

# COMPLEXITY OF DEEP COMPUTATIONS VIA TOPOLOGY OF FUNCTION SPACES

EDUARDO DUEÑEZ<sup>1</sup> JOSÉ IOVINO<sup>1</sup> TONATIUH MATOS-WIEDERHOLD<sup>2</sup>  
LUCIANO SALVETTI<sup>2</sup> FRANKLIN D. TALL<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Texas at San Antonio  
<sup>2</sup>Department of Mathematics, University of Toronto

**ABSTRACT.** We use topological methods to study complexity of deep computations and limit 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. We use the theory of Rosenthal compacta to characterize approximability of deep computations, both deterministically and probabilistically.

## 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

Date: December 30, 2025.

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

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

<sup>32</sup> topology, the classification of Rosenthal compacta pioneered by Todorčević [Tod99];  
<sup>33</sup> in logic, the classification of theories developed by Shelah [She90]; and in statistical  
<sup>34</sup> learning, the notion PAC learning and VC dimension pioneered by Vapkins and  
<sup>35</sup> Chervonenkis [VC74, VC71].

<sup>36</sup> In a previous paper [ADIW24], we introduced the concept of limits of computations,  
<sup>37</sup> which we called *ultracomputations* (given they arise as ultrafilter limits of  
<sup>38</sup> standard computations) and *deep computations* (following usage in machine learn-  
<sup>39</sup> ing [BKK]). There is a technical difference between both designations, but in this  
<sup>40</sup> paper, to simplify the nomenclature, we will ignore the difference and use only the  
<sup>41</sup> term “deep computation”.

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

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

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

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

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

<sup>75</sup> Going beyond Todorčević’s trichotomy, we invoke a more recent heptachotomy  
<sup>76</sup> for Rosenthal compacta obtained by Argyros, Dodos and Kanellopoulos [ADK08].  
<sup>77</sup> Argyros, Dodos and Kanellopoulos identified seven fundamental “prototypes” of  
<sup>78</sup> separable Rosenthal compacta, and proved that any non-metrizable separable Rosen-  
<sup>79</sup> 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 then 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 In section 1, we present the basic topological and combinatorial preliminaries,  
 90 and in section 2, we introduce the structural/model-theoretic viewpoint (no previ-  
 91 ous exposure to model theory is needed). Section 3 is devoted to the classification  
 92 of deep computations, and the final section, section 4, presents the probabilistic  
 93 viewpoint.

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

98

## CONTENTS

99      Introduction	1
100     1. General topological preliminaries: From continuity to Baire class 1	4
101        1.1. From Rosenthal’s dichotomy to the Bourgain-Fremlin-Talagrand	
102            dichotomy to Shelah’s NIP	5
103        1.2. NIP as universal dividing line between polynomial and exponential	
104            complexity	7
105        1.3. Rosenthal compacta	8
106        1.4. The special case $B_1(X, \mathbb{R}^{\mathcal{P}})$ with $\mathcal{P}$ countable.	8
107     2. Compositional computation structures: A structural approach to	
108            floating-point computation	10
109     3. Classifying deep computations	12
110        3.1. NIP, Rosenthal compacta, and deep computations	12
111        3.2. The Todorčević trichotomy and levels of PAC learnability	13
112        3.3. The Argyros-Dodos-Kanellopoulos heptachotomy, and approximability	
113            of deep computation by minimal classes	15
114     4. Measure-theoretic versions of NIP and universal Monte Carlo	
115            computability of deep computations	16
116        4.1. A measure-theoretic version of NIP	16
117        4.2. Universal Monte Carlo computability of deep computations	17
118        4.3. Bourgain-Fremlin-Talagrand, NIP, and universal Monte Carlo	
119            computability of deep computations	18
120        4.4. Talagrand stability, Fremlin’s dichotomy, NIP, and universal Monte	
121            Carlo computability of deep computations	18
122        References	20

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

Recall that a subset of a topological space is  $F_\sigma$  if it is a countable union of closed sets, and  $G_\delta$  if it is a countable intersection of closed sets. Note that in a metrizable space, every open set is  $F_\sigma$ ; equivalently, every closed set is  $G_\delta$ .

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

137 In this paper, we shall often discuss subspaces, and so there is a pertinent subtlety  
 138 of the definitions worth mentioning: *completely metrizable space* is not the same as  
 139 *complete metric space*; for an illustrative example, the interval  $(0, 1)$  with the metric  
 140 inherited from the reals is not complete, but it is Polish since it is homeomorphic to  
 141 the real line. Being Polish is a topological property while being metrically complete  
 142 is not.

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

**Fact 1.1.** A subset  $A$  of a Polish space  $X$  is itself Polish in the subspace topology if and only if it is a  $G_\delta$  set. In particular, closed subsets and open subsets of Polish spaces are also Polish spaces.

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. When  $Y = \mathbb{R}$ , we denote this collection simply as  $C_p(X)$ . A natural question is, how do topological properties of  $X$  translate into  $C_p(X)$  and vice versa? These questions, and in general the study of these spaces, are the concern of  $C_p$ -theory, an active field of research in general topology which was pioneered by A. V. Arhangel'skii and his students in the 1970's and 1980's. This field has found many applications in model theory and functional analysis. Recent surveys on the topics include [HT23] and [Tka11].

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

168 A topological space  $X$  is *perfectly normal* if it is normal and every closed subset  
 169 of  $X$  is a  $G_\delta$  (equivalently, every open subset of  $X$  is a  $G_\delta$ ). Note that every  
 170 metrizable space is perfectly normal.

