

# 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 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.

## 1. 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], etc.). In this paper, we combine ideas of topology 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].

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

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

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

<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, real-valued functions that  
<sup>50</sup> arise as the pointwise limit of a sequence of continuous are called *functions of the*  
<sup>51</sup> *first Baire class*, or *Baire class 1* functions, or *Baire-1* for short; Baire class 1 form  
<sup>52</sup> a step above simple continuity in the hierarchy of functions studied in real analysis  
<sup>53</sup> (Baire class 0 functions being continuous functions). Intuitively, Baire-1 functions  
<sup>54</sup> represent functions with “controlled” discontinuities, so they are crucial in topology  
<sup>55</sup> and set 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 the fundamental “prototypes” of sepa-  
<sup>78</sup> rable Rosenthal compacta, and proved that any non-metrizable separable Rosenthal  
<sup>79</sup> compactum must contain a “canonical” embedding of one of these prototypes. They  
<sup>80</sup> showed that if a separable Rosenthal compactum is not hereditarily separable, it  
<sup>81</sup> must contain an uncountable discrete subspace of the size of the continuum.

<sup>82</sup> We believe that the results presented in this paper show practitioners of com-  
<sup>83</sup> putation, or topology, or descriptive set theory, or model theory, how classification  
<sup>84</sup> invariants used in their field translate into classification invariants of other fields.  
<sup>85</sup> 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. The necessary topological background  
 88 beyond undergraduate topology is covered in section 2.

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 2. GENERAL TOPOLOGICAL PRELIMINARIES

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

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

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

106 In this paper, we shall discuss subspaces, and so there is a pertinent subtlety of  
 107 the definitions worth mentioning: *completely metrizable space* is not the same as  
 108 *complete metric space*; for an illustrative example, the interval  $(0, 1)$  with the metric  
 109 inherited from the reals is not complete, but it is Polish since that is homeomorphic  
 110 to the real line. Being Polish is a topological property.

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

115 **Fact 2.1.** *A subset  $A$  of a Polish space  $X$  is itself Polish in the subspace topology  
 116 if and only if it is a  $G_\delta$  set. In particular, closed subsets and open subsets of Polish  
 117 spaces are also Polish spaces.*

118 Given two topological spaces  $X$  and  $Y$  we denote by  $C_p(X, Y)$  the set of all  
 119 continuous functions  $f : X \rightarrow Y$  endowed with the topology of pointwise convergence.  
 120 When  $Y = \mathbb{R}$ , we denote this collection simply as  $C_p(X)$ . A natural question is, how  
 121 do topological properties of  $X$  translate to  $C_p(X)$  and vice versa? These questions,  
 122 and in general the study of these spaces, are the concern of  $C_p$ -theory, an active  
 123 field of research in general topology which was pioneered by A. V. Arhangel'skiĭ  
 124 and his students in the 1970's and 1980's. This field has found many applications in  
 125 model theory and functional analysis. Recent surveys on the topics include [HT23]  
 126 and [Tka11].

127 A *Baire class 1* function between topological spaces is a function that can be  
 128 expressed as the pointwise limit of a sequence of continuous functions. If  $X$  and  $Y$   
 129 are topological spaces, the Baire class 1 functions  $f : X \rightarrow Y$  endowed with the  
 130 topology of pointwise convergence is denoted  $B_1(X, Y)$ . As above, in the special  
 131 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

132 hierarchy of functions was introduced by French mathematician René-Louis Baire  
 133 in his 1899 doctoral thesis, *Sur les fonctions de variables réelles*. His work moved  
 134 away from the 19th-century preoccupation with "pathological" functions toward a  
 135 constructive classification based on pointwise limits.

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

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

141 **Fact 2.2** (Baire). *If  $X$  is perfectly normal, then the following conditions are equiv-*  
 142 *alent for a function  $f : X \rightarrow \mathbb{R}$ :*

- 143 •  $f$  is a Baire class 1 function, that is,  $f \in B_1(X)$ .
- 144 •  $f^{-1}[U]$  is an  $F_\sigma$  subset of  $X$  whenever  $U \subseteq Y$  is open.
- 145 •  $f$  is a pointwise limit of continuous functions.
- 146 • For every closed  $F \subseteq X$ , the restriction  $f|_F$  has a point of continuity.

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

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

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

155 **Lemma 2.3.** *Let  $X$  be a Polish space and  $A \subseteq B_1(X)$  be pointwise bounded. The*  
 156 *following are equivalent:*

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

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

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

164 (ii) $\Rightarrow$ (iii) Assume that  $A$  is relatively countably compact in  $B_1(X)$  and that  
 165  $f \in \overline{A} \setminus B_1(X)$ . By Fact 2.2, there are countable  $D_0, D_1 \subseteq X$  with  $\overline{D_0} = \overline{D_1}$ , and  
 166  $a < b$  such that  $D_0 \subseteq f^{-1}(-\infty, a]$  and  $D_1 \subseteq f^{-1}[b, \infty)$ . We claim that there is a  
 167 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,  
 168 use the countability to enumerate  $D_0 \cup D_1$  as  $\{x_n\}_{n \in \mathbb{N}}$ . Then for each positive  $n$   
 169 find  $f_n \in A$  with  $|f_n(x_i) - f(x_i)| < \frac{1}{n}$  for all  $i \leq n$ . The claim follows.

170 By relative countable compactness of  $A$ , there is an accumulation point  $g \in$   
 171  $B_1(X)$  of  $\{f_n\}_{n \in \mathbb{N}}$ . It is straightforward to show that since  $f$  and  $g$  agree on  $D_0 \cup D_1$ ,  
 172  $g$  does not have a point of continuity on the closed set  $\overline{D_0} = \overline{D_1}$ , which contradicts  
 173 Fact 2.2.

174 (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  
 175  $\prod_{x \in X} [-M_x, M_x]$ ; Tychonoff's theorem states that the product of compact spaces

176 is always compact, and since closed subsets of compact spaces are compact,  $\bar{A}$  must  
 177 be compact, as desired.  $\square$

178 **2.1. From Rosenthal's dichotomy to Shelah's NIP.** The fundamental idea  
 179 that connects the rich theory here presented to real-valued computations is the  
 180 concept of an *approximation*. In the reals, points of closure from some subset  
 181 can always be approximated by points inside the set, via a convergent sequence.  
 182 For more complicated spaces, such as  $C_p(X)$ , this fails in remarkable ways. To  
 183 see an example, consider the Cantor space  $X = 2^{\mathbb{N}}$ , and for each  $n \in \mathbb{N}$  define  
 184  $p_n : X \rightarrow \{0, 1\}$  by  $p_n(x) = x(n)$  for each  $x \in X$ . Then  $p_n$  is continuous for each  $n$ ,  
 185 but one can show (see Chapter 1.1 of [Tod97] for details) that the only continuous  
 186 functions in the closure of  $\{p_n\}_{n \in \mathbb{N}}$  are the functions  $p_n$  themselves; moreover, none  
 187 of the subsequences of  $\{p_n\}_{n \in \mathbb{N}}$  converge. In some sense, this example is the worst  
 188 possible scenario for convergence. The topological space obtained from this closure  
 189 is well-known: it is the *Stone-Čech compactification* of the discrete space of natural  
 190 numbers, or  $\beta\mathbb{N}$  for short, and it is an important object of study in general topology.

