DEEP COMPUTATIONS AND NIP

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

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

Abstract. This paper revisits and extends a bridge between functional analysis and model theory, emphasizing its relevance to the theoretical foundations of machine learning. We show that the compactness behavior of families of Baire class 1 functions mirrors the learnability conditions in the sense of *Probably Approximately Correct* (PAC) learning, and that the failure of compactness corresponds to the presence of infinite *Vapnik-Chervonenkis* (VC) dimension. From this perspective, Rosenthal compacta emerge as the natural topological counterpart of PAC-learnable concept classes, while NIP vs. IP structures capture the precise boundary between analytical regularity and combinatorial intractability. These parallels suggest a unified framework linking compactness, definability, and learnability, exemplifying how the topology of function spaces encodes the algorithmic and epistemic limits of prediction.

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 zero, 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 the machine learning literature (e.g., neural ODE's [CRBD] or deep equilibrium models [BKK]). In this paper, we combine ideas of topology and model theory to study these limit phenomena from a more general 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. This allows us to view computations in a given computational model as elements of a space of real-valued functions, called the *space of types* of the model, and thereby to utilize the vast theory of topology of function spaces, known as $\mathsf{C_p}$-theory, to obtain results about complexity of topological limits of computations. As we indicate next, recent classification results for topological spaces of functions provide and elegant and powerful machinery to classify computations according to their level "tameness" or "wildness", with the former corresponding to polynomial approximability and the latter to or exponential approximability. The viewpoint of spaces of types, which we borrow 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 due Shelah [She90]; and in statistical learning, the notion PAC learning and VC dimension pioneered by Vapkins and Chervonenkis [**?**, VC15].

In a previous paper [**?**], we introduced the concept of limits of computations, which we called *ultracomputations* (given they arise as ultrafilter limits of standard computations) and *deep computations* (following usage in machine learning [BKK]).

There is a technical difference between both, but in this paper, to simplify the nomenclature, we will ignore the difference and use only the term "deep computation".

In [**?**], we investigated deep computations (or ultracomputations) that are (real-valued) continuous functions. Under our model-theoretic Rosetta stone, polynomial approximability in the sense of computation becomes identified with the notion of continuous extendability in the sense of topology, and to the notion of *stability* in the sense of model theory.

In this paper, we follow the general approach, i.e., we investigate ultracomputations are pointwise limits of continuous functions. In topology, real-valued functions that arise as the pointwise of a sequence of continuous are called *Baire class 1* functions, or *Baire-1* for short; they form a step above simple continuity in the hierarchy of functions studied in real analysis (Baire class 0 functions being continuous functions). Intuitively, Baire-1 functions represent functions with "controlled" discontinuities, and they are therefore crucial in topology and set theory.

In the first paper, which focused on continuous deep computations, we invoked a classical result of Grothendieck from late 50s [Gro52] to obtain a new polynomial-vs-exponential dichotomy for deep computations. In this paper, which focuses on general Baire-1 computations, we invoke a celebrated result of Todorčević from the late 90s, for Rosenthal compacta [Tod99], to obtain a new trichotomy of general deep computations. Through the aforementioned Rosetta stone, Rosenthal compacta in topology correspond to the important concept of No Independence Property (known as "NIP") in model theory [She71, She90], and to the concept of Probably Approximately Correct learning (known as "PAC learnability") in statistical learning theory [Val84]. We then go beyond Todorčević's trichotomy, and invoke a more recent heptachotomy for minimal families from the early 2000s [**?**]].

We believe that the results presented here show practitioners of computation, or topology or, or model theory, how classification invariants in their field translate into classification invariants of other fields. However, in the interest of accessibility, we do not assume previous familiarity with high level topology or model theory, or computing. The only technical prerequisite of the paper is undergraduate-level topology. The necessary topological preliminaries are included in setction 3.

Throughout the paper, we focus on classical computation; however, the results presented here can be extended, using contemporary model-theortic machinery, to quantum computation and open quantum systems. This extension will be addressed in a forthcoming paper.

2. Motivation

Suppose that A is a subset of the real line $\mathbb{R}$ and that $\overline{A}$ is its *closure*. It is a well-known fact that any point of closure of A, say $x \in \overline{A}$, can be *approximated* by points inside of A, in the sense that a sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq A$ must exist with the property that $\lim_{n\to\infty} x_n = x$. For most applications we wish to approximate objects more complicated than points, such as functions.