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

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

- 175 •  *$f$  is a Baire class 1 function, that is,  $f$  is a pointwise limit of continuous  
 176 functions..*
- 177 •  *$f^{-1}[U]$  is an  $F_\sigma$  subset of  $X$  whenever  $U \subseteq Y$  is open.*
- 178 • *For every closed  $F \subseteq X$ , the restriction  $f|_F$  has a point of continuity.*

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

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

181 A subset  $L$  of a topological space  $X$  is *relatively compact* in  $X$  if the closure  
 182 of  $L$  in  $X$  is compact. Relatively compact subsets of  $B_1(X)$  (for  $X$  Polish) have  
 183 been objects of interest for researchers in Analysis and Topological Dynamics. We  
 184 begin with the following well-known result. Recall that a set  $A \subseteq \mathbb{R}^X$  of real-  
 185 valued functions is *pointwise bounded* if for every  $x \in X$  there is  $M_x > 0$  such that  
 186  $|f(x)| < M_x$  for all  $f \in A$ . We include a proof for the reader's convenience:

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

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

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

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

196 (ii) $\Rightarrow$ (iii) Assume that  $A$  is relatively countably compact in  $B_1(X)$  and that  
 197  $f \in \overline{A} \setminus B_1(X)$ . By Fact 1.2, there are countable  $D_0, D_1 \subseteq X$  with  $\overline{D_0} = \overline{D_1}$ , and  
 198  $a < b$  such that  $D_0 \subseteq f^{-1}(-\infty, a]$  and  $D_1 \subseteq f^{-1}[b, \infty)$ . We claim that there is a  
 199 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,  
 200 use the countability to enumerate  $D_0 \cup D_1$  as  $\{x_n\}_{n \in \mathbb{N}}$ . Then for each positive  $n$   
 201 find  $f_n \in A$  with  $|f_n(x_i) - f(x_i)| < \frac{1}{n}$  for all  $i \leq n$ . The claim follows.

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

206 (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  
 207  $\prod_{x \in X} [-M_x, M_x]$ ; Tychonoff's theorem states that the product of compact spaces  
 208 is always compact, and since closed subsets of compact spaces are compact,  $\overline{A}$  must  
 209 be compact, as desired.  $\square$

210 **1.1. From Rosenthal's dichotomy to the Bourgain-Fremlin-Talagrand di-  
 211 chotomy 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 in a Banach spaces. I

**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  $\beta\mathbb{N}$ .*

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

The genesis of Theorem 1.4 was Rosenthal’s  $\ell_1$  theorem, which states that the only reason why Banach space can fail to have an isomorphic copy of  $\ell_1$  (the space of absolutely summable sequences) is the presence of a bounded sequence with no weakly Cauchy subsequence. The theorem is famous for connecting diverse areas of mathematics, namely, Banach space geometry, Ramsey theory, set theory, and topology of function spaces.

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

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

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

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

248 **Definition 1.6.** We shall say that a set  $A \subseteq \mathbb{R}^X$  satisfies the *Independence Property*, or IP for short, if it satisfies the following condition: There exists every 250  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and  $a < b$  such that for every pair of disjoint sets  $E, F \subseteq \mathbb{N}$ , we 251 have

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

252 If  $A$  satisfies the negation of this condition, we will say that  $A$  *satisfies NIP*, or 253 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

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

254 To summarize, the particular case of Theorem 1.5 for  $X$  compact can be stated  
 255 in the following way:

256 **Theorem 1.8.** *Let  $X$  be a compact Polish space. Then, for every pointwise bounded  
 257  $A \subseteq C_p(X)$ , one and exactly one of the following two conditions must hold:*  
 258 (i)  $\overline{A} \subseteq B_1(X)$ .  
 259 (ii)  $A$  has NIP.

260 The Independence Property was first isolated by Saharon Shelah in model theory  
 261 as a dividing line between theories whose models are “tame” (corresponding to  
 262 NIP) theories of models are “wild” (corresponding to IP). See [She71, Definition  
 263 4.1],[She90]. We will discuss this dividing line in more detail in the next section.

264 **1.2. NIP as universal dividing line between polynomial and exponential  
 265 complexity.** The particular case of the BFT dichotomy (Theorem 1.5) when  $A$   
 266 consists of  $\{0, 1\}$ -valued (i.e., {Yes, No}-valued) strings was discovered indepen-  
 267 dently, around 1971-1972 in many foundational contexts related to polynomial  
 268 (“tame”) vs exponential (“wild”) complexity: In model theory, by Saharon She-  
 269 lah [She71],[She90], in combinatorics, by Norbert Sauer [Sau72], and Shelah [She72,  
 270 She90], and in statistical learning, by Vladimir Vapnik and Alexey Chervonenkis [VC71,  
 271 VC74].

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

282 In Shelah’s classification program [She90], theories without the indepen-  
 283 dence property (called NIP theories, or dependent theories) have a well-  
 284 behaved, “tame” structure; the number of types over a set of parameters  
 285 of size  $\kappa$  of such a theory is of polynomially or similar “slow” growth on  $\kappa$ .  
 286 In contrast, Theories with the Independence Property (called IP theories)  
 287 are considered “intractable” or “wild”. A theory with the Independence  
 288 Property produces the maximum possible number of types over a set of  
 289 parameters; for a set of parameters of cardinality  $\kappa$ , the theory will have  
 290  $2^{2^\kappa}$ -many distinct types.

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

295  $N \in \mathbb{N}$  such that for every  $A \subseteq S$  with  $|A| \geq N$ , one has

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

296 (yielding polynomial complexity). This answered a question of Erdős.

297 **In machine learning:** Readers familiar with statistical learning may rec-  
298 ognize the Sauer-Shelah lemma as the dichotomy discovered and proved  
299 slightly earlier (1971) by Vapnik and Chervonenkis [VC71, VC74] to ad-  
300 dress the problem of uniform convergence in statistics. The least integer  
301  $N$  given by the preceding paragraph, when it exists, is called the *VC-*  
302 *dimension* of  $\mathcal{F}$ . This is a core concept in machine learning. If such an  
303 integer  $N$  does not exist, we say that the VC-dimension of  $\mathcal{F}$  is infinite. The  
304 lemma provides upper bounds on the number of data points (sample size  $m$ )  
305 needed to learn a concept class with VC dimension  $d \in \mathbb{N}$  by showing this  
306 number grows polynomially with  $m$  and  $d$  (namely,  $\sum_{i=0}^d \binom{m}{i} \approx O(m^d)$ ),  
307 not exponentially. The Fundamental Theorem of Statistical Learning states  
308 that a hypothesis class is PAC-learnable (PAC stands for “Probably Ap-  
309 proximately Correct”) if and only if its VC dimension is finite.