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

194 **Theorem 2.4** (Rosenthal's Dichotomy, 1974). *If  $X$  is Polish and  $\{f_n\} \subseteq C_p(X)$   
 195 is pointwise bounded, then either  $\{f_n\}_{n \in \mathbb{N}}$  contains a convergent subsequence or a  
 196 subsequence whose closure (in  $\mathbb{R}^X$ ) is homeomorphic to  $\beta\mathbb{N}$ .*

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

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

206 **Theorem 2.5** (The BFT Dichotomy. Bourgain-Fremlin-Talagrand, [BFT78, The-  
 207 oreom 4G]). *Let  $X$  be a Polish space and  $A \subseteq C_p(X)$  be pointwise bounded. The  
 208 following are equivalent:*

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

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

210 **Definition 2.6.** We shall say that a set  $A \subseteq \mathbb{R}^X$  has the *Independence Property*, or  
 211 IP for short, if it satisfies the following condition: There exists every  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$   
 212 and  $a < b$  such that for every pair of disjoint sets  $E, F \subseteq \mathbb{N}$ , we have

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

213 If  $A$  satisfies the negation of this condition, we will say that  $A$  *satisfies NIP*, or  
 214 that has the NIP.

*Remark 2.7.* Note that if  $X$  is compact and  $A \subseteq C_p(X)$ , then  $A$  has 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

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

215 To summarize, the particular case of Theorem 2.8 when for  $X$  compact can be  
216 stated in the following way:

217 **Theorem 2.8.** *Let  $X$  be a compact Polish space. Then, for every pointwise bounded  
218  $A \subseteq C_p(X)$ , one and exactly one of the following two conditions must hold:*

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

221 The Independence Property was first isolated by Saharon Shelah in model theory  
222 as a dividing line between theories whose models are “tame” (corresponding to  
223 NIP) theories of models are “wild” (corresponding to IP). See [She71, Definition  
224 4.1],[She90].

225 **2.2. A universal polynomial vs exponential dichotomy.** The particular case  
226 of the BSF Dichotomy (Theorem 2.8) when  $A$  consists of  $\{0, 1\}$ -valued (i.e., {Yes, No}-  
227 valued strings was discovered independently, around 1971-1972 in many founda-  
228 tional contexts related to polynomial (“tame”) vs exponential (“wild”) complexity:  
229 In model theory by Saharon Shelah [She71],[She90], in combinatorics by Norbert  
230 Sauer [Sau72] and Shelah [She72, She90], and in statistical learning, by Vladimir  
231 Vapnik and Alexey Chervonenkis [VC71, VC74].

232 Saharon Shelah’s classification theory is a foundational program in mathematical  
233 logic devised to categorize first-order theories based on the complexity and structure  
234 of their models. A theory  $T$  is considered classifiable in Shelah’s sense if the number  
235 of non-isomorphic models of  $T$  of a given cardinality can be described by a bounded  
236 number of numerical invariants. In contrast, a theory  $T$  is unclassifiable if the  
237 number of models of  $T$  of a given cardinality is the maximum possible number.  
238 This number is directly impacted by the number of “types” over of parameters in  
239 models of  $T$ ; a controlled number of types is a characteristic of a classifiable theory.

240 In Shelah’s classification program [She90], theories without the independence  
241 property (called NIP theories, or dependent theories) have a well-behaved, “tame”  
242 structure; the number of types over a set of parameters of size  $\kappa$  of such a theory  
243 is of polynomially or similar “slow” growth on  $\kappa$ . Theories with the Independence  
244 Property (called IP theories), in contrast, are considered “intractable” or “wild”.  
245 A theory with the independence property produces the maximum possible number  
246 of types over a set of parameters; for a set of parameters of cardinality  $\kappa$ , the theory  
247 will have  $2^{2^\kappa}$ -many distinct types.

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

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

252 for every  $A \subseteq S$  such that  $|A| \geq N$ (yielding polynomial complexity). This answered  
253 a question of Erdős.

Readers familiar with statistical learning may recognize the Sauer-Shelah lemma as the dichotomy discovered and proved slightly (1971) earlier by Vapknis and Chervonenkis [VC71, VC74] to address the problem of uniform convergence in statistics. The integer  $N$  given by the preceding paragraph, when it exists, is called the *VC-dimension* of  $\mathcal{F}$ . This is a core concept in machine learning. The lemma provides upper bounds on the number of data points (sample size  $m$ ) needed to learn a concept class with VC dimension  $d$  by showing this number grows polynomially with  $m$  and  $d$  (namely,  $\sum_{i=0}^d \binom{m}{i} \approx O(m^d)$ ), not exponentially.

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

**Lemma 2.9.** *Let  $X$  be a Polish space and  $\mathcal{P}$  be a countable set. Then,  $f \in B_1(X, \mathbb{R}^{\mathcal{P}})$  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 2.10.** *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 2.9, 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 2.9 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 2.8.

**Theorem 2.11.** *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  $K \subseteq X$ :*

- (1)  $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^\mathcal{P})$ .
- (2)  $\pi_P \circ A|_K$  has 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 2.10 we get  $\widehat{A|_{\mathcal{P} \times K}} \subseteq B_1(\mathcal{P} \times K)$ . By Theorem 2.8, 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

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

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

(2) $\Rightarrow$ (1) Fix  $f \in \overline{A|_K}$ . By lemma 2.9 it suffices to show that  $\pi_P \circ f \in B_1(K)$  for all  $P \in \mathcal{P}$ . By (2),  $\pi_P \circ A|_K$  has the NIP. Hence, by Theorem 2.8 we have  $\pi_P \circ \overline{A|_K} \subseteq B_1(K)$ . But then  $\pi_P \circ f \in \pi_P \circ \overline{A|_K} \subseteq B_1(K)$ .  $\square$

Lastly, a simple but significant result that helps understand the operation of restricting a set of functions to a specific subspace of the domain space  $X$ , of course in the context of the NIP, is that we may always assume that said subspace is closed. Concretely, whether we take its closure or not has no effect on the NIP:

**Lemma 2.12.** *Assume that  $X$  is Hausdorff and that  $A \subseteq C_p(X)$ . The following are equivalent for every  $L \subseteq X$ :*

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

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

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

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

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

By definition of closure:

