f^{-1}[V] = \bigcap_{P \in \mathcal{P}'} (\pi_P \circ f)^{-1}[U_P]$$

320 is an F_{σ} set. Since \mathcal{P} is countable, $\mathbb{R}^{\mathcal{P}}$ is second countable so every open set U in
321 $\mathbb{R}^{\mathcal{P}}$ is a countable union of basic open sets. Hence, $f^{-1}[U]$ is F_{σ} . \square

322 Below we consider \mathcal{P} with the discrete topology. For each $f : X \rightarrow \mathbb{R}^{\mathcal{P}}$ denote
323 $\hat{f}(P, x) := \pi_P \circ f(x)$ for all $(P, x) \in \mathcal{P} \times X$. Similarly, for each $g : \mathcal{P} \times X \rightarrow \mathbb{R}$ denote
324 $\check{g}(x)(P) := g(P, x)$. Given $A \subseteq (\mathbb{R}^{\mathcal{P}})^X$, we denote \hat{A} as the set of all \hat{f} such that
325 $f \in A$. Note that the map $(\mathbb{R}^{\mathcal{P}})^X \rightarrow \mathbb{R}^{\mathcal{P} \times X}$ given by $f \mapsto \hat{f}$ is a homeomorphism
326 and its inverse is given by $g \mapsto \check{g}$.

327 **Lemma 1.11.** *Let X be a Polish space and \mathcal{P} be countable. Then, $f \in B_1(X, \mathbb{R}^{\mathcal{P}})$
328 if and only if $\hat{f} \in B_1(\mathcal{P} \times X)$.*

Proof. (\Rightarrow) By Lemma 1.10, given an open set of reals U , we have $f^{-1}[\pi_P^{-1}[U]]$ is
 F_{σ} for every $P \in \mathcal{P}$. Given that \mathcal{P} is a discrete countable space, we observe that

$$\hat{f}^{-1}[U] = \bigcup_{P \in \mathcal{P}} (\{P\} \times f^{-1}[\pi_P^{-1}[U]])$$

329 is an F_{σ} as well.

(\Leftarrow) By lemma 1.10 it suffices to show that $\pi_P \circ f \in B_1(X)$ for all $P \in \mathcal{P}$. Fix an open $U \subseteq \mathbb{R}$. Write $\hat{f}^{-1}[U] = \bigcup_{n \in \mathbb{N}} F_n$ where F_n is closed in $\mathcal{P} \times X$. Then,

$$(\pi_P \circ f)^{-1}[U] = \bigcup_{n \in \mathbb{N}} \{x \in X : (P, x) \in F_n\}$$

330 which is F_σ . □

331 Given $A \subseteq Y^X$ and $K \subseteq X$ we write $A|_K := \{f|_K : f \in A\}$, i.e., the set of
332 all restrictions of functions in A to K . The following Theorem is a slightly more
333 general version of Theorem 1.8.

334 **Theorem 1.12.** *Assume that \mathcal{P} is countable, X is a Polish space, and $A \subseteq$
335 $C_p(X, \mathbb{R}^\mathcal{P})$ is such that $\pi_P \circ A$ is pointwise bounded for all $P \in \mathcal{P}$. The follow-
336 ing are equivalent for every compact $K \subseteq X$:*

- 337 (1) $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^\mathcal{P})$.
338 (2) $\pi_P \circ A|_K$ satisfies the NIP for every $P \in \mathcal{P}$.

Proof. (1) \Rightarrow (2). Let $P \in \mathcal{P}$. Fix $\{f_n\}_{n \in \mathbb{N}} \subseteq A$ and $a < b$. By (1), we have $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^\mathcal{P})$. Applying the homeomorphism $f \mapsto \hat{f}$ and using lemma 1.11 we get $\hat{A}|_{\mathcal{P} \times K} \subseteq B_1(\mathcal{P} \times K)$. By Theorem 1.8, there is $I \subseteq \mathbb{N}$ such that

$$(\mathcal{P} \times K) \cap \bigcap_{n \in I} \hat{f}_n^{-1}(-\infty, a] \cap \bigcap_{n \notin I} \hat{f}_n^{-1}[b, \infty) = \emptyset$$

Hence,

$$K \cap \bigcap_{n \in I} (\pi_P \circ f_n)^{-1}(-\infty, a] \cap \bigcap_{n \notin I} (\pi_P \circ f_n)^{-1}[b, \infty) = \emptyset$$

By the compactness of K , there are finite $E \subseteq I$ and $F \subseteq \mathbb{N} \setminus I$ such that

$$K \cap \bigcap_{n \in E} (\pi_P \circ f_n)^{-1}(-\infty, a] \cap \bigcap_{n \in F} (\pi_P \circ f_n)^{-1}[b, \infty) = \emptyset$$

339 Thus, $\pi_P \circ A|_L$ satisfies the NIP.

340 (2) \Rightarrow (1) Fix $f \in \overline{A|_K}$. By lemma 1.10 it suffices to show that $\pi_P \circ f \in B_1(K)$
341 for all $P \in \mathcal{P}$. By (2), $\pi_P \circ A|_K$ satisfies the NIP. Hence, by Theorem 1.8 we have
342 $\pi_P \circ A|_K \subseteq B_1(K)$. But then $\pi_P \circ f \in \overline{\pi_P \circ A|_K} \subseteq B_1(K)$. □

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

346 **Lemma 1.13.** *Assume that X is Hausdorff and that $A \subseteq C_p(X)$. The following
347 are equivalent for every $L \subseteq X$:*

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

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

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

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

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

By definition of closure:

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

350 This contradicts (i). □

351 2. COMPOSITIONAL COMPUTATION STRUCTURES.