310 **1.3. Rosenthal compacta.** The comprehensiveness of Theorem 1.5, attested by  
311 the examples outlined in the preceding section, led to the following definition (iso-  
312 lated by Gilles Godefroy [God80]):

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

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

322 **1.4. The special case  $B_1(X, \mathbb{R}^{\mathcal{P}}$  with  $\mathcal{P}$  countable.** Our goal now is to charac-  
323 terize relatively compact subsets of  $B_1(X, Y)$  for the particular case when  $Y = \mathbb{R}^{\mathcal{P}}$   
324 with  $\mathcal{P}$  countable. Given  $P \in \mathcal{P}$  we denote the projection map onto the  $P$ -coordinate  
325 by  $\pi_P : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}$ . From a high-level topological interpretation, the next lemma  
326 states that, in this context, the spaces  $\mathbb{R}$  and  $\mathbb{R}^{\mathcal{P}}$  are really not that different,  
327 and that if we understand the Baire class 1 functions of one space, then we also  
328 understand the functions of both.

329 **Lemma 1.10.** Let  $X$  be a Polish space and  $\mathcal{P}$  be a countable set. Then,  $f \in$   
330  $B_1(X, \mathbb{R}^{\mathcal{P}})$  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,

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

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

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

**Lemma 1.11.** *Let  $X$  be a Polish space and  $\mathcal{P}$  be countable. Then,  $f \in B_1(X, \mathbb{R}^\mathcal{P})$  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  $F_\sigma$  for every  $P \in \mathcal{P}$ . Given that  $\mathcal{P}$  is a discrete countable space, we observe that

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

is an  $F_\sigma$  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,

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

which is  $F_\sigma$ .  $\square$

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 1.5.

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

- 348 (i)  $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^\mathcal{P})$ .
- 349 (ii)  $\pi_P \circ A|_K$  satisfies the NIP for every  $P \in \mathcal{P}$ .

*Proof.* (i) $\Rightarrow$ (ii). Let  $P \in \mathcal{P}$ . Fix  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and  $a < b$ . By (i), 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 1.5, 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

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

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

(ii) $\Rightarrow$ (i) Fix  $f \in \overline{A|_K}$ . By lemma 1.10 it suffices to show that  $\pi_P \circ f \in B_1(K)$  for all  $P \in \mathcal{P}$ . By (ii),  $\pi_P \circ A|_K$  satisfies the NIP. Hence, by Theorem 1.5 we have  $\overline{\pi_P \circ A|_K} \subseteq B_1(K)$ . But then,  $\pi_P \circ f \in \overline{\pi_P \circ A|_K} \subseteq B_1(K)$ .  $\square$

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

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

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

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

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

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

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

By definition of closure:

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

361 This contradicts (i).

## 2. COMPOSITIONAL COMPUTATION STRUCTURES: A STRUCTURAL APPROACH TO FLOATING-POINT COMPUTATION

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

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

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

<sup>369</sup> For each  $P \in \mathcal{P}$ , we call real value  $P(v)$  the  $P$ -th feature of  $v$ . A transition of a  
<sup>370</sup> computation states structure  $(L, \mathcal{P})$  is a map  $f : L \rightarrow L$ .

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

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

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

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

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

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

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

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

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

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  $r_\bullet \in R$  and  $\gamma \in \Delta$ . Elements in  $\overline{\Delta} \subseteq \mathcal{L}_{\text{sh}}$  are called (real-valued) *deep computations* or *ultracomputations*.

A tenet of our approach is that a map  $f : L \rightarrow \mathcal{L}$  is to be considered “effectively computable” if, for each  $Q \in \mathcal{P}$ , the output feature  $Q \circ f : L \rightarrow \mathbb{R}$  is a *definable* predicate in the following sense: Given any arbitrary  $\varepsilon > 0$  and any  $K \subseteq L$  wherein every input feature  $P(v)$  remains bounded in magnitude there is an  $\varepsilon$ -approximating continuous “algebraic” operator  $\varphi(P_1, \dots, P_n)$  of finitely many input predicates  $P_1, \dots, P_n \in \mathcal{P}$ , such that the following holds: for all  $v \in K$ , the output feature  $Q(f(v))$  is  $\varepsilon$ -approximated by  $\varphi(P_1(v), \dots, P_n(v))$ . By “algebraic”, we mean that, *aside from the primitively computable  $P_1, \dots, P_n$ , the approximating operator  $\varphi(P_1, \dots, P_n)$  uses only the also primitively computable operations of  $\mathbb{R}^{\mathcal{P}}$*  as an algebra, i.e., vector addition, vector multiplication, and scalar addition.

It is shown in [ADIW24]) that:

- (1) For a definable  $f : L \rightarrow \mathcal{L}$ , the approximating operators  $\varphi$  may be taken to be *polynomials* of the input features, and
- (2) Definable transforms  $f : L \rightarrow \mathcal{L}$  are precisely those that extend to continuous  $\tilde{f} : \mathcal{L} \rightarrow \mathcal{L}$

This motivates the following definition.

**Definition 2.4.** We say that a CCS  $(L, \mathcal{P}, \Gamma)$  satisfies the *Extendability Axiom* if 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  $s_\bullet$  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* of  $\gamma$ .