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

This contradicts (i).  $\square$

## 303 3. NIP IN THE CONTEXT OF COMPUTATION

304 In this section, we study what the NIP tell us in the context of deep computations as defined in [ADIW24]. We say a structure  $(L, \mathcal{P}, \Gamma)$  is a *Compositional*  
 305 *Computation Structure* (CCS) if  $L \subseteq \mathbb{R}^{\mathcal{P}}$  is a subspace of  $\mathbb{R}^{\mathcal{P}}$ , with the pointwise  
 306 convergence topology, and  $\Gamma \subseteq L^L$  is a semigroup under composition. The motivation  
 307 for CCS comes from (continuous) model theory, where  $\mathcal{P}$  is a fixed collection  
 308 of predicates and  $L$  is a (real-valued) structure. Every point in  $L$  is identified with  
 309 its “type”, which is the tuple of all values the point takes on the predicates from  
 310  $\mathcal{P}$ , i.e., an element of  $\mathbb{R}^{\mathcal{P}}$ . In this context, elements of  $\mathcal{P}$  are called *features*. In the  
 311 discrete model theory framework, one views the space of complete-types as a sort of  
 312 compactification of the structure  $L$ . In this context, we don’t want to consider only  
 313 points in  $L$  (realized types) but in its closure  $\overline{L}$  (possibly unrealized types). The  
 314 problem is that the closure  $\overline{L}$  is not necessarily compact, an assumption that turns  
 315 out to be very useful in the context of continuous model theory. To bypass this  
 316 problem in a framework for deep computations, Alva, Dueñez, Iovino and Walton  
 317 introduced in [ADIW24] the concept of *shards*, which essentially consists in covering  
 318 (a large fragment) of the space  $\overline{L}$  by compact, and hence pointwise-bounded,  
 319 subspaces (shards). We shall give the formal definition next.

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

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

323 For an illustrative example, we can frame Newton’s polynomial root approximation  
 324 method in the context of a CCS (see Example 5.6 of [ADIW24] for details) as  
 325 follows. Begin by considering the extended complex numbers  $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$  with  
 326 the usual Riemann sphere topology that makes it into a compact space (where  
 327 unbounded sequences converge to  $\infty$ ). In fact, not only is this space compact  
 328 but it is covered by the shard given by the sizer  $(1, 1, 1)$  (the unit sphere is con-  
 329 tained in the cube  $[-1, 1]^3$ ). The space  $\hat{\mathbb{C}}$  is homeomorphic to the usual unit  
 330 sphere  $S^2 := \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$  of  $\mathbb{R}^3$ , by means of the stereographic  
 331 projection and its inverse  $\hat{\mathbb{C}} \rightarrow S^2$ . This function is regarded as a triple of pred-  
 332 icates  $x, y, z : \hat{\mathbb{C}} \rightarrow [-1, 1]$  where each will map an extended complex number to  
 333 its corresponding real coordinate on the cube  $[-1, 1]^3$ . Now fix the cubic com-  
 334 plex polynomial  $p(s) := s^3 - 1$ , and consider the map which performs one step  
 335 in Newton’s method at a particular (extended) complex number  $s$ , for finding  
 336 a root of  $p$ ,  $\gamma_p : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ . The explicit inner workings of  $\gamma_p$  are irrelevant for  
 337 this example, except for the fact that it is a continuous mapping. It follows that  
 338  $(S^2, \{x, y, z\}, \{\gamma_p^k : k \in \mathbb{N}\})$  is a CCS. The idea is that repeated applications of  
 339  $\gamma_p(s), \gamma_p \circ \gamma_p(s), \gamma_p \circ \gamma_p \circ \gamma_p(s), \dots$  would approximate a root of  $p$  provided  $s$  was  
 340 a good enough initial guess.

341 The  $r_{\bullet}$ -type-shard is defined as  $\mathcal{L}[r_{\bullet}] = \overline{L[r_{\bullet}]}$  and  $\mathcal{L}_{sh}$  is the union of all type-  
 342 shards. Notice that  $\mathcal{L}_{sh}$  is not necessarily equal to  $\mathcal{L} = \overline{L}$ , unless  $\mathcal{P}$  is countable  
 343 (see [ADIW24]). A *transition* is a map  $f : L \rightarrow L$ , in particular, every element  
 344 in the semigroup  $\Gamma$  is a transition (these are called *realized computations*). In  
 345 practice, one would like to work with “definable” computations, i.e., ones that can  
 346 be described by a computer. In this topological framework, being continuous is an  
 347 expected requirement. However, as in the case of complete-types in model theory,

we will work with “unrealized computations”, i.e., maps  $f : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$ . Note that continuity of a computation does not imply that it can be continuously extended to  $\mathcal{L}_{sh}$ .

Suppose that a transition map  $f : L \rightarrow \mathcal{L}$  can be extended continuously to a map  $\tilde{f} : \mathcal{L} \rightarrow \mathcal{L}$ . Then, the Stone-Weierstrass theorem implies that the feature  $\pi_P \circ f$  (here  $P$  is a fixed predicate, and the feature is hence continuous) can be uniformly approximated by polynomials on the compact set  $\mathcal{L}[r_\bullet]$ . Theorem 2.2 in [ADIW24] formalizes the converse of this fact, in the sense that transitions maps that are not continuously extendable in this fashion cannot be obtained from simple constructions involving predicates. Under this framework, the features  $\pi_P \circ f$  of such transitions  $f$  are not approximable by polynomials, and so they are understood as “non-computable” since, again, we expect the operations computers carry out to be determined by elementary algebra corresponding to polynomials (namely addition and multiplication). Therefore it is crucial we assume some extendibility conditions.

We say that the CCS  $(L, \mathcal{P}, \Gamma)$  satisfies the *Extendibility Axiom* if for all  $\gamma \in \Gamma$ , there is  $\tilde{\gamma} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$  such that for every sizer  $r_\bullet$  there is an  $s_\bullet$  such that  $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$  is continuous. For a deeper discussion about this axiom, we refer the reader to [ADIW24].

A collection  $R$  of sizers is called *exhaustive* if  $\mathcal{L}_{sh} = \bigcup_{r_\bullet \in R} \mathcal{L}[r_\bullet]$ . 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  $\Delta$  are called *real-valued computations* (in this article we will refer to them simply as *computations*) and elements in  $\tilde{\Delta} \subseteq \mathcal{L}_{sh}^L$  are called (real-valued) *deep computations* or *ultracomputations*. By  $\tilde{\Delta}$  we denote the set of all extensions  $\tilde{\gamma}$  for  $\gamma \in \Delta$ . For a more complete description of this framework, we refer the reader to [ADIW24].