Suppose we wish to build a neural network that decides, given an 8 by 8 black-and-white image of a hand-written scribble, what single decimal digit the scribble represents. Maybe there exists f, a function representing an optimal solution to this classifier. Thus if X is the set of all (possible) images, then for $I \in X$, $f(I) \in \{0, 1, 2, \dots, 9\}$ is the "best" (or "good enough" for whatever deployment is needed) possible guess. Training the neural network involves approximating f until its guesses are within an acceptable error range. In general, f might be a function defined on a more complicated topological space X.

Often computers' viable operations are restricted (addition, subtraction, multiplication, division, etc.) and so we want to approximate a complicated function using simple functions (like polynomials). The problem is that, in contrast with mere points, functions in the closure of a set of functions need not be approximable (meaning the pointwise limit of a sequence of functions) by functions in the set.

Functions that are the pointwise limit of continuous functions are *Baire class 1 functions*, and the set of all of these is denoted by $B_1(X)$. Notice that these are not necessarily continuous themselves! A set of Baire class 1 functions, A, will be relatively compact if its closure consists of just Baire class 1 functions (we delay the formal definition of *relatively compact* until Section 3, but the fact mentioned here is sufficient). The Bourgain-Fremlin-Talagrand (BFT) theorem reveals a precise correspondence between relative compactness in $B_1(X)$ and the model-theoretic notion of *Non-Independence Property* (NIP). This was realized by Pierre Simon in [?].

Simon's insight was to view definable families of functions as sets of real-valued functions on type spaces and to interpret relative compactness in $B_1(X)$ as a form of "tame behavior" under ultrafilter limits. From this perspective, NIP theories are those whose definable families behave like relatively compact sets of Baire class 1 functions, avoiding the wild, $\beta\mathbb{N}$-like configurations that witness instability. This observation opened a new bridge between analysis and logic: topological compactness corresponds to the absence of combinatorial independence. Simon's later developments connected these ideas to *Keisler measures* and *empirical averages*, allowing tools from functional analysis to be used to study learnability and definable types. This reinterpretation of model-theoretic tameness through the lens of the BFT theorem has made NIP a central notion not only in stability theory but also in contemporary connections with learning theory and ergodic analysis.

Historically, the notion of NIP arises from Shelah's foundational work on the classification theory of models. In his seminal book *Unstable Theories* [?], Shelah introduced the independence property as a key dividing line within unstable structures, identifying the class of *stable* theories inside those in which this property fails. Fix a first-order formula $\varphi(x, y)$ in a language L and a model M of an L-theory T. We say that $\varphi(x, y)$ has the *independence property (IP)* in M if there is a sequence $(c_i)_{i\in\mathbb{N}} \subseteq M^{|x|}$ such that for every $S \subseteq \mathbb{N}$ there is $a_S \in M^{|y|}$ with

$$\forall i \in \mathbb{N}, \qquad M \models \varphi(c_i, a_S) \quad \iff \quad i \in S.$$

The formula $\phi(x, y)$ has the IP if it does so in some model M, and the formula has the *non-independence property (NIP)* if it does not have the IP. The latter notion of NIP generalizes stability by forbidding the full combinatorial independence pattern while allowing certain controlled forms of unstability. Thus, Simon's interpretation of the BFT theorem can be viewed as placing Shelah's dividing line into a topological-analytic framework, connecting the earliest notions of stability to compactness phenomena in spaces of Baire class 1 functions.

One of the most important innovations in Machine Learning is the mathematical notion, introduced by Turing Awardee Leslie Valiant in the 1980s, of 'probably approximately correct learning', or PAC-learning for short [?]. We give a standard but short overview of these concepts in the context that is relevant to this work.

Consider the following important idea in data classification. Suppose that A is a set and that $\mathcal{C}$ is a collection of sets. We say that $\mathcal{C}$ *shatters* A if every subset of A is of the form $C \cap A$ for some $C \in \mathcal{C}$. For a classical geometric example, if A is the set of four points on the Euclidean plane of the form $(\pm 1, \pm 1)$, then the collection of all half-planes does not shatter A, the collection of all open balls does not shatter A, but the collection of all convex sets shatters A. While A need not be finite, it will usually be assumed to be so in Machine Learning applications. A finer way to distinguish collections of sets that shatter a given set from those that do not is by the *Vapnik-Chervonenkis dimension (VC-dimension)*, which is equal to the cardinality of the largest finite set shattered by the collection, in case it exists, or to infinity otherwise.