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

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

427 For an illustrative example, we can frame Newton's polynomial root approximation  
 428 method in the context of a CCS (see Example 5.6 of [ADIW24] for details) as  
 429 follows. Begin by considering the extended complex numbers  $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$  with  
 430 the usual Riemann sphere topology that makes it into a compact space (where  
 431 unbounded sequences converge to  $\infty$ ). In fact, not only is this space compact,  
 432 but it is covered by the shard given by the sizer  $(1, 1, 1)$  (the unit sphere is con-  
 433 tained in the cube  $[-1, 1]^3$ ). The space  $\hat{\mathbb{C}}$  is homeomorphic to the usual unit sphere  
 434  $S^2 := \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$  of  $\mathbb{R}^3$ , by means of the stereographic pro-  
 435 jection and its inverse  $\hat{\mathbb{C}} \rightarrow S^2$ . This function is regarded as a triple of predi-  
 436 cates  $x, y, z : \hat{\mathbb{C}} \rightarrow [-1, 1]$  where each will map an extended complex number to  
 437 its corresponding real coordinate on the cube  $[-1, 1]^3$ . Now fix the cubic com-  
 438 plex polynomial  $p(s) := s^3 - 1$ , and consider the map which performs one step  
 439 in Newton's method at a particular (extended) complex number  $s$ , for finding a  
 440 root of  $p$ ,  $\gamma_p : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ . The explicit inner workings of  $\gamma_p$  are irrelevant for this  
 441 example, except for the fact that it is a continuous mapping. It follows that  
 442  $(S^2, \{x, y, z\}, \{\gamma_p^k : k \in \mathbb{N}\})$  is a CCS. The idea is that repeated applications of  
 443  $\gamma_p(s), \gamma_p \circ \gamma_p(s), \gamma_p \circ \gamma_p \circ \gamma_p(s), \dots$  would approximate a root of  $p$  provided  $s$  was a  
 444 good enough initial guess.

### 445 3. CLASSIFYING DEEP COMPUTATIONS

446 **3.1. NIP, Rosenthal compacta, and deep computations.** Under what condi-  
 447 tions are deep computations Baire class 1, and thus well-behaved according to our  
 448 framework, on type-shards? The following theorem says that, under the assump-  
 449 tion that  $\mathcal{P}$  is countable, the space of deep computations is a Rosenthal compactum  
 450 (when restricted to shards) if and only if the set of computations satisfies the NIP  
 451 feature by feature. Hence, we can import the theory of Rosenthal compacta into  
 452 this framework of deep computations.

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

- 457 (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

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

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

463 *Proof.* Since  $\mathcal{P}$  is countable,  $\mathcal{L}[r_\bullet] \subseteq \mathbb{R}^\mathcal{P}$  is Polish. Also, the Extendability Axiom  
 464 implies that  $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$  is a pointwise bounded set of continuous functions for all  
 465  $P \in \mathcal{P}$ . Hence, Theorem 1.12 and Lemma 1.13 prove the equivalence of (i) and (ii).  
 466 If (i) holds and  $f \in \overline{\Delta}$ , then write  $f = \text{Ulim}_i \gamma_i$  as an ultralimit. Define  $\tilde{f} := \text{Ulim}_i \tilde{\gamma}_i$ .

467 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  
468 deep computation is a pointwise limit of a countable sequence of computations  
469 follows from the fact that for a Polish space  $X$  every compact subset of  $B_1(X)$  is  
470 Fréchet-Urysohn (that is, a space where topological closures coincide with sequential  
471 closures, see Theorem 3F in [BFT78] or Theorem 4.1 in [Deb13]).  $\square$

472 **3.2. The Todorčević trichotomy and levels of PAC learnability.** Given a  
473 countable set  $\Delta$  of computations satisfying the NIP on features and shards (con-  
474 dition (ii) of Theorem 3.1) we have that  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  (for a fixed sizer  $r_\bullet$ ) is a separable  
475 Rosenthal compactum (see Definition 1.9). Todorčević proved a remarkable tri-  
476 chotomy for Rosenthal compacta [Tod99], and later Argyros, Dodos, Kanellopou-  
477 los [ADK08] proved an heptachotomy that refined Todorčević's classification. In  
478 this section, inspired by the work of Glasner and Megrelishvili [GM22], we study  
479 ways in which this classification allows us obtain different levels of PAC-learnability  
480 and NIP.

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

486 **Definition 3.2.** Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendability Axiom and  $R$   
487 be an exhaustive collection of sizers. Let  $\Delta \subseteq \Gamma$  be an  $R$ -confined countable set of  
488 computations satisfying the NIP on shards and features (condition (ii) in Theorem  
489 3.1). We say that  $\Delta$  is:

- 490 (i) NIP<sub>1</sub> if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is first countable for every  $r_\bullet \in R$ .
- 491 (ii) NIP<sub>2</sub> if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is hereditarily separable for every  $r_\bullet \in R$ .
- 492 (iii) NIP<sub>3</sub> if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is metrizable for every  $r_\bullet \in R$ .

493 Observe that NIP<sub>3</sub>  $\Rightarrow$  NIP<sub>2</sub>  $\Rightarrow$  NIP<sub>1</sub>  $\Rightarrow$  NIP. A natural question that would con-  
494 tinue this work is to find examples of CCS that separate these levels of NIP. In  
495 [Tod99], Todorčević isolates three canonical examples of Rosenthal compacta that  
496 witness the failure of the converse implications above.

497 We now present some separable and non-separable examples of Rosenthal com-  
498 pacta.

499 **Examples 3.3.**

- 500 (1) *Alexandroff compactification of a discrete space of size continuum.* For  
501 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  
502  $\delta_a(x) = 0$  otherwise. Let  $A(2^\mathbb{N}) = \{\delta_a : a \in 2^\mathbb{N}\} \cup \{0\}$ , where 0 is the zero  
503 map. Notice that  $A(2^\mathbb{N})$  is a compact subset of  $B_1(2^\mathbb{N})$ , in fact  $\{\delta_a : a \in 2^\mathbb{N}\}$   
504 is a discrete subspace of  $B_1(2^\mathbb{N})$  and its pointwise closure is precisely  $A(2^\mathbb{N})$ .  
505 Hence, this is a Rosenthal compactum which is not first countable. Notice  
506 that this space is also not separable.
- 507 (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  $v_s(x) = 0$   
508 otherwise. Let  $\hat{A}(2^\mathbb{N})$  be the pointwise closure of  $\{v_s : s \in 2^{<\mathbb{N}}\}$ , i.e.,  
509  $\hat{A}(2^\mathbb{N}) = A(2^\mathbb{N}) \cup \{v_s : s \in 2^{<\mathbb{N}}\}$ . Note that this space is a separable  
510 Rosenthal compactum which is not first countable.