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

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

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

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

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

*Proof.* Since  $\mathcal{P}$  is countable, then  $\mathcal{L}[r_\bullet] \subseteq \mathbb{R}^\mathcal{P}$  is Polish. Also, the Extendibility Axiom implies that  $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$  is a pointwise bounded set of continuous functions

for all  $P \in \mathcal{P}$ . Hence, Theorem 2.11 and Lemma 2.12 prove the equivalence of (1) and (2). If (1) holds and  $f \in \overline{\Delta}$ , then write  $f = \underline{\text{Ulim}}_i \gamma_i$  as an ultralimit. Define  $\tilde{f} := \underline{\text{Ulim}}_i \tilde{\gamma}_i$ . Hence, for all  $r_\bullet \in R$  we have  $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ . That every deep computation is a pointwise limit of a countable sequence of computations follows from the fact that for a Polish space  $X$  every compact subset of  $B_1(X)$  is Fréchet-Urysohn (that is, a space where topological closures coincide with sequential closures, see Theorem 3F in [BFT78] or Theorem 4.1 in [Deb13]).  $\square$

Given a countable set  $\Delta$  of computations satisfying the NIP on features and shards (condition (2) of Theorem 3.1) we have that  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  (for a fixed sizer  $r_\bullet$ ) is a separable *Rosenthal compactum* (compact subset of  $B_1(P \times \mathcal{L}[r_\bullet])$ ). The work of Todorčević ([Tod99]) and Argyros, Dodos, Kanellopoulos ([ADK08]) culminates in a trichotomy theorem for separable Rosenthal Compacta. Inspired by the work of Glasner and Megrelishvili ([GM22]), we are interested to see how this allows us to classify and obtain different levels of PAC-learnability (NIP).

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

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

- (i)  $\text{NIP}_1$  if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is first countable for every  $r_\bullet \in R$ .
- (ii)  $\text{NIP}_2$  if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is hereditarily separable for every  $r_\bullet \in R$ .
- (iii)  $\text{NIP}_3$  if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is metrizable for every  $r_\bullet \in R$ .

Observe that  $\text{NIP}_3 \Rightarrow \text{NIP}_2 \Rightarrow \text{NIP}_1 \Rightarrow \text{NIP}$ . A natural question that would continue this work is to find examples of CCS that separate these levels of NIP. In [Tod99], Todorčević isolates 3 canonical examples of Rosenthal compacta that witness the failure of the converse implications above. We now present some separable and non-separable examples of Rosenthal compacta:

- (1) *Alexandroff compactification of a discrete space of size continuum.* For each  $a \in 2^\mathbb{N}$  consider the map  $\delta_a : 2^\mathbb{N} \rightarrow \mathbb{R}$  given by  $\delta_a(x) = 1$  if  $x = a$  and  $\delta_a(x) = 0$  otherwise. Let  $A(2^\mathbb{N}) = \{\delta_a : a \in 2^\mathbb{N}\} \cup \{0\}$ , where  $0$  is the zero map. Notice that  $A(2^\mathbb{N})$  is a compact subset of  $B_1(2^\mathbb{N})$ , in fact  $\{\delta_a : a \in 2^\mathbb{N}\}$  is a discrete subspace of  $B_1(2^\mathbb{N})$  and its pointwise closure is precisely  $A(2^\mathbb{N})$ . Hence, this is a Rosenthal compactum which is not first countable. Notice that this space is also not separable.
- (2) *Extended Alexandroff compactification.* For each finite binary sequence  $s \in 2^{<\mathbb{N}}$ , let  $v_s : 2^\mathbb{N} \rightarrow \mathbb{R}$  be given by  $v_s(x) = 1$  if  $x$  extends  $s$  and  $v_s(x) = 0$  otherwise. Let  $\hat{A}(2^\mathbb{N})$  be the pointwise closure of  $\{v_s : s \in 2^{<\mathbb{N}}\}$ , i.e.,  $\hat{A}(2^\mathbb{N}) = A(2^\mathbb{N}) \cup \{v_s : s \in 2^{<\mathbb{N}}\}$ . Note that this space is a separable Rosenthal compactum which is not first countable.
- (3) *Split Cantor.* Let  $<$  be the lexicographic order in the space of infinite binary sequences, i.e.,  $2^\mathbb{N}$ . For each  $a \in 2^\mathbb{N}$  let  $f_a^- : 2^\mathbb{N} \rightarrow \mathbb{R}$  be given by

435        $f_a^-(x) = 1$  if  $x < a$  and  $f_a^-(x) = 0$  otherwise. Let  $f_a^+ : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  be given  
 436       by  $f_a^+(x) = 1$  if  $x \leq a$  and  $f_a^+(x) = 0$  otherwise. The split Cantor is the  
 437       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  
 438       compactum. One example of a countable dense subset is the set of all  $f_a^+$   
 439       and  $f_a^-$  where  $a$  is an infinite binary sequence that is eventually constant.  
 440       Moreover, it is hereditarily separable but it is not metrizable.

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

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

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

441       Let  $D(K) = \{g_a^0 : a \in K\} \cup \{g_a^1 : a \in K\}$ . Notice that  $D(K)$  is a first  
 442       countable Rosenthal compactum. It is not separable if  $K$  is uncountable.  
 443       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

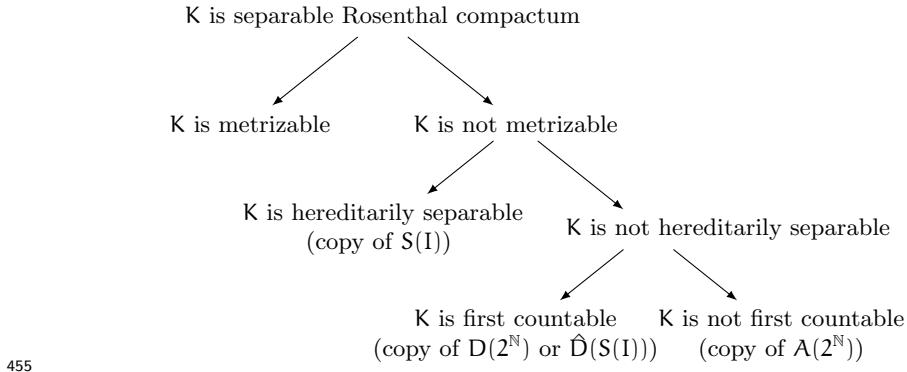
$$h_t(x) = \begin{cases} 0, & x < a_t \\ 1/2, & a_t \leq x \leq b_t \\ 1, & b_t < x \end{cases}$$

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

448       **Theorem 3.3** (Todorčević's Trichotomy, [Tod99], Theorem 3 in [ADK08]). *Let  $K$   
 449       be a separable Rosenthal Compactum.*

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

454       In other words, we have the following classification:



455

456 Lastly, the definitions provided here for  $NIP_i$  ( $i = 1, 2, 3$ ) are topological.

457 **Question 3.4.** Is there a non-topological characterization for  $NIP_i$ ,  $i = 1, 2, 3$ ?

458 More can be said about the nature of the embeddings in Todorčević's Trichotomy.  
 459 Given a separable Rosenthal compactum  $K$ , there is typically more than one countable dense subset of  $K$ . We can view a separable Rosenthal compactum as the accumulation points of a countable family of pointwise bounded real-valued functions.  
 460 The choice of the countable families is not important when a bijection between them can be lifted to a homeomorphism of their closures. To be more precise:

461 **Definition 3.5.** Given a Polish space  $X$ , a countable set  $I$  and two pointwise bounded families  $\{f_i : i \in I\} \subseteq \mathbb{R}^X$ ,  $\{g_i : i \in I\} \subseteq \mathbb{R}^X$  indexed 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  $\overline{\{g_i : i \in I\}}$ .