A concrete illustration of these ideas appears when considering threshold classifiers on the real line. Let $\mathcal{H}$ be the collection of all indicator functions h_t given by $h_t(x) = 1$ if $x \leq t$ and $h_t(x) = 0$ otherwise. Each h_t is a Baire class 1 function, and the family $\mathcal{H}$ is relatively compact in $B_1(\mathbb{R})$. In model-theoretic terms, $\mathcal{H}$ is NIP, since no configuration of points and thresholds can realize the full independence pattern of a binary matrix. By contrast, the family of parity functions $\{x \mapsto (-1)^{\langle w,x \rangle} : w \in \{0,1\}^n\}$ on $\{0,1\}^n$ (here $\langle w, x\rangle$ is the usual vector dot product) has the independence property and fails relative compactness in $B_1(X)$, capturing the analytical meaning of instability. This dichotomy mirrors the behavior of concept classes with finite versus infinite VC dimension in statistical learning theory.

Going back to the model theoretic framework, let

$$\mathcal{F}_\varphi(M) := \{\varphi(M, a) : a \in M^{|y|}\}$$

be the family of subsets of $M^{|x|}$ defined by instances of the formula φ, where $\varphi(M, a)$ is the set of $|x|$-tuples c in M for which $M \models \varphi(c, a)$. The fundamental theorem of statistical learning states that a binary hypothesis class is PAC-learnable if and only if it has finite VC-dimension, and the subsequent theorem connects the rest of the concepts presented in this section.

Theorem 2.1 (Laskowski)**.** *The formula $\varphi(x, y)$ has the NIP if and only if $\mathcal{F}_\varphi(M)$ has finite VC-dimension.*

For two simple examples of formulas satisfying the NIP, consider first the language $L = \{<\}$ and the model $M = (\mathbb{R}, <)$ of the reals with their usual linear order. Take the formula $\varphi(x, y)$ to mean $x < y$, then $\varphi(M, a) = (-\infty, a)$, and so $\mathcal{F}_\varphi(M)$ is just the set of left open rays. The VC-dimension of this collection is 1, since it can shatter a single point, but no two point set can be shattered since the rays are downwards closed. Now in contrast, the collection of open intervals, given by the formula $\varphi(x; y_1, y_2) := (y_1 < x) \wedge (x < y_2)$, has VC-dimension 2.

In this work, we study the corresponding notions of NIP (and hence PAC-learnability) in the context of Compositional Computation Structures (CCS) introduced in [**?**].

3. General topological preliminaries

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

A *Polish space* is a separable and completely metrizable topological space. The most important examples are the reals $\mathbb{R}$, the Cantor space $2^{\mathbb{N}}$ (the set of all infinite binary sequences, endowed with the product topology), and the Baire space $\mathbb{N}^{\mathbb{N}}$ (the set of all infinite sequences of naturals, also with the product topology). Countable products of Polish spaces are Polish; this includes spaces like $\mathbb{R}^{\mathbb{N}}$, the space of sequences of real numbers. A subspace of a Polish space is itself Polish if and only if it is a G_δ-set, that is, it can be written as the intersection of a countable family of open subsets; in particular, closed subsets and open subsets of Polish spaces are also Polish spaces.

In this work we talk a lot about subspaces, and so there is a pertinent subtlety of the definitions worth mentioning: *completely metrizable space* is not the same as *complete metric space*; for an illustrative example, notice that $(0,1)$ is homeomorphic to the real line, and thus a Polish space (being Polish is a topological property), but with the metric inherited from the reals, as a subspace, $(0,1)$ is **not** a complete metric space. In summary, a Polish space has its topology generated by *some* complete metric, but other metrics generating the same topology might not be. In practice, such as when studying descriptive set theory, one finds that we can often keep the metric implicit.