- 512     (3) *Split Cantor.* Let  $<$  be the lexicographic order in the space of infinite  
 513       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  
 514        $f_a^-(x) = 1$  if  $x < a$  and  $f_a^-(x) = 0$  otherwise. Let  $f_a^+ : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  be given  
 515       by  $f_a^+(x) = 1$  if  $x \leq a$  and  $f_a^+(x) = 0$  otherwise. The split Cantor is the  
 516       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  
 517       compactum. One example of a countable dense subset is the set of all  $f_a^+$   
 518       and  $f_a^-$  where  $a$  is an infinite binary sequence that is eventually constant.  
 519       Moreover, it is hereditarily separable, but it is not metrizable.
- 520     (4) *Alexandroff Duplicate.* Let  $K$  be any compact metric space and consider  
 521       the Polish space  $X = C(K) \sqcup K$ , i.e., the disjoint union of  $C(K)$  (with its  
 522       supremum norm topology) and  $K$ . For each  $a \in K$  define  $g_a^0, g_a^1 : X \rightarrow \mathbb{R}$   
 523       as follows:

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

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

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

- 523     (5) *Extended Alexandroff Duplicate of the split Cantor.* For each finite binary  
 524       sequence  $t \in 2^{<\mathbb{N}}$  let  $a_t \in 2^{\mathbb{N}}$  be the sequence starting with  $t$  and ending  
 525       with 0's and let  $b_t \in 2^{\mathbb{N}}$  be the sequence starting with  $t$  and ending with  
 526       1's. Define  $h_t : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  by

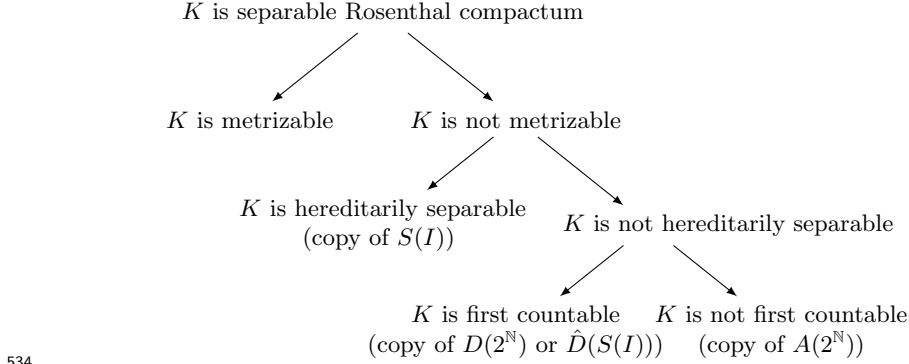
$$h_t(x) = \begin{cases} 0, & x < a_t \\ 1/2, & a_t \leq x \leq b_t \\ 1, & b_t < x \end{cases}$$

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

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

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

533     We thus have the following classification:



535     The definitions provided here for  $\text{NIP}_i$  ( $i = 1, 2, 3$ ) are topological. This raises  
536     the following question:

537     **Question 3.5.** Is there a non-topological characterization for  $\text{NIP}_i$ ,  $i = 1, 2, 3$ ?

538     3.3. **The Argyros-Dodos-Kanellopoulos heptachotomy, and approxima-**  
539     **bility of deep computation by minimal classes.** In the three separable three  
540     cases given in 3.3, namely,  $(\hat{A}(2^N), S(2^N)$  and  $\hat{D}(S(2^N))$ ), the countable dense sub-  
541     sets are indexed by the binary tree  $2^{<\mathbb{N}}$ . This choice of index is useful for two  
542     reasons:

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

553     **Definition 3.6.** Let  $X$  be a Polish space.

- 554        (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  
555           pointwise families by  $I$ , we say that  $\{f_i : i \in I\}$  and  $\{g_i : i \in I\}$  are  
556           *equivalent* if and only if the map  $f_i \mapsto g_i$  is extended to a homeomorphism  
557           from  $\overline{\{f_i : i \in I\}}$  to  $\overline{\{g_i : i \in I\}}$ .
- 558        (2) If  $\{f_t : t \in 2^{<\mathbb{N}}\}$  is a pointwise bounded family, we say that  $\{f_t : t \in 2^{<\mathbb{N}}\}$   
559           is *minimal* if and only if for every dyadic subtree  $\{s_t : t \in 2^{<\mathbb{N}}\}$  of  $2^{<\mathbb{N}}$ ,  
560            $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$  is equivalent to  $\{f_t : t \in 2^{<\mathbb{N}}\}$ .

561     One of the main results in [ADK08] is that, up to equivalence, there are seven  
562     minimal families of Rosenthal compacta and that for every relatively compact  $\{f_t :  
563     t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$  there is a dyadic subtree  $\{s_t : t \in 2^{<\mathbb{N}}\}$  such that  $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$   
564     is equivalent to one of the minimal families. We shall describe the seven minimal  
565     families next. We follow the same notation as in [ADK08]. For any node  $t \in 2^{<\mathbb{N}}$ ,  
566     let us denote by  $t^\frown 0^\infty$  ( $t^\frown 1^\infty$ ) the infinite binary sequence starting with  $t$  and  
567     continuing with all 0's (respectively, all 1's). Fix a regular dyadic subtree  $R = \{s_t :  
568     t \in 2^{<\mathbb{N}}\}$  of  $2^{<\mathbb{N}}$  (i.e., a dyadic subtree such that every level of  $R$  is contained