462 Notice that in the separable examples discussed before ( $\hat{A}(2^N)$ ,  $S(2^N)$  and  $\hat{D}(S(2^N))$ ), the countable dense subsets are indexed by the binary tree  $2^{<\mathbb{N}}$ . This choice of index is useful because the Ramsey theory of perfect subsets of the Cantor space  $2^{\mathbb{N}}$  can be imported 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].

463 **Definition 3.6.** Given a Polish space  $X$  and a pointwise bounded family  $\{f_t : t \in 2^{<\mathbb{N}}\}$ . 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  $\{f_t : t \in 2^{<\mathbb{N}}\}$ .

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

489  $\{x \in 2^{\mathbb{N}} : a < x\}$ . Given two maps  $f, g : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  we denote by  $(f, g) : 2^{\mathbb{N}} \sqcup 2^{\mathbb{N}} \rightarrow \mathbb{R}$   
490 the function which is  $f$  on the first copy of  $2^{\mathbb{N}}$  and  $g$  on the second copy of  $2^{\mathbb{N}}$ .

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

498 **Theorem 3.7** (Heptacotomy of minimal families, Theorem 2 in [ADK08]). *Let  
499  $X$  be Polish. For every relatively compact  $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$ , there exists  $i =$   
500  $1, 2, \dots, 7$  and a regular dyadic subtree  $\{s_t : t \in 2^{<\mathbb{N}}\}$  of  $2^{<\mathbb{N}}$  such that  $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$   
501 is equivalent to  $D_i$ . Moreover, all  $D_i$  are minimal and mutually non-equivalent.*

502 **3.2. NIP and definability by universally measurable functions.** We now  
503 turn to the question: what happens when  $\mathcal{P}$  is uncountable? Notice that the count-  
504 ability assumption is crucial in the proof of Theorem 2.11 essentially because it  
505 makes  $\mathbb{R}^{\mathcal{P}}$  a Polish space. For the uncountable case, we may lose Baire-1 definabil-  
506 ity so we shall replace  $B_1(X)$  by a bigger class. Recall that the purpose of studying  
507 the class of Baire-1 functions is that a pointwise limit of continuous functions is not  
508 necessarily continuous. In [BFT78], J. Bourgain, D.H. Fremlin and M. Talagrand  
509 characterized the Non-Independence Property of a set of continuous functions with  
510 various notions of compactness in function spaces containing  $C(X)$ , such as  $B_1(X)$ .  
511 In this section we will replace  $B_1(X)$  with the larger space  $M_r(X)$  of universally  
512 measurable functions. The development of this section is based on Theorem 2F in  
513 [BFT78]. We now give the relevant definitions. Readers with little familiarity with  
514 measure theory can review the appendix for standard definitions appearing in this  
515 subsection.

516 Given a Hausdorff space  $X$  and a measurable space  $(Y, \Sigma)$ , we say that  $f : X \rightarrow Y$   
517 is *universally measurable* (with respect to  $\Sigma$ ) if  $f^{-1}(E)$  is universally measurable  
518 for every  $E \in \Sigma$ , i.e.,  $f^{-1}(E)$  is  $\mu$ -measurable for every Radon probability measure  
519  $\mu$  on  $X$ . When  $Y = \mathbb{R}$  we will always take  $\Sigma = \mathcal{B}(\mathbb{R})$ , the Borel  $\sigma$ -algebra of  $\mathbb{R}$ .  
520 In that case, a function  $f : X \rightarrow \mathbb{R}$  is universally measurable if and only if  $f^{-1}(U)$   
521 is  $\mu$ -measurable for every Radon probability measure  $\mu$  on  $X$  and every open set  
522  $U \subseteq \mathbb{R}$ . Following [BFT78], the collection of all universally measurable real-valued  
523 functions will be denoted by  $M_r(X)$ . In the context of deep computations, we will  
524 be interested in transition maps from a state space  $L \subseteq \mathbb{R}^{\mathcal{P}}$  to itself. There are two  
525 natural  $\sigma$ -algebras one can consider in the product space  $\mathbb{R}^{\mathcal{P}}$ : the Borel  $\sigma$ -algebra,  
526 i.e., the  $\sigma$ -algebra generated by open sets in  $\mathbb{R}^{\mathcal{P}}$ ; and the cylinder  $\sigma$ -algebra, i.e.,  
527 the  $\sigma$ -algebra generated by Borel cylinder sets or equivalently basic open sets in  
528  $\mathbb{R}^{\mathcal{P}}$ . Note that when  $\mathcal{P}$  is countable, both  $\sigma$ -algebras coincide but in general the  
529 cylinder  $\sigma$ -algebra is strictly smaller. We will use the cylinder  $\sigma$ -algebra to define  
530 universally measurable maps  $f : \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}^{\mathcal{P}}$ . The reason for this choice is because of  
531 the following characterization:

532 **Lemma 3.8.** *Let  $X$  be a Hausdorff space and  $Y = \prod_{i \in I} Y_i$  be any product of  
533 measurable spaces  $(Y_i, \Sigma_i)$  for  $i \in I$ . Let  $\Sigma_Y$  be the cylinder  $\sigma$ -algebra generated by  
534 the measurable spaces  $(Y_i, \Sigma_i)$ . Let  $f : X \rightarrow Y$ . The following are equivalent:*

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

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

542        The previous lemma says that a transition map is universally measurable if and  
 543        only if it is universally measurable on all its features. In other words, we can check  
 544        measurability of a transition just by checking measurability in all its features. We  
 545        will denote by  $M_r(X, \mathbb{R}^P)$  the collection of all universally measurable functions  
 546         $f : X \rightarrow \mathbb{R}^P$  (with respect to the cylinder  $\sigma$ -algebra), endowed with the topology of  
 547        pointwise convergence.

548        **Definition 3.9.** Let  $(L, \mathcal{P}, \Gamma)$  be a CCS. We say that a transition  $f : L \rightarrow L$  is  
 549        *universally measurable shard-definable* if and only if there exists  $\tilde{f} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$   
 550        extending  $f$  such that for every sizer  $r_\bullet$  there is a sizer  $s_\bullet$  such that the restriction  
 551         $\tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$  is universally measurable, i.e.  $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow [-s_P, s_P]$   
 552        is  $\mu$ -measurable for every Radon probability measure  $\mu$  on  $\mathcal{L}[r_\bullet]$ .