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

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

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

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

359 Intuitively, L is the set of states of a computation, and the predicates $P \in \mathcal{P}$
360 are primitives that are given and accepted as computational. We think of each
361 state $v \in L$ as being uniquely characterized by its type $\text{tp}(v)$. Thus, in practice,
362 we identify L with a subset of $\mathbb{R}^{\mathcal{P}}$. A typical case will be when $L = \mathbb{R}^{\mathbb{N}}$ or $L = \mathbb{R}^n$
363 for some positive integer n and there is a predicate $P_i(v) = v_i$ for each of the
364 coordinates v_i of v . We regard the space of types as a topological space, endowed
365 with the topology of pointwise convergence inherited from $\mathbb{R}^{\mathcal{P}}$. In particular, for
366 each $P \in \mathcal{P}$, the projection map $v \mapsto P(v)$ is continuous.

367 **Definition 2.1.** Given a computation states structure (L, \mathcal{P}) , any element of $\mathbb{R}^{\mathcal{P}}$
368 in the image of L under the map $v \mapsto \text{tp}(v)$ will be called a *realized type*. The
369 topological closure of the set of realized types in $\mathbb{R}^{\mathcal{P}}$ (endowed with the point-
370 wise convergence topology) will be called the *space of types* of (L, \mathcal{P}) , denoted \mathcal{L} .
371 Elements of $\mathcal{L} \setminus L$ will be called *unrealized types*.

372 In traditional model theory, the space of types of a structure is viewed as a sort
373 of compactification of the structure, and the compactness of type spaces plays a
374 central role. However, the space \mathcal{L} defined above is not necessarily compact. To
375 bypass this obstacle, we follow the idea introduced in [ADIW24] of covering \mathcal{L} by
376 “thin” compact subspaces that we call *shards*. The formal definition of shard is
377 next.

378 **Definition 2.2.** A *sizer* is a tuple $r_{\bullet} = (r_P)_{P \in \mathcal{P}}$ of positive real numbers indexed
379 by \mathcal{P} . Given a sizer r_{\bullet} , we define the r_{\bullet} -*shard* as:

$$L[r_{\bullet}] = L \cap \prod_{P \in \mathcal{P}} [-r_P, r_P].$$

380 For a sizer r_{\bullet} , the r_{\bullet} -*type shard* is defined as $\mathcal{L}[r_{\bullet}] = \overline{L[r_{\bullet}]}$. We define \mathcal{L}_{sh} , as
381 the union of all type-shards.

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

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

386 The elements of the semigroup Γ are called the *computations* of the structure
 387 (L, \mathcal{P}, Γ) .

388 If $\Delta \subseteq \Gamma$, we say that $\Delta \subseteq \Gamma$ is *R-confined* if $\gamma|_{L[r_\bullet]} : L[r_\bullet] \rightarrow L[r_\bullet]$ for every
 389 $r_\bullet \in R$ and $\gamma \in \Delta$. Elements in $\overline{\Delta} \subseteq \mathcal{L}_{\text{sh}}$ are called (real-valued) *deep computations*
 390 or *ultracomputations*.

391 A tenet of our approach is that a map $f : L \rightarrow \mathcal{L}$ is to be considered “effectively
 392 computable” if, for each $Q \in \mathcal{P}$, the output feature $Q \circ f : L \rightarrow \mathbb{R}$ is a *definable*
 393 predicate in the following sense:

394 Given any arbitrary $\varepsilon > 0$ and any $K \subseteq L$ wherein every input feature $P(v)$
 395 remains bounded in magnitude there is an ε -approximating continuous “algebraic”
 396 operator $\varphi(P_1, \dots, P_n)$ of finitely many input predicates $P_1, \dots, P_n \in \mathcal{P}$, such that
 397 the following holds: for all $v \in K$, the output feature $Q(f(v))$ is ε -approximated
 398 by $\varphi(P_1(v), \dots, P_n(v))$. By “algebraic”, we mean that, aside from the primitives
 399 P_1, \dots, P_n , the approximating operator $\varphi(P_1, \dots, P_n)$ uses only the algebra operations
 400 of $\mathbb{R}^\mathcal{P}$, i.e., vector addition, vector multiplication, and scalar addition.

401 It is shown in [ADIW24]) that:

- 402 (1) For a definable $f : L \rightarrow \mathcal{L}$, the approximating operators φ may be taken to
 403 be *polynomials* of the input features, and
 404 (2) Definable transforms $f : L \rightarrow \mathcal{L}$ are precisely those that extend to contin-
 405 uous $\tilde{f} : \mathcal{L} \rightarrow \mathcal{L}$ (this is the property of *extendibility* mentioned above).

406 This motivates the following definition.

407 **Definition 2.4.** We say that a CCS (L, \mathcal{P}, Γ) satisfies the *Extendability Axiom* if
 408 for all $\gamma \in \Gamma$, there is $\tilde{\gamma} : \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$ such that for every sizer r_\bullet there is a sizer s_\bullet
 409 such that $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$ is continuous. We refer to $\tilde{\gamma}$ as a *free extension*
 410 of γ .

411 By the preceding remarks, the Extendability Axiom says that the elements of
 412 the semigroup Γ are definable. For the rest of the paper, fix for each $\gamma \in \Gamma$ a free
 413 extension $\tilde{\gamma}$ of γ . For any $\Delta \subseteq \Gamma$, let $\tilde{\Delta}$ denote $\{\tilde{\gamma} : \gamma \in \Delta\}$.