in a level of  $2^{<\mathbb{N}}$ ) with the property that for all  $s, s' \in R$ ,  $s^\frown 0^\infty \neq s' \frown 0^\infty$  and  $s^\frown 1^\infty \neq s' \frown 1^\infty$ . Given  $t \in 2^{<\mathbb{N}}$ , let  $v_t$  be the characteristic function of the set  $\{x \in 2^\mathbb{N} : x \text{ extends } t\}$ . Let  $<$  be the lexicographic order in  $2^\mathbb{N}$ . Given  $a \in 2^\mathbb{N}$ , let  $f_a^+ : 2^\mathbb{N} \rightarrow \{0, 1\}$  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)  $D_2 = \{s_t^\frown 0^\infty : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_2} = 2^{\leq N}$ .
- (3)  $D_3 = \{f_{s_t^\frown 0^\infty}^+ : t \in 2^{<\mathbb{N}}\}$ . This is a discrete in  $\overline{D_3} = S(2^\mathbb{N})$ .
- (4)  $D_4 = \{f_{s_t^\frown 1^\infty}^- : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_4} = S(2^\mathbb{N})$ .
- (5)  $D_5 = \{v_t : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_5} = \hat{A}(2^\mathbb{N})$ .
- (6)  $D_6 = \{(v_{s_t}, s_t^\frown 0^\infty) : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_6} = \hat{D}(2^\mathbb{N})$ .
- (7)  $D_7 = \{(v_{s_t}, f_{s_t^\frown 0^\infty}^+) : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_7} = \hat{D}(S(2^\mathbb{N}))$

**Theorem 3.7** (Heptacotomy 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, all  $D_i$  are minimal and mutually non-equivalent.*

#### 4. MEASURE-THEORETIC VERSIONS OF NIP AND UNIVERSAL MONTE CARLO COMPUTABILITY OF DEEP COMPUTATIONS

The countability assumption on  $\mathcal{P}$  played a crucial role in the proof of Theorem 1.12, as it makes  $\mathbb{R}^\mathcal{P}$  a Polish space. In this section, we do not assume that  $\mathcal{P}$  is countable. We replace deterministic computability by measure-theoretic ('Monte Carlo') computability.

**4.1. A measure-theoretic version of NIP.** Recall that the *raison d'être* of the class of Baire-1 functions is to have a class that contains the continuous functions but is closed under pointwise limits, and that for perfectly normal  $X$ , a function  $f$  is in  $B_1(X, Y)$  if and only if  $f^{-1}[U]$  is an  $F_\sigma$  subset of  $X$  for every open  $U \subseteq Y$  (see Fact 1.2). This motivates the following definition:

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

**Remark 4.2.** A function  $f : X \rightarrow \mathbb{R}$  is universally measurable if and only if  $f^{-1}(U)$  is  $\mu$ -measurable for every Radon probability measure  $\mu$  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].

612 **Notation 4.3.** Following [BFT78], the collection of all universally measurable real-  
 613 valued functions will be denoted by  $M_r(X)$ .

614 In the context of deep computations, we will be interested in transition maps of  
 615 a state space  $L \subseteq \mathbb{R}^{\mathcal{P}}$  into itself. There are two natural  $\sigma$ -algebras one can consider  
 616 in the product space  $\mathbb{R}^{\mathcal{P}}$ : the Borel  $\sigma$ -algebra, i.e., the  $\sigma$ -algebra generated by open  
 617 sets in  $\mathbb{R}^{\mathcal{P}}$ , and the cylinder  $\sigma$ -algebra, i.e., the  $\sigma$ -algebra generated by the sub-basic  
 618 open sets in  $\mathbb{R}^{\mathcal{P}}$ . Note that when  $\mathcal{P}$  is countable, both  $\sigma$ -algebras coincide, but in  
 619 general the cylinder  $\sigma$ -algebra is strictly smaller. We will use the cylinder  $\sigma$ -algebra  
 620 to define universally measurable maps  $f : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}^{\mathcal{P}}$ . The reason for this choice is  
 621 the following characterization:

622 **Lemma 4.4.** Let  $X$  be a Hausdorff space and  $Y = \prod_{i \in I} Y_i$  be any product of  
 623 measurable spaces  $(Y_i, \Sigma_i)$  for  $i \in I$ . Let  $\Sigma_Y$  be the cylinder  $\sigma$ -algebra generated by  
 624 the measurable spaces  $(Y_i, \Sigma_i)$ . Let  $f : X \rightarrow Y$ . The following are equivalent:

- 625 (i)  $f : X \rightarrow Y$  is universally measurable (with respect to  $\Sigma_Y$ ).
- 626 (ii)  $\pi_i \circ f : X \rightarrow Y_i$  is universally measurable (with respect to  $\Sigma_i$ ) for all  $i \in I$ .

627 *Proof.* (i) $\Rightarrow$ (ii) is clear since the projection maps  $\pi_i$  are measurable and the com-  
 628 position of measurable functions is measurable. To prove (ii) $\Rightarrow$ (i), suppose that  
 629  $C = \prod_{i \in I} C_i$  is a measurable cylinder and let  $J$  be the finite set of  $i \in I$  such that  
 630  $C_i \neq Y_i$ . 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  
 631 measurable set by assumption.  $\square$

632 The preceding lemma says that a transition map is universally measurable if and  
 633 only if it is universally measurable on all its features; in other words, we can check  
 634 measurability of a transition just by checking measurability feature by feature. We  
 635 will denote by  $M_r(X, \mathbb{R}^{\mathcal{P}})$  the collection of all universally measurable functions  
 636  $f : X \rightarrow \mathbb{R}^{\mathcal{P}}$  (with respect to the cylinder  $\sigma$ -algebra), endowed with the topology  
 637 of pointwise convergence.