Given two topological spaces X and Y we denote by $B_1(X,Y)$ the set of all functions $f : X \to Y$ such that for all open $U \subseteq Y$, $f^{-1}[U]$ is an F_σ subset of X (that is, a countable union of closed sets); we call these types of functions *Baire class 1 functions*. When $Y = \mathbb{R}$ we simply denote this collection by $B_1(X)$. We endow $B_1(X,Y)$ with the topology of pointwise convergence (the topology inherited from the product topology of Y^X). By $C_p(X,Y)$ we denote the set of all continuous functions $f : X \to Y$ with the topology of pointwise convergence. Similarly, $C_p(X) := C_p(X,\mathbb{R})$. A natural question is, how do topological properties of X translate to $C_p(X)$ 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'skiĭ and his students in the 1970's and 1980's. This field has found many exciting applications in model theory and functional analysis. Good recent surveys on the topics include [?] and [?]. We begin with the following:

Fact 3.1. *If all open subsets of* X *are* F_σ *(in particular if* X *is metrizable), then* $C_p(X,Y) \subseteq B_1(X,Y)$.

The proof of the following fact (due to Baire) can be found in Section 10 of [?].

Fact 3.2 (Baire)**.** *If* X *is a complete metric space, then the following are equivalent:*

(i) f *is a Baire class 1 function, that is,* $f \in B_1(X)$.
(ii) f *is a pointwise limit of continuous functions.*
(iii) For every closed $F \subseteq X$*, the restriction* $f|_F$ *has a point of continuity.*

Moreover, if X *is Polish and* $f \notin B_1(X)$*, then there exists countable* $D_0, D_1 \subseteq X$ *and reals* $a < b$ *such that* $\overline{D_0} = \overline{D_1}$, $D_0 \subseteq f^{-1}(-\infty, a]$ *and* $D_1 \subseteq f^{-1}[b, \infty)$.

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

Lemma 3.3. *Let* X *be a Polish space and* $A \subseteq B_1(X)$ *be pointwise bounded. The following are equivalent:*

(i) A *is relatively compact in* $B_1(X)$.

(ii) A *is relatively countably compact in* $B_1(X)$*, i.e., every countable subset of* A *has an accumulation point in* $B_1(X)$.

(iii) $\overline{A} \subseteq B_1(X)$*, where* $\overline{A}$ *denotes the closure in* $\mathbb{R}^X$.

Proof. By definition, being pointwise bounded means that there is, for each $x \in X$, $M_x > 0$ such that, for every $f \in A$, $|f(x)| \leq M_x$.

(i)⇒(ii) holds in general.

(ii)⇒(iii) Assume that A is relatively countably compact in $B_1(X)$ and that $f \in \overline{A} \setminus B_1(X)$. By Fact 3.2, there are countable $D_0, D_1 \subseteq X$ with $\overline{D_0} = \overline{D_1}$, and $a < b$ such that $D_0 \subseteq f^{-1}(-\infty, a]$ and $D_1 \subseteq f^{-1}[b, \infty)$. We claim that there is a sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ such that for all $x \in D_0 \cup D_1$, $\lim_{n\to\infty} f_n(x) = f(x)$. Indeed, use the countability to enumerate $D_0 \cup D_1$ as $\{x_n\}_{n\in\mathbb{N}}$. Then find, for each positive n, $f_n \in A$ with $|f_n(x_i) - f(x_i)| < \frac{1}{n}$ for all $i \leq n$. The claim follows.

By relative countable compactness of A, there is an accumulation point $g \in B_1(X)$ of $\{f_n\}_{n\in\mathbb{N}}$. It is straightforward to show that since f and g agree on $D_0 \cup D_1$, g does not have a point of continuity on the closed set $\overline{D_0} = \overline{D_1}$, which contradicts Fact 3.2.

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

3.1. **From Rosenthal's dichotomy to NIP.** The fundamental idea that connects the rich theory here presented to real-valued computations is the concept of an *approximation*. In the reals, points of closure from 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 a remarkably intriguing way. Let us show an example that is actually the protagonist of a celebrated result. Consider the Cantor space $X = 2^{\mathbb{N}}$ and let $p_n(x) = x(n)$ define a continuous mapping $X \to \{0, 1\}$. Then one can show (see Chapter 1.1 of [?] for details) that, perhaps surprisingly, 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. Topologists refer to it as 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.

Theorem 3.4 (Rosenthal's Dichotomy)**.** *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 will either contain a subsequence that converges or a subsequence whose closure is essentially the same as the example mentioned in the previous paragraphs (the worst 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.

If we intend to generalize our results from $C_p(X)$ to the bigger 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 worst possible way (which in this context, is the IP!). The theorem is usually not phrased as a dichotomy but rather as an equivalence (with the NIP instead):

Theorem 3.5 (Bourgain-Fremlin-Talagrand, Theorem 4G in [BFT78])**.** *Let* X *be a Polish space and* $A \subseteq