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

415 For an illustrative example, we can frame Newton’s polynomial root approxima-
 416 tion method in the context of a CCS (see Example 5.6 of [ADIW24] for details) as
 417 follows. Begin by considering the extended complex numbers $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ with
 418 the usual Riemann sphere topology that makes it into a compact space (where
 419 unbounded sequences converge to ∞). In fact, not only is this space compact
 420 but it is covered by the shard given by the sizer $(1, 1, 1)$ (the unit sphere is con-
 421 tained in the cube $[-1, 1]^3$). The space $\hat{\mathbb{C}}$ is homeomorphic to the usual unit sphere
 422 $S^2 := \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ of \mathbb{R}^3 , by means of the stereographic pro-
 423 jection and its inverse $\hat{\mathbb{C}} \rightarrow S^2$. This function is regarded as a triple of predi-
 424 cates $x, y, z : \hat{\mathbb{C}} \rightarrow [-1, 1]$ where each will map an extended complex number to
 425 its corresponding real coordinate on the cube $[-1, 1]^3$. Now fix the cubic com-
 426 plex polynomial $p(s) := s^3 - 1$, and consider the map which performs one step
 427 in Newton’s method at a particular (extended) complex number s , for finding a

428 root of p , $\gamma_p : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$. The explicit inner workings of γ_p are irrelevant for this
 429 example, except for the fact that it is a continuous mapping. It follows that
 430 $(S^2, \{x, y, z\}, \{\gamma_p^k : k \in \mathbb{N}\})$ is a CCS. The idea is that repeated applications of
 431 $\gamma_p(s), \gamma_p \circ \gamma_p(s), \gamma_p \circ \gamma_p \circ \gamma_p(s), \dots$ would approximate a root of p provided s was a
 432 good enough initial guess.

433 **3. CLASSIFYING DEEP COMPUTATIONS**

434 **3.1. NIP, Rosenthal compacta, and deep computations.** Under what conditions
 435 are deep computations Baire class 1, and thus well-behaved according to our
 436 framework, on type-shards? The following theorem says that, under the assumption
 437 that \mathcal{P} is countable, the space of deep computations is a Rosenthal compactum
 438 (when restricted to shards) if and only if the set of computations satisfies the NIP
 439 feature by feature. Hence, we can import the theory of Rosenthal compacta into
 440 this framework of deep computations.

441 **Theorem 3.1.** *Let (L, \mathcal{P}, Γ) be a compositional computational structure (Definition 2.3) satisfying the Extendability Axiom (Definition 2.4) with \mathcal{P} countable. Let
 442 R be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be R -confined. The following
 443 are equivalent.*

- 444 (1) $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ for all $r_\bullet \in R$.
 445 (2) $\pi_P \circ \Delta|_{L[r_\bullet]}$ 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

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

446 Moreover, if any (hence all) of the preceding conditions hold, then every deep
 447 computation $f \in \overline{\Delta}$ can be extended to a Baire-1 function on shards, i.e., there is
 448 $\tilde{f} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$ 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
 449 each shard every deep computation is the pointwise limit of a countable sequence of
 450 computations.

451 *Proof.* Since \mathcal{P} is countable, $\mathcal{L}[r_\bullet] \subseteq \mathbb{R}^\mathcal{P}$ is Polish. Also, the Extendability Axiom
 452 implies that $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$ is a pointwise bounded set of continuous functions for all
 453 $P \in \mathcal{P}$. Hence, Theorem 1.12 and Lemma 1.13 prove the equivalence of (1) and (2).
 454 If (1) holds and $f \in \overline{\Delta}$, then write $f = \text{Ulim}_i \gamma_i$ as an ultralimit. Define $\tilde{f} := \text{Ulim}_i \tilde{\gamma}_i$.
 455 Hence, for all $r_\bullet \in R$ we have $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$. That every
 456 deep computation is a pointwise limit of a countable sequence of computations
 457 follows from the fact that for a Polish space X every compact subset of $B_1(X)$ is
 458 Fréchet-Urysohn (that is, a space where topological closures coincide with sequential
 459 closures, see Theorem 3F in [BFT78] or Theorem 4.1 in [Deb13]). \square

460 **3.2. The Todorčević trichotomy and levels of PAC learnability.** Given a
 461 countable set Δ of computations satisfying the NIP on features and shards (con-
 462 dition (2) of Theorem 3.1) we have that $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ (for a fixed sizer r_\bullet) is a separable
 463 *Rosenthal compactum* (see Definition 1.9). Todorčević proved a trichotomy for
 464 Rosenthal compacta [Tod99], and later Argyros, Dodos, Kanellopoulos [ADK08]
 465 proved an heptachotomy that refined Todorčević's classification. In this section,

466 inspired by the work of Glasner and Megrelishvili [GM22], we study ways in which
 467 this classification allows us obtain different levels of PAC-learnability and NIP

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

473 **Definition 3.2.** Let (L, \mathcal{P}, Γ) be a CCS satisfying the Extendability Axiom and R
 474 be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be an R -confined countable set of
 475 computations satisfying the NIP on shards and features (condition (2) in Theorem
 476 3.1). We say that Δ is:

- 477 (i) NIP₁ if $\overline{\Delta|_{\mathcal{L}[r_\bullet]}}$ is first countable for every $r_\bullet \in R$.
- 478 (ii) NIP₂ if $\overline{\Delta|_{\mathcal{L}[r_\bullet]}}$ is hereditarily separable for every $r_\bullet \in R$.
- 479 (iii) NIP₃ if $\overline{\Delta|_{\mathcal{L}[r_\bullet]}}$ is metrizable for every $r_\bullet \in R$.