638 We will need the following result about NIP and universally measurable func-  
 639 tions:

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

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

646 **4.2. Universal Monte Carlo computability of deep computations.** We now  
 647 wish to define the concept of a deep computation being computable except a set  
 648 of arbitrarily small measure “no matter which reasonable way you try to measure  
 649 things on its domain” (see the remarks following definition 4.1). This is the concept  
 650 of *universal Monte Carlo computability* defined below (Definition 4.6). To motivate  
 651 the definition, we need to recall two facts:

- 652 (1) Littlewood's second principle states that every Lebesgue measurable func-  
 653 tion is “nearly continuous”. The formal version of this, which is Luzin's  
 654 theorem, states that if  $(X, \Sigma, \mu)$  a Radon measure space and  $Y$  be a second-  
 655 countable topological space (e.g.,  $Y = \mathbb{R}^{\mathcal{P}}$  with  $\mathcal{P}$  countable) equipped with  
 656 a Borel 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.

(2) Computability of deep computations can be characterized in terms of continuous extendibility of computations. This is at the core of [ADIW24].

These two facts motivate the following definition:

**Definition 4.6.** Let  $(L, \mathcal{P}, \Gamma)$  be a CCS. We say that a transition  $f : L \rightarrow L$  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  $\mu$ -measurable for every Radon probability measure  $\mu$  on  $\mathcal{L}[r_\bullet]$ .

**4.3. Bourgain-Fremlin-Talagrand, NIP, and universal Monte Carlo computability of deep computations.** Theorem 4.5 immediately yields the following.

**Theorem 4.7.** Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendability 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 is universally Monte Carlo computable.

*Proof.* By the Extendability Axiom, Theorem 4.5 and lemma 1.13 we have that  $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq M_r(\mathcal{L}[r_\bullet])$  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  $\Delta$ . 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 4.8.** Under the same assumptions of the preceding theorem, suppose that every deep computation of  $\Delta$  is universally Monte Carlo computable. Must  $\pi_P \circ \Delta|_{L[r_\bullet]}$  have the NIP for all  $P \in \mathcal{P}$  and all  $r_\bullet \in R$ ?

**4.4. Talagrand stability, Fremlin's dichotomy, NIP, and universal 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  $\mu$  be a Radon probability measure on  $X$ . Given a  $\mu$ -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  $\mu$ -stable* if and only if for every  $\mu$ -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  $\mu^*$  denotes the outer measure (we work with outer since the sets  $D_k(A, E, a, b)$  need not be  $\mu$ -measurable). This is certainly the case when  $A$  is a countable set of continuous (or  $\mu$ -measurable) functions.

**Notation 4.9.** For a measure  $\mu$  on a set  $X$ , the set of all  $\mu$ -measurable functions will be denoted by  $\mathcal{M}^0(X, \mu)$ .

695     The following lemma establishes that Talagrand stability is a way to ensure that  
 696     deep computations are definable by measurable functions.

697     **Lemma 4.10.** *If  $A$  is Talagrand  $\mu$ -stable, then  $\overline{A}$  is also Talagrand  $\mu$ -stable and  
 698      $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$ .*

699     *Proof.* First, observe that a subset of a  $\mu$ -stable set is  $\mu$ -stable. To show that  $\overline{A}$  is  
 700      $\mu$ -stable, observe that  $D_k(\overline{A}, E, a, b) \subseteq D_k(A, E, a', b')$  where  $a < a' < b' < b$  and  $E$   
 701     is a  $\mu$ -measurable set with positive measure. It suffices to show that  $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$ .  
 702     Suppose that there exists  $f \in \overline{A}$  such that  $f \notin \mathcal{M}^0(X, \mu)$ . By a characterization  
 703     of measurable functions (see 413G in [Fre03]), there exists a  $\mu$ -measurable set  $E$   
 704     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,  
 705      $\{f\}$  is not  $\mu$ -stable, but we argued before that a subset of a  $\mu$ -stable set must be  
 706      $\mu$ -stable.  $\square$

709     We say that  $A$  is *universally Talagrand stable* if  $A$  is Talagrand  $\mu$ -stable for every  
 710     Radon probability measure  $\mu$  on  $X$ . An argument similar to the proof of 4.5, yields  
 711     the following:

712     **Theorem 4.11.** *Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendability Axiom. If  
 713      $\pi_P \circ \Delta|_{L[r_\bullet]}$  is universally Talagrand stable for all  $P \in \mathcal{P}$  and all sizers  $r_\bullet$ , then  
 714     every deep computation is universally Monte Carlo computable.*

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

717     **Lemma 4.12** (Fremlin's Dichotomy, 463K in [Fre03]). *If  $(X, \Sigma, \mu)$  is a perfect  
 718      $\sigma$ -finite measure space (in particular, for  $X$  compact and  $\mu$  a Radon probability  
 719     measure on  $X$ ) and  $\{f_n : n \in \mathbb{N}\}$  be a sequence of real-valued measurable functions  
 720     on  $X$ , then one (and only one) of the following conditions holds:*

- 721       (i)  $\{f_n : n \in \mathbb{N}\}$  has a subsequence that converges  $\mu$ -almost everywhere,
- 722       (ii)  $\{f_n : n \in \mathbb{N}\}$  has a subsequence with no  $\mu$ -measurable accumulation point in  
 723        $\mathbb{R}^X$ .

724     The preceding lemma can be considered as a measure-theoretic version of Rosen-  
 725     thal's dichotomy. Combining this dichotomy with Theorem 4.5, we get the following  
 726     result:

727     **Theorem 4.13.** *Let  $X$  be a Hausdorff space and  $A \subseteq C(X)$  be pointwise bounded.  
 728     The following are equivalent:*

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

735     *Proof.* Notice that the equivalence (i)-(iii) is Theorem 4.5. Notice that the equiva-  
 736     lence of (iii) and (iv) is Fremlin's Dichotomy (Theorem 4.12).  $\square$

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

739 **Proposition 4.14.** *Let  $X$  be a compact Hausdorff space and  $A \subseteq C(X)$  be point-  
740 wise bounded. If  $A$  is universally Talagrand stable, then  $A$  satisfies the NIP.*

741 *Proof.* By Theorem 4.5, it suffices to show that  $A$  is relatively countably compact in  
742  $\mathcal{M}^0(X, \mu)$  for all Radon probability measure  $\mu$  on  $X$ . Since  $A$  is Talagrand  $\mu$ -stable  
743 for any such  $\mu$ , we have  $\overline{A} \subseteq \mathcal{M}^0(X, \mu)$ . In particular,  $A$  is relatively countably  
744 compact in  $\mathcal{M}^0(X, \mu)$ .  $\square$

745 **Question 4.15.** Is the converse true?