553        We will need the following result about NIP and universally measurable func-  
 554        tions:

555        **Theorem 3.10** (Bourgain-Fremlin-Talagrand, Theorem 2F in [BFT78]). *Let  $X$  be a  
 556        Hausdorff space and  $A \subseteq C(X)$  be pointwise bounded. The following are equivalent:*

- 557        (i)  $\overline{A} \subseteq M_r(X)$ .  
 558        (ii) *For every compact  $K \subseteq X$ ,  $A|_K$  has the NIP.*  
 559        (iii) *For every Radon measure  $\mu$  on  $X$ ,  $A$  is relatively countably compact in  
 560               $\mathcal{L}^0(X, \mu)$ , i.e., every countable subset of  $A$  has an accumulation point in  
 561               $\mathcal{L}^0(X, \mu)$ .*

562        Theorem 2.8 immediately yields the following.

563        **Theorem 3.11.** *Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendibility Axiom. Let  $R$   
 564        be an exhaustive collection of sizers. Let  $\Delta \subseteq \Gamma$  be  $R$ -confined. If  $\pi_P \circ \Delta|_{L[r_\bullet]}$  has  
 565        the NIP for all  $P \in \mathcal{P}$  and all  $r_\bullet \in R$ , then every deep computation is universally  
 566        measurable shard-definable.*

567        *Proof.* By the Extendibility Axiom, Theorem 2.8 and lemma 2.12 we have that  
 568         $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq M_r(\mathcal{L}[r_\bullet])$  for all  $r_\bullet \in R$  and  $P \in \mathcal{P}$ . Let  $f \in \overline{\Delta}$  be a deep computation.  
 569        Write  $f = \mathcal{U} \lim_i \gamma_i$  as an ultralimit of computations in  $\Delta$ . Define  $\tilde{f} := \mathcal{U} \lim_i \tilde{\gamma}_i$ .  
 570        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  $\pi_P \circ f|_{\mathcal{L}[r_\bullet]} \in$   
 571         $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq M_r(\mathcal{L}[r_\bullet])$ .  $\square$

572        **Question 3.12.** Under the same assumptions of the previous Theorem, suppose  
 573        that every deep computation of  $\Delta$  is universally measurable shard-definable. Must  
 574         $\pi_P \circ \Delta|_{L[r_\bullet]}$  have the NIP for all  $P \in \mathcal{P}$  and all  $r_\bullet \in R$ ?

575 **3.3. Talagrand stability and definability by universally measurable functions.** There is another notion closely related to NIP, introduced by Talagrand  
 576 in [Tal84] while studying Pettis integration. Suppose that  $X$  is a compact Haus-  
 577 dorff space and  $A \subseteq \mathbb{R}^X$ . Let  $\mu$  be a Radon probability measure on  $X$ . Given a  
 578  $\mu$ -measurable set  $E \subseteq X$ , a positive integer  $k$  and real numbers  $a < b$ . we write:

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

580 We say that  $A$  is *Talagrand  $\mu$ -stable* if and only if for every  $\mu$ -measurable  
 581 set  $E \subseteq X$  of positive measure and for every  $a < b$  there is  $k \geq 1$  such that  
 582  $(\mu^{2k})^*(D_k(A, E, a, b)) < (\mu(E))^{2k}$ . Notice that we work with the outer measure  
 583 because it is not necessarily true that the sets  $D_k(A, E, a, b)$  are  $\mu$ -measurable.  
 584 This is certainly the case when  $A$  is a countable set of continuous (or  $\mu$ -measurable)  
 585 functions.

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

589 **Lemma 3.13.** *If  $A$  is Talagrand  $\mu$ -stable, then  $\bar{A}$  is also Talagrand  $\mu$ -stable and  
 590  $\bar{A} \subseteq \mathcal{L}^0(X, \mu)$ .*

591 *Proof.* First, observe that a subset of a  $\mu$ -stable set is  $\mu$ -stable. To show that  $\bar{A}$   
 592 is  $\mu$ -stable, observe that  $D_k(\bar{A}, E, a, b) \subseteq D_k(A, E, a', b')$  where  $a < a' < b' <$   
 593  $b$  and  $E$  is a  $\mu$ -measurable set with positive measure. It suffices to show that  
 594  $\bar{A} \subseteq \mathcal{L}^0(X, \mu)$ . Suppose that there exists  $f \in \bar{A}$  such that  $f \notin \mathcal{L}^0(X, \mu)$ . By a  
 595 characterization of measurable functions (see 413G in [Fre03]), there exists a  $\mu$ -  
 596 measurable set  $E$  of positive measure and  $a < b$  such that  $\mu^*(P) = \mu^*(Q) = \mu(E)$   
 597 where  $P = \{x \in E : f(x) \leq a\}$  and  $Q = \{x \in E : f(x) \geq b\}$ . Then, for any  $k \geq 1$ :  
 598  $(P \times Q)^k \subseteq D_k(\{f\}, E, a, b)$  so  $(\mu^{2k})^*(D_k(\{f\}, E, a, b)) = (\mu^*(P)\mu^*(Q))^k = (\mu(E))^{2k}$ .  
 599 Thus,  $\{f\}$  is not  $\mu$ -stable, but we argued before that a subset of a  $\mu$ -stable set must  
 600 be  $\mu$ -stable.  $\square$

601 We say that  $A$  is *universally Talagrand stable* if  $A$  is Talagrand  $\mu$ -stable for  
 602 every Radon probability measure  $\mu$  on  $X$ . A similar argument as before, yields the  
 603 following:

604 **Theorem 3.14.** *Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendibility Axiom. If  $\pi_P \circ$   
 605  $\Delta|_{L[r_\bullet]}$  is universally Talagrand stable for all  $P \in \mathcal{P}$  and all sizers  $r_\bullet$ , then every  
 606 deep computation is universally measurable sh-definable.*

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

609 **Lemma 3.15** (Fremlin's Dichotomy, 463K in [Fre03]). *If  $(X, \Sigma, \mu)$  is a perfect  $\sigma$ -  
 610 finite measure space (in particular, for  $X$  compact and  $\mu$  a Radon probability measure  
 611 on  $X$ ) and  $\{f_n : n \in \mathbb{N}\}$  be a sequence of real-valued measurable functions on  $X$ , then  
 612 either:*

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

616     The preceding lemma can be considered as the measure theoretic version of  
 617     Rosenthal's Dichotomy. Combining this dichotomy with the Theorem 3.10 we get  
 618     the following result:

619     **Theorem 3.16.** *Let  $X$  be a Hausdorff space and  $A \subseteq C(X)$  be pointwise bounded.  
 620     The following are equivalent:*

- 621       (i)  $\overline{A} \subseteq M_r(X)$ .
- 622       (ii) *For every compact  $K \subseteq X$ ,  $A|_K$  has the NIP.*
- 623       (iii) *For every Radon measure  $\mu$  on  $X$ ,  $A$  is relatively countably compact in  
                    $L^0(X, \mu)$ , i.e., every countable subset of  $A$  has an accumulation point in  
                    $L^0(X, \mu)$ .*
- 626       (iv) *For every Radon measure  $\mu$  on  $X$  and every sequence  $\{f_n : n \in \mathbb{N}\} \subseteq A$ ,  
                   there is a subsequence that converges  $\mu$ -almost everywhere.*

628     *Proof.* Notice that the equivalence (i)-(iii) is Theorem 3.10. Notice that the equivalence of (iii) and (iv) is Fremlin's Dichotomy Theorem.  $\square$

630     **Lemma 3.17.** *Let  $X$  be a compact Hausdorff space and  $A \subseteq C(X)$  be pointwise bounded. If  $A$  is universally Talagrand stable, then  $A$  has the NIP.*

632     *Proof.* By Theorem 3.10, it suffices to show that  $A$  is relatively countably compact in  $L^0(X, \mu)$  for all Radon probability measure  $\mu$  on  $X$ . Since  $A$  is Talagrand  $\mu$ -stable for any such  $\mu$ , then  $\overline{A} \subseteq L^0(X, \mu)$ . In particular,  $A$  is relatively countably compact in  $L^0(X, \mu)$ .  $\square$

636     **Question 3.18.** Is the converse true?