480 Observe that NIP₃ \Rightarrow NIP₂ \Rightarrow NIP₁ \Rightarrow NIP. A natural question that would
 481 continue this work is to find examples of CCS that separate these levels of NIP. In
 482 [Tod99], Todorčević isolates three canonical examples of Rosenthal compacta that
 483 witness the failure of the converse implications above.

484 We now present some separable and non-separable examples of Rosenthal com-
 485 pacta:

486 Examples 3.3.

- 487 (1) *Alexandroff compactification of a discrete space of size continuum.* For
 488 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
 489 $\delta_a(x) = 0$ otherwise. Let $A(2^\mathbb{N}) = \{\delta_a : a \in 2^\mathbb{N}\} \cup \{0\}$, where 0 is the zero
 490 map. Notice that $A(2^\mathbb{N})$ is a compact subset of $B_1(2^\mathbb{N})$, in fact $\{\delta_a : a \in 2^\mathbb{N}\}$
 491 is a discrete subspace of $B_1(2^\mathbb{N})$ and its pointwise closure is precisely $A(2^\mathbb{N})$.
 492 Hence, this is a Rosenthal compactum which is not first countable. Notice
 493 that this space is also not separable.
- 494 (2) *Extended Alexandroff compactification.* For each finite binary sequence $s \in$
 495 $2^{<\mathbb{N}}$, let $v_s : 2^\mathbb{N} \rightarrow \mathbb{R}$ be given by $v_s(x) = 1$ if x extends s and $v_s(x) = 0$
 496 otherwise. Let $\hat{A}(2^\mathbb{N})$ be the pointwise closure of $\{v_s : s \in 2^{<\mathbb{N}}\}$, i.e.,
 497 $\hat{A}(2^\mathbb{N}) = A(2^\mathbb{N}) \cup \{v_s : s \in 2^{<\mathbb{N}}\}$. Note that this space is a separable
 498 Rosenthal compactum which is not first countable.
- 499 (3) *Split Cantor.* Let $<$ be the lexicographic order in the space of infinite
 500 binary sequences, i.e., $2^\mathbb{N}$. For each $a \in 2^\mathbb{N}$ let $f_a^- : 2^\mathbb{N} \rightarrow \mathbb{R}$ be given by
 501 $f_a^-(x) = 1$ if $x < a$ and $f_a^-(x) = 0$ otherwise. Let $f_a^+ : 2^\mathbb{N} \rightarrow \mathbb{R}$ be given
 502 by $f_a^+(x) = 1$ if $x \leq a$ and $f_a^+(x) = 0$ otherwise. The split Cantor is the
 503 space $S(2^\mathbb{N}) = \{f_a^- : a \in 2^\mathbb{N}\} \cup \{f_a^+ : a \in 2^\mathbb{N}\}$. This is a separable Rosenthal
 504 compactum. One example of a countable dense subset is the set of all f_a^+
 505 and f_a^- where a is an infinite binary sequence that is eventually constant.
 506 Moreover, it is hereditarily separable but it is not metrizable.
- 507 (4) *Alexandroff Duplicate.* Let K be any compact metric space and consider
 508 the Polish space $X = C(K) \sqcup K$, i.e., the disjoint union of $C(K)$ (with its
 509 supremum norm topology) and K . For each $a \in K$ define $g_a^0, g_a^1 : X \rightarrow \mathbb{R}$

as follows:

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

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

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

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

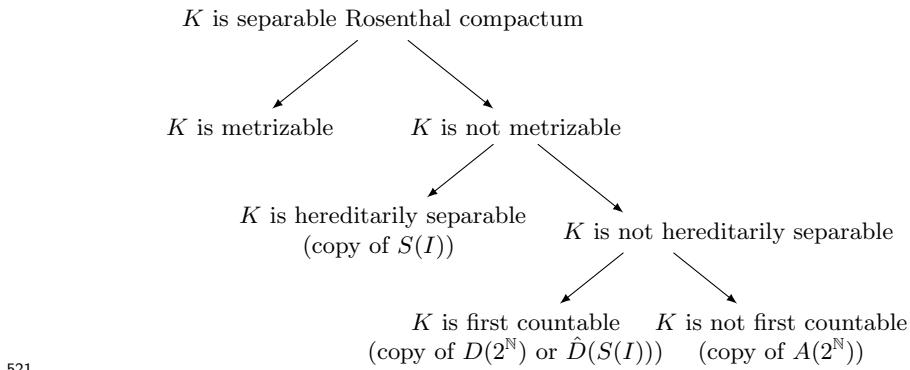
$$h_t(x) = \begin{cases} 0, & x < a_t \\ 1/2, & a_t \leq x \leq b_t \\ 1, & b_t < x \end{cases}$$

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

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

- 516 (i) If K is hereditarily separable but non-metrizable, then $S(2^{\mathbb{N}})$ embeds into K .
 517 (ii) If K is first countable but not hereditarily separable, then either $D(2^{\mathbb{N}})$ or
 518 $\hat{D}(S(2^{\mathbb{N}}))$ embeds into K .
 519 (iii) If K is not first countable, then $A(2^{\mathbb{N}})$ embeds into K .