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

749 **Theorem 4.16** (Talagrand, Theorem 9-3-1(a) in [Tal84]). *Let  $X$  be a compact  
750 Hausdorff space and  $A \subseteq M_r(X)$  be countable and pointwise bounded. Assume that  
751  $[0, 1]$  is not the union of  $< \mathfrak{c}$  closed measure zero sets. If  $A$  satisfies the NIP, then  
752  $A$  is universally Talagrand stable.*

753 **Theorem 4.17** (Fremlin, Shelah, [FS93]). *It is consistent that there exists a count-  
754 able pointwise bounded set of Lebesgue measurable functions with the NIP which is  
755 not Talagrand stable with respect to Lebesgue measure.*

756

## REFERENCES

- 757 [ADIW24] Samson Alva, Eduardo Dueñez, Jose Iovino, and Claire Walton. Approximability of deep equilibria. *arXiv preprint arXiv:2409.06064*, 2024. Last revised 20 May 2025, version 3.
- 760 [ADK08] Spiros A. Argyros, Pandelis Dodos, and Vassilis Kanellopoulos. A classification of separable Rosenthal compacta and its applications. *Dissertationes Mathematicae*, 449:1–55, 2008.
- 763 [Ark91] A. V. Arkhangel'skii. *Topological Function Spaces*. Springer, New York, 1st edition, 1991.
- 765 [BD19] Shai Ben-David. Understanding machine learning through the lens of model theory. *The Bulletin of Symbolic Logic*, 25(3):319–340, 2019.
- 767 [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.
- 769 [BKK] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. Preprint. <https://arxiv.org/abs/1909.01377>, <https://implicit-layers-tutorial.org/>.
- 771 [CRBD] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. Preprint. <https://arxiv.org/abs/1806.07366>.
- 773 [Deb13] Gabriel Debs. Descriptive aspects of Rosenthal compacta. In *Recent Progress in General Topology III*, pages 205–227. Springer, 2013.
- 775 [Fre00] David H. Fremlin. *Measure Theory, Volume 1: The Irreducible Minimum*. Torres Fremlin, Colchester, UK, 2000. Second edition 2011.
- 777 [Fre01] David H. Fremlin. *Measure Theory, Volume 2: Broad Foundations*. Torres Fremlin, Colchester, UK, 2001.
- 779 [Fre03] David H. Fremlin. *Measure Theory, Volume 4: Topological Measure Spaces*. Torres Fremlin, Colchester, UK, 2003.
- 781 [FS93] David H. Fremlin and Saharon Shelah. Pointwise compact and stable sets of measurable functions. *Journal of Symbolic Logic*, 58(2):435–455, 1993.
- 783 [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.
- 786 [God80] Gilles Godefroy. Compacts de Rosenthal. *Pacific J. Math.*, 91(2):293–306, 1980.
- 787 [Gro52] A. Grothendieck. Critères de compacité dans les espaces fonctionnels généraux. *Amer. J. Math.*, 74:168–186, 1952.

- 789 [HT23] Clovis Hamel and Franklin D. Tall.  $C_p$ -theory for model theorists. In Jose Iovino, editor, *Beyond First Order Model Theory, Volume II*, chapter 5, pages 176–213. Chapman and Hall/CRC, 2023.
- 790 [Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, 1995.
- 791 [Kha20] Karim Khanaki. Stability, *nip*, and *nsop*; model theoretic properties of formulas via topological properties of function spaces. *Mathematical Logic Quarterly*, 66(2):136–149, 2020.
- 792 [Pap02] E. Pap, editor. *Handbook of measure theory. Vol. I, II*. North-Holland, Amsterdam, 2002.
- 793 [Ros74] Haskell P. Rosenthal. A characterization of Banach spaces containing  $l^1$ . *Proc. Nat. Acad. Sci. U.S.A.*, 71:2411–2413, 1974.
- 794 [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.
- 795 [Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972.
- 796 [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.
- 797 [She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.
- 798 [She78] Saharon Shelah. *Unstable Theories*, volume 187 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1978.
- 799 [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.
- 800 [Sim15a] Pierre Simon. *A Guide to NIP Theories*, volume 44 of *Lecture Notes in Logic*. Cambridge University Press, 2015.
- 801 [Sim15b] Pierre Simon. Rosenthal compacta and NIP formulas. *Fundamenta Mathematicae*, 231(1):81–92, 2015.
- 802 [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.
- 803 [Tal87] Michel Talagrand. The Glivenko-Cantelli problem. *The Annals of Probability*, 15(3):837–870, 1987.
- 804 [Tka11] Vladimir V. Tkachuk. *A  $C_p$ -Theory Problem Book: Topological and Function Spaces*. Problem Books in Mathematics. Springer, 2011.
- 805 [Tod97] Stevo Todorcevic. *Topics in Topology*, volume 1652 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 1997.
- 806 [Tod99] Stevo Todorcevic. Compact subsets of the first Baire class. *Journal of the American Mathematical Society*, 12(4):1179–1212, 1999.
- 807 [Val84] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.
- 808 [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.
- 809 [VC74] Vladimir N Vapnik and Alexey Ya Chervonenkis. *Theory of Pattern Recognition*. Nauka, Moscow, 1974. German Translation: Theorie der Zeichenerkennung, Akademie-Verlag, Berlin, 1979.
- 810 [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.