637     There is a delicate point in this question, as it may be sensitive to set-theoretic axioms (even assuming countability of  $A$ ).

639     **Theorem 3.19** (Talagrand, Theorem 9-3-1(a) in [Tal84]). *Let  $X$  be a compact Hausdorff space and  $A \subseteq M_r(X)$  be countable and pointwise bounded. Assume that  $[0, 1]$  is not the union of  $< c$  closed measure zero sets. If  $A$  has the NIP, then  $A$  is universally Talagrand stable.*

643     **Theorem 3.20** (Fremlin, Shelah, [FS93]). *It is consistent that there exists a countable pointwise bounded set of Lebesgue measurable functions with the NIP which is not Talagrand stable with respect to Lebesgue measure.*

#### 646 APPENDIX: MEASURE THEORY

647     Given a set  $X$ , a collection  $\Sigma$  of subsets of  $X$  is called a  $\sigma$ -algebra if  $\Sigma$  contains  $X$  and is closed under complements and countable unions. Hence, for example, a  $\sigma$ -algebra is also closed under countable intersections. Intuitively, a  $\sigma$ -algebra is a collection of sets in which we can define a  $\sigma$ -additive measure. We call sets in a  $\sigma$ -algebra  $\Sigma$  measurable sets and the pair  $(X, \Sigma)$  a measurable space. If  $X$  is a topological space, there is a natural  $\sigma$ -algebra of subsets of  $X$ , namely the Borel  $\sigma$ -algebra  $\mathcal{B}(X)$ , i.e., the smallest  $\sigma$ -algebra containing all open subsets of  $X$ . Given two measurable spaces  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$ , we say that a function  $f : X \rightarrow Y$  is measurable if and only if  $f^{-1}(E) \in \Sigma_X$  for every  $E \in \Sigma_Y$ . In particular, we say that  $f : X \rightarrow \mathbb{R}$  is measurable if  $f^{-1}(E) \in \Sigma_X$  for all  $E \in \mathcal{B}(\mathbb{R})$  (equivalently,  $E$  open in  $\mathbb{R}$ ).

Given a measurable space  $(X, \Sigma)$ , a  $\sigma$ -additive measure is a non-negative function  $\mu : \Sigma \rightarrow \mathbb{R}$  with the property that  $\mu(\emptyset) = 0$  and  $\mu(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu(A_n)$  whenever  $\{A_n : n \in \mathbb{N}\} \subseteq \Sigma$  is pairwise disjoint. We call  $(X, \Sigma, \mu)$  a *measure space*. A  $\sigma$ -additive measure is called a *probability measure* if  $\mu(X) = 1$ . A measure  $\mu$  is *complete* if for every  $A \subseteq B \in \Sigma$ ,  $\mu(B) = 0$  implies  $A \in \Sigma$ . In words, subsets of measure-zero sets are always measurable (and hence, by the monotonicity of  $\mu$ , have measure zero as well). A measure  $\mu$  is  $\sigma$ -*finite* if  $X = \bigcup_{n=1}^{\infty} X_n$  where  $\mu(X_n) < \infty$  for all  $n \in \mathbb{N}$  (i.e.,  $X$  can be decomposed into countably many finite measure sets). A measure  $\mu$  is *perfect* if for every measurable  $f : X \rightarrow \mathbb{R}$  and every measurable set  $E$  with  $\mu(E) > 0$ , there exists a compact  $K \subseteq f(E)$  such that  $\mu(f^{-1}(K)) > 0$ . We say that a property  $\phi(x)$  about  $x \in X$  holds  $\mu$ -almost everywhere if  $\mu(\{x \in X : \phi(x) \text{ does not hold}\}) = 0$ .

A special example of the preceding concepts is that of a *Radon measure*. If  $X$  is a Hausdorff topological space, then a measure  $\mu$  on the Borel sets of  $X$  is called a *Radon measure* if

- for every open set  $U$ ,  $\mu(U)$  is the supremum of  $\mu(K)$  over all compact  $K \subseteq U$ , that is, the measure of open sets may be approximated via compact sets; and
- every point of  $X$  has a neighborhood  $U \ni x$  for which  $\mu(U)$  is finite.

Perhaps the most famous example of a Radon measure on  $\mathbb{R}$  is the Lebesgue measure of Borel sets. If  $X$  is finite,  $\mu(A) := |A|$  (the cardinality of  $A$ ) defines a Radon measure on  $X$ . Every Radon measure is perfect (see 451A, 451B and 451C in [Fre03]).

While not immediately obvious, sets can be measurable according to one measure, but non-measurable according to another. Given a measure space  $(X, \Sigma, \mu)$  we say that a set  $E \subseteq X$  is  $\mu$ -measurable if there are  $A, B \in \Sigma$  such that  $A \subseteq E \subseteq B$  and  $\mu(B \setminus A) = 0$ . The set of all  $\mu$ -measurable sets is a  $\sigma$ -algebra containing  $\Sigma$  and it is denoted by  $\Sigma_{\mu}$ . A set  $E \subseteq X$  is *universally measurable* if it is  $\mu$ -measurable for every Radon probability measure on  $X$ . It follows that Borel sets are universally measurable. We say that  $f : X \rightarrow \mathbb{R}$  is  $\mu$ -measurable if  $f^{-1}(E) \in \Sigma_{\mu}$  for all  $E \in \mathcal{B}(\mathbb{R})$  (equivalently,  $E$  open in  $\mathbb{R}$ ). The set of all  $\mu$ -measurable functions is denoted by  $\mathcal{L}^0(X, \mu)$ .