520 We thus have the following classification:



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

524 Question 3.5. Is there a non-topological characterization for NIP_i , $i = 1, 2, 3$?

525 **3.3. The Argyros-Dodos-Kanellopoulos heptachotomy, and approxima-
 526 bility of deep computation by minimal classes.** In the three separable three
 527 cases given in 3.3, namely, $(\hat{A}(2^{\mathbb{N}}), S(2^{\mathbb{N}}) \text{ and } \hat{D}(S(2^{\mathbb{N}})))$, the countable dense sub-
 528 sets are indexed by the binary tree $2^{<\mathbb{N}}$. This choice of index is useful for two
 529 reasons:

- 530 (1) Our emphasis is computational. Elements of $2^{<\mathbb{N}}$ represent finite bitstrings,
 i.e., standard computations, while Rosenthal compacta represent deep com-
 putations, i.e., limits of finite computations. Mathematically, deep computa-
 tions are pointwise limits of standard computations; however, computa-
 tionally, we are interested in the manner (and the efficiency) in which the
 approximations can occur.
- 536 (2) The Ramsey theory of perfect subsets of the Cantor space $2^{\mathbb{N}}$ can be im-
 ported to analyze the behavior of the accumulation points. Since $2^{<\mathbb{N}}$ is
 countable, we can always choose this index for the countable dense subsets.
 This is done in [ADK08].

540 **Definition 3.6.** Let X be a Polish space.

- 541 (1) If I is a countable and $\{f_i : i \in I\} \subseteq \mathbb{R}^X$, $\{g_i : i \in I\} \subseteq \mathbb{R}^X$ are two
 pointwise families by I , we say that $\{f_i : i \in I\}$ and $\{g_i : i \in I\}$ are
 equivalent if and only if the map $f_i \mapsto g_i$ is extended to a homeomorphism
 from $\overline{\{f_i : i \in I\}}$ to $\overline{\{g_i : i \in I\}}$.
- 545 (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 $\{f_t : t \in 2^{<\mathbb{N}}\}$.

548 One of the main results in [ADK08] is that, up to equivalence, there are seven
 549 minimal families of Rosenthal compacta and that for every relatively compact $\{f_t :
 550 t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$ there is a dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ such that $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$
 551 is equivalent to one of the minimal families. We shall describe the minimal families
 552 next. We follow the same notation as in [ADK08]. For any node $t \in 2^{<\mathbb{N}}$, let us
 553 denote by $t^\frown 0^\infty$ ($t^\frown 1^\infty$) the infinite binary sequence starting with t and continuing
 554 will all 0's (respectively, all 1's). Fix a regular dyadic subtree $R = \{s_t : t \in 2^{<\mathbb{N}}\}$
 555 of $2^{<\mathbb{N}}$ (i.e., a dyadic subtree such that every level of R is contained in a level of
 556 $2^{<\mathbb{N}}$) with the property that for all $s, s' \in R$, $s^\frown 0^\infty \neq s' \frown 0^\infty$ and $s^\frown 1^\infty \neq s' \frown 1^\infty$.
 557 Given $t \in 2^{<\mathbb{N}}$, let v_t be the characteristic function of the set $\{x \in 2^{\mathbb{N}} : x \text{ extends } t\}$.
 558 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
 559 characteristic function of $\{x \in 2^{\mathbb{N}} : a \leq x\}$ and let $f_a^- : 2^{\mathbb{N}} \rightarrow \{0, 1\}$ be the
 560 characteristic function of $\{x \in 2^{\mathbb{N}} : a < x\}$. Given two maps $f, g : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ we
 561 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
 562 g on the second copy of $2^{\mathbb{N}}$.

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

570 **Theorem 3.7** (Heptacotomy of minimal families, Theorem 2 in [ADK08]). *Let*
 571 *X be Polish. For every relatively compact $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$, there exists*
 572 *i = 1, 2, \dots, 7* *and a regular dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$ such that $\{f_{s_t} :$*
 573 *t \in 2^{<\mathbb{N}}\}* *is equivalent to D_i . Moreover, all D_i are minimal and mutually non-*
 574 *equivalent.*

575 **3.4. Bourgain-Fremlin-Talagrand, NIP, and essential computability of**
 576 **deep computations.** We now turn to the question: what happens when \mathcal{P} is
 577 uncountable? Notice that the countability assumption is crucial in the proof of
 578 Theorem 1.12 essentially because it makes $\mathbb{R}^{\mathcal{P}}$ a Polish space. For the uncountable
 579 case, we may lose Baire-1 definability so we shall replace $B_1(X)$ by a larger class.
 580 Recall that the *raison d'être* of the class of Baire-1 functions is to have a class that
 581 is contains the continuous functions but is closed under pointwise limits, and that (

582 Fact 1.2) for perfectly normal X , a function f is in $B_1(X, Y)$ if and only if $f^{-1}[U]$
 583 is an F_{σ} subset of X for every open $U \subseteq Y$. This motivates the following definition:

584 **Definition 3.8.** Given a Hausdorff space X and a measurable space (Y, Σ) , we say
 585 that $f : X \rightarrow Y$ is *universally measurable* (with respect to Σ) if $f^{-1}(E)$ is Borel
 586 for every $E \in \Sigma$, i.e., $f^{-1}(E)$ is μ -measurable for every Radon probability measure
 587 μ on X . When $Y = \mathbb{R}$ we will always take $\Sigma = \mathcal{B}(\mathbb{R})$, the Borel σ -algebra of \mathbb{R} .
 588 In this case, a function $f : X \rightarrow \mathbb{R}$ is universally measurable if and only if $f^{-1}(U)$
 589 is μ -measurable for every Radon probability measure μ on X and every open set
 590 $U \subseteq \mathbb{R}$.

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

597 Following [BFT78], the collection of all universally measurable real-valued func-
 598 tions will be denoted by $M_r(X)$. In the context of deep computations, we will be
 599 interested in transition maps from a state space $L \subseteq \mathbb{R}^{\mathcal{P}}$ to itself. There are two
 600 natural σ -algebras one can consider in the product space $\mathbb{R}^{\mathcal{P}}$: the Borel σ -algebra,
 601 i.e., the σ -algebra generated by open sets in $\mathbb{R}^{\mathcal{P}}$, and the cylinder σ -algebra, i.e.,
 602 the σ -algebra generated by the sub-basic open sets in $\mathbb{R}^{\mathcal{P}}$. Note that when \mathcal{P} is
 603 countable, both σ -algebras coincide but in general the cylinder σ -algebra is strictly
 604 smaller. We will use the cylinder σ -algebra to define universally measurable maps
 605 $f : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}^{\mathcal{P}}$. The reason for this choice is the following characterization:

606 **Lemma 3.9.** *Let X be a Hausdorff space and $Y = \prod_{i \in I} Y_i$ be any product of*
 607 *measurable spaces (Y_i, Σ_i) for $i \in I$. Let Σ_Y be the cylinder σ -algebra generated by*
 608 *the measurable spaces (Y_i, Σ_i) . Let $f : X \rightarrow Y$. The following are equivalent:*

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

611 *Proof.* (i) \Rightarrow (ii) is clear since the projection maps π_i are measurable and the com-
 612 position of measurable functions is measurable. To prove (ii) \Rightarrow (i), suppose that
 613 $C = \prod_{i \in I} C_i$ is a measurable cylinder and let J be the finite set of $i \in I$ such that
 614 $C_i \neq Y_i$. 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 a universally
 615 measurable set by assumption. \square

616 The preceding lemma says that a transition map is universally measurable if and
 617 only if it is universally measurable on all its features. In other words, we can check
 618 measurability of a transition just by checking measurability feature by feature. We
 619 will denote by $M_r(X, \mathbb{R}^{\mathcal{P}})$ the collection of all universally measurable functions
 620 $f : X \rightarrow \mathbb{R}^{\mathcal{P}}$ (with respect to the cylinder σ -algebra), endowed with the topology
 621 of pointwise convergence.

622 We now wish to define the concept of a deep computation being computable
 623 except a set of arbitrarily small measure “no matter which reasonable way you try
 624 to measure things on its domain” (see the remarks following definition). This is
 625 definition below. To motivate the definition, we need to recall two facts:

- 626 (1) Littlewoood’s second principle states that every Lebesgue measurable func-
 627 tion is “nearly continuous”. The formal version of this, which is Luzin’s
 628 theorem, states that if (X, Σ, μ) a Radon measure space and Y be a second-
 629 countable topological space (e.g., $Y = \mathbb{R}^{\mathcal{P}}$ with \mathcal{P} countable) equipped with
 630 a Borel algebra, then any given $f : X \rightarrow Y$ is measurable if and only if for
 631 every $E \in \Sigma$ and every $\varepsilon > 0$ there exists a closed $F \subseteq E$ such that the
 632 restriction $f|F$ is continuous.
- 633 (2) Computability of deep computations can is characterized in terms of con-
 634 tinuous extendibility of computations. This is at the core of [ADIW24].

635 These facts motivate the following definition:

636 **Definition 3.10.** Let (L, \mathcal{P}, Γ) be a CCS. We say that a transition $f : L \rightarrow L$
 637 is *universally essentially computable* if and only if there exists $\tilde{f} : \mathcal{L}_{\text{sh}} \rightarrow \mathcal{L}_{\text{sh}}$
 638 extending f such that for every sizer r_{\bullet} there is a sizer s_{\bullet} such that the restriction
 639 $\tilde{f}|_{\mathcal{L}[r_{\bullet}]} : \mathcal{L}[r_{\bullet}] \rightarrow \mathcal{L}[s_{\bullet}]$ is universally measurable, i.e., $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_{\bullet}]} : \mathcal{L}[r_{\bullet}] \rightarrow [-s_P, s_P]$
 640 is μ -measurable for every Radon probability measure μ on $\mathcal{L}[r_{\bullet}]$.

641 For a measure μ on aX , the set of all μ -measurable functions will denoted by
 642 $\mathcal{M}^0(X, \mu)$.

643 We will need the following result about NIP and universally measurable func-
 644 tions:

645 **Theorem 3.11** (Bourgain-Fremlin-Ta set lagrand, Theorem 2F in [BFT78]). *Let*
 646 X be a Hausdorff space and $A \subseteq C(X)$ be pointwise bounded. *The following are*
 647 *equivalent:*

- 648 (i) $\overline{A} \subseteq M_r(X)$.
- 649 (ii) For every compact $K \subseteq X$, $A|_K$ satisfies the NIP.
- 650 (iii) For every Radon measure μ on X , A is relatively countably compact in
 651 $\mathcal{M}^0(X, \mu)$, i.e., every countable subset of A has an accumulation point in
 652 $\mathcal{M}^0(X, \mu)$.