Recall that if  $\{X_i : i \in I\}$  is a collection of topological spaces indexed by some set  $I$ , then the product space  $X := \prod_{i \in I} X_i$  is endowed with the topology generated by *cylinders*, that is, sets of the form  $\prod_{i \in I} U_i$  where each  $U_i$  is open in  $X_i$ , and  $U_i = X_i$  except for finitely many indices  $i \in I$ . If each space is measurable, say we pair  $X_i$  with a  $\sigma$ -algebra  $\Sigma_i$ , then there are multiple ways to interpret the product space  $X$  as a measurable space, but the interpretation we care about in this paper is the so called *cylinder  $\sigma$ -algebra*, as used in Lemma 3.8. Namely, let  $\Sigma$  be the  $\sigma$ -algebra generated by sets of the form

$$\prod_{i \in I} C_i, \quad C_i \in \Sigma_i, \quad C_i = X_i \text{ for all but finitely many } i \in I.$$

We remark that when  $I$  is uncountable and  $\Sigma_i = \mathcal{B}(X_i)$  for all  $i \in I$ , then  $\Sigma$  is, in general, strictly **smaller** than  $\mathcal{B}(X)$ .

692

## REFERENCES

- 693 [ADIW24] Samson Alva, Eduardo Dueñez, Jose Iovino, and Claire Walton. Approximability of  
694 deep equilibria. *arXiv preprint arXiv:2409.06064*, 2024. Last revised 20 May 2025,  
695 version 3.
- 696 [ADK08] Spiros A. Argyros, Pandelis Dodos, and Vassilis Kanellopoulos. A classification of sepa-  
697 rable Rosenthal compacta and its applications. *Dissertationes Mathematicae*, 449:1–55,  
698 2008.
- 699 [Ark91] A. V. Arkhangel'skii. *Topological Function Spaces*. Springer, New York, 1st edition,  
700 1991.
- 701 [BD19] Shai Ben-David. Understanding machine learning through the lens of model theory.  
702 *The Bulletin of Symbolic Logic*, 25(3):319–340, 2019.
- 703 [BFT78] J. Bourgain, D. H. Fremlin, and M. Talagrand. Pointwise compact sets of Baire-  
704 measurable functions. *American Journal of Mathematics*, 100(4):845–886, 1978.
- 705 [BKK] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. Preprint.  
706 <https://arxiv.org/abs/1909.01377>, <https://implicit-layers-tutorial.org/>.
- 707 [CRBD] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural  
708 ordinary differential equations. Preprint. <https://arxiv.org/abs/1806.07366>.
- 709 [Deb13] Gabriel Debs. Descriptive aspects of Rosenthal compacta. In *Recent Progress in Gen-  
710 eral Topology III*, pages 205–227. Springer, 2013.
- 711 [Fre00] David H. Fremlin. *Measure Theory, Volume 1: The Irreducible Minimum*. Torres  
712 Fremlin, Colchester, UK, 2000. Second edition 2011.
- 713 [Fre01] David H. Fremlin. *Measure Theory, Volume 2: Broad Foundations*. Torres Fremlin,  
714 Colchester, UK, 2001.
- 715 [Fre03] David H. Fremlin. *Measure Theory, Volume 4: Topological Measure Spaces*. Torres  
716 Fremlin, Colchester, UK, 2003.
- 717 [FS93] David H. Fremlin and Saharon Shelah. Pointwise compact and stable sets of measurable  
718 functions. *Journal of Symbolic Logic*, 58(2):435–455, 1993.
- 719 [GM22] Eli Glasner and Michael Megrelishvili. Todorčević' trichotomy and a hierarchy in the  
720 class of tame dynamical systems. *Transactions of the American Mathematical Society*,  
721 375(7):4513–4548, 2022.
- 722 [Gro52] A. Grothendieck. Critères de compacité dans les espaces fonctionnels généraux. *Amer.  
723 J. Math.*, 74:168–186, 1952.
- 724 [HT23] Clovis Hamel and Franklin D. Tall.  $C_p$ -theory for model theorists. In Jose Iovino, edi-  
725 tor, *Beyond First Order Model Theory, Volume II*, chapter 5, pages 176–213. Chapman  
726 and Hall/CRC, 2023.
- 727 [Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts  
728 in Mathematics*. Springer-Verlag, 1995.
- 729 [Kha20] Karim Khanaki. Stability, nfp, and nsop; model theoretic properties of formulas via  
730 topological properties of function spaces. *Mathematical Logic Quarterly*, 66(2):136–  
731 149, 2020.
- 732 [RPK19] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: a  
733 deep learning framework for solving forward and inverse problems involving nonlinear  
734 partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.
- 735 [Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–  
736 147, 1972.
- 737 [She71] Saharon Shelah. Stability, the f.c.p., and superstability; model theoretic properties of  
738 formulas in first order theory. *Ann. Math. Logic*, 3(3):271–362, 1971.
- 739 [She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories  
740 in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.
- 741 [She78] Saharon Shelah. *Unstable Theories*, volume 187 of *Studies in Logic and the Foun-  
742 dations of Mathematics*. North-Holland, 1978.
- 743 [She90] Saharon Shelah. *Classification theory and the number of nonisomorphic models*, vol-  
744 ume 92 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Pub-  
745 lishing Co., Amsterdam, second edition, 1990.
- 746 [Sim15a] Pierre Simon. *A Guide to NIP Theories*, volume 44 of *Lecture Notes in Logic*. Cam-  
747 bridge University Press, 2015.

- 748 [Sim15b] Pierre Simon. Rosenthal compacta and NIP formulas. *Fundamenta Mathematicae*,  
 231(1):81–92, 2015.
- 749 [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.
- 750 [Tal87] Michel Talagrand. The Glivenko-Cantelli problem. *The Annals of Probability*,  
 15(3):837–870, 1987.
- 751 [Tka11] Vladimir V. Tkachuk. *A  $C_p$ -Theory Problem Book: Topological and Function Spaces*.  
 Problem Books in Mathematics. Springer, 2011.
- 752 [Tod97] Stevo Todorcevic. *Topics in Topology*, volume 1652 of *Lecture Notes in Mathematics*.  
 Springer Berlin, Heidelberg, 1997.
- 753 [Tod99] Stevo Todorcevic. Compact subsets of the first Baire class. *Journal of the American  
 Mathematical Society*, 12(4):1179–1212, 1999.
- 754 [Val84] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*,  
 27(11):1134–1142, 1984.
- 755 [VC71] Vladimir N Vapnik and Alexey Ya Chervonenkis. On the uniform convergence of rel-  
 ative frequencies of events to their probabilities. *Theory of Probability & Its Applica-  
 tions*, 16(2):264–280, 1971.
- 756 [VC74] Vladimir N Vapnik and Alexey Ya Chervonenkis. *Theory of Pattern Recognition*.  
 Nauka, Moscow, 1974. German Translation: Theorie der Zeichenerkennung, Akademie-  
 Verlag, Berlin, 1979.
- 757 [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.
- 758 [WYP22]
- 759 [WYP22]
- 760 [WYP22]
- 761 [WYP22]
- 762 [WYP22]
- 763 [WYP22]
- 764 [WYP22]
- 765 [WYP22]
- 766 [WYP22]
- 767 [WYP22]
- 768 [WYP22]
- 769 [WYP22]
- 770 [WYP22]