653 Theorem 1.8 immediately yields the following.

654 **Theorem 3.12.** *Let (L, \mathcal{P}, Γ) be a CCS satisfying the Extendability Axiom. Let*
 655 *R be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be R -confined. If $\pi_P \circ \Delta|_{L[r_{\bullet}]}$
 656 *satisfies the NIP for all $P \in \mathcal{P}$ and all $r_{\bullet} \in R$, then every deep computation is*
 657 *universally essentially computable.**

658 *Proof.* By the Extendability Axiom, Theorem 1.8 and lemma 1.13 we have that
 659 $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_{\bullet}]} \subseteq M_r(\mathcal{L}[r_{\bullet}])$ 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, for all $r_\bullet \in R$ and $P \in \mathcal{P}$ $\pi_P \circ \tilde{\gamma}_i|_{\mathcal{L}[r_\bullet]} \in M_r(\mathcal{L}[r_\bullet])$ for all i , so $\pi_P \circ f|_{\mathcal{L}[r_\bullet]} \in \pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq M_r(\mathcal{L}[r_\bullet])$. \square

Question 3.13. Under the same assumptions of the preceding theorem, suppose that every deep computation of Δ is universally essentially computable. Must $\pi_P \circ \Delta|_{L[r_\bullet]}$ have the NIP for all $P \in \mathcal{P}$ and all $r_\bullet \in R$?

3.5. Talagrand stability, NIP, and essential 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 $k \geq 1$ such that

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

where μ^* denotes the outer measure (we work with outer since the sets $D_k(A, E, a, b)$ need not be μ -measurable). This is certainly the case when A is a countable set of continuous (or μ -measurable) functions.

The following lemma establishes that Talagrand stability is a way to ensure that deep computations are definable by measurable functions. We include a proof for the reader's convenience.

Lemma 3.14. *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, observe that a subset of a μ -stable set is μ -stable. To show that \overline{A} is μ -stable, observe that $D_k(\overline{A}, E, a, b) \subseteq D_k(A, E, a', b')$ where $a < a' < b' < b$ and E is a μ -measurable set with positive measure. 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 a characterization of measurable functions (see 413G in [Fre03]), 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, but we argued before that a subset of a μ -stable set must be μ -stable. \square

We say that A is *universally Talagrand stable* if A is Talagrand μ -stable for every Radon probability measure μ on X . An argument similar to the proof of 3.11, yields the following:

Theorem 3.15. *Let (L, \mathcal{P}, Γ) be a CCS satisfying the Extendability 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 essentially computable.*

It is then natural to ask: what is the relationship between Talagrand stability and the NIP? The following dichotomy will be useful.

699 **Lemma 3.16** (Fremlin's Dichotomy, 463K in [Fre03]). *If (X, Σ, μ) is a perfect
700 σ -finite measure space (in particular, for X compact and μ a Radon probability
701 measure on X) and $\{f_n : n \in \mathbb{N}\}$ be a sequence of real-valued measurable functions
702 on X , then either*

- 703 (i) $\{f_n : n \in \mathbb{N}\}$ has a subsequence that converges μ -almost everywhere, or
704 (ii) $\{f_n : n \in \mathbb{N}\}$ has a subsequence with no μ -measurable accumulation point
705 in \mathbb{R}^X .

706 The preceding lemma can be considered as a measure-theoretic version of Rosen-
707 thal's Dichotomy. Combining this dichotomy with the Theorem 3.11 we get the
708 following result:

709 **Theorem 3.17.** *Let X be a Hausdorff space and $A \subseteq C(X)$ be pointwise bounded.
710 The following are equivalent:*

- 711 (i) $\overline{A} \subseteq M_r(X)$.
712 (ii) For every compact $K \subseteq X$, $A|_K$ satisfies the NIP.
713 (iii) For every Radon measure μ on X , A is relatively countably compact in
714 $\mathcal{M}^0(X, \mu)$, i.e., every countable subset of A has an accumulation point in
715 $\mathcal{M}^0(X, \mu)$.
716 (iv) For every Radon measure μ on X and every sequence $\{f_n : n \in \mathbb{N}\} \subseteq A$,
717 there is a subsequence that converges μ -almost everywhere.

718 *Proof.* Notice that the equivalence (i)-(iii) is Theorem 3.11. Notice that the equiv-
719 alence of (iii) and (iv) is Fremlin's Dichotomy (Theorem 3.16). \square

720 Finally, it is natural to ask what the connection is between Talagrand stability
721 and NIP.

722 **Proposition 3.18.** *Let X be a compact Hausdorff space and $A \subseteq C(X)$ be point-
723 wise bounded. If A is universally Talagrand stable, then A satisfies the NIP.*

724 *Proof.* By Theorem 3.11, it suffices to show that A is relatively countably compact
725 in $\mathcal{M}^0(X, \mu)$ for all Radon probability measure μ on X . Since A is Talagrand
726 μ -stable for any such μ , we have $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$. In particular, A is relatively
727 countably compact in $\mathcal{M}^0(X, \mu)$. \square

728 **Question 3.19.** Is the converse true?

729 The following two results suggest that the precise connection between Talagrand
730 stability and NIP may be sensitive to set-theoretic axioms (even assuming count-
731 ability of A).

732 **Theorem 3.20** (Talagrand, Theorem 9-3-1(a) in [Tal84]). *Let X be a compact
733 Hausdorff space and $A \subseteq M_r(X)$ be countable and pointwise bounded. Assume that
734 $[0, 1]$ is not the union of $< \mathfrak{c}$ closed measure zero sets. If A satisfies the NIP, then
735 A is universally Talagrand stable.*

736 **Theorem 3.21** (Fremlin, Shelah, [FS93]). *It is consistent that there exists a count-
737 able pointwise bounded set of Lebesgue measurable functions with the NIP which is
738 not Talagrand stable with respect to Lebesgue measure.*

739

REFERENCES

- 740 [ADIW24] Samson Alva, Eduardo Dueñez, Jose Iovino, and Claire Walton. Approximability of
 741 deep equilibria. *arXiv preprint arXiv:2409.06064*, 2024. Last revised 20 May 2025,
 742 version 3.
- 743 [ADK08] Spiros A. Argyros, Pandelis Dodos, and Vassilis Kanellopoulos. A classification of sepa-
 744 rable Rosenthal compacta and its applications. *Dissertationes Mathematicae*, 449:1–55,
 745 2008.
- 746 [Ark91] A. V. Arkhangel'skii. *Topological Function Spaces*. Springer, New York, 1st edition,
 747 1991.
- 748 [BD19] Shai Ben-David. Understanding machine learning through the lens of model theory.
 749 *The Bulletin of Symbolic Logic*, 25(3):319–340, 2019.
- 750 [BFT78] J. Bourgain, D. H. Fremlin, and M. Talagrand. Pointwise compact sets of Baire-
 751 measurable functions. *American Journal of Mathematics*, 100(4):845–886, 1978.
- 752 [BKK] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. Preprint.
 753 <https://arxiv.org/abs/1909.01377>, <https://implicit-layers-tutorial.org/>.
- 754 [CRBD] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural
 755 ordinary differential equations. Preprint. <https://arxiv.org/abs/1806.07366>.
- 756 [Deb13] Gabriel Debs. Descriptive aspects of Rosenthal compacta. In *Recent Progress in Gen-
 757 eral Topology III*, pages 205–227. Springer, 2013.
- 758 [Fre00] David H. Fremlin. *Measure Theory, Volume 1: The Irreducible Minimum*. Torres
 759 Fremlin, Colchester, UK, 2000. Second edition 2011.
- 760 [Fre01] David H. Fremlin. *Measure Theory, Volume 2: Broad Foundations*. Torres Fremlin,
 761 Colchester, UK, 2001.
- 762 [Fre03] David H. Fremlin. *Measure Theory, Volume 4: Topological Measure Spaces*. Torres
 763 Fremlin, Colchester, UK, 2003.
- 764 [FS93] David H. Fremlin and Saharon Shelah. Pointwise compact and stable sets of measurable
 765 functions. *Journal of Symbolic Logic*, 58(2):435–455, 1993.
- 766 [GM22] Eli Glasner and Michael Megrelishvili. Todorčević' trichotomy and a hierarchy in the
 767 class of tame dynamical systems. *Transactions of the American Mathematical Society*,
 768 375(7):4513–4548, 2022.
- 769 [God80] Gilles Godefroy. Compacts de Rosenthal. *Pacific J. Math.*, 91(2):293–306, 1980.
- 770 [Gro52] A. Grothendieck. Critères de compacité dans les espaces fonctionnels généraux. *Amer.
 771 J. Math.*, 74:168–186, 1952.
- 772 [HT23] Clovis Hamel and Franklin D. Tall. C_p -theory for model theorists. In Jose Iovino, editor,
 773 *Beyond First Order Model Theory, Volume II*, chapter 5, pages 176–213. Chapman
 774 and Hall/CRC, 2023.
- 775 [Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts
 776 in Mathematics*. Springer-Verlag, 1995.
- 777 [Kha20] Karim Khanaki. Stability, nipp, and nsop; model theoretic properties of formulas via
 778 topological properties of function spaces. *Mathematical Logic Quarterly*, 66(2):136–
 779 149, 2020.
- 780 [Pap02] E. Pap, editor. *Handbook of measure theory. Vol. I, II*. North-Holland, Amsterdam,
 781 2002.
- 782 [Ros74] Haskell P. Rosenthal. A characterization of Banach spaces containing l^1 . *Proc. Nat.
 783 Acad. Sci. U.S.A.*, 71:2411–2413, 1974.
- 784 [RPK19] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: a
 785 deep learning framework for solving forward and inverse problems involving nonlinear
 786 partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.
- 787 [Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–
 788 147, 1972.
- 789 [She71] Saharon Shelah. Stability, the f.c.p., and superstability; model theoretic properties of
 790 formulas in first order theory. *Ann. Math. Logic*, 3(3):271–362, 1971.
- 791 [She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories
 792 in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.
- 793 [She78] Saharon Shelah. *Unstable Theories*, volume 187 of *Studies in Logic and the Foun-
 794 dations of Mathematics*. North-Holland, 1978.

- 795 [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.
- 796 [Sim15a] Pierre Simon. *A Guide to NIP Theories*, volume 44 of *Lecture Notes in Logic*. Cambridge University Press, 2015.
- 797 [Sim15b] Pierre Simon. Rosenthal compacta and NIP formulas. *Fundamenta Mathematicae*, 231(1):81–92, 2015.
- 798 [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.
- 799 [Tal87] Michel Talagrand. The Glivenko-Cantelli problem. *The Annals of Probability*, 15(3):837–870, 1987.
- 800 [Tka11] Vladimir V. Tkachuk. *A Cp-Theory Problem Book: Topological and Function Spaces*. Problem Books in Mathematics. Springer, 2011.
- 801 [Tod97] Stevo Todorcevic. *Topics in Topology*, volume 1652 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 1997.
- 802 [Tod99] Stevo Todorcevic. Compact subsets of the first Baire class. *Journal of the American Mathematical Society*, 12(4):1179–1212, 1999.
- 803 [Val84] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.
- 804 [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.
- 805 [VC74] Vladimir N Vapnik and Alexey Ya Chervonenkis. *Theory of Pattern Recognition*. Nauka, Moscow, 1974. German Translation: Theorie der Zeichenerkennung, Akademie-Verlag, Berlin, 1979.
- 806 [WYP22] Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why PINNs fail to train: a neural tangent kernel perspective. *J. Comput. Phys.*, 449:Paper No. 110768, 28, 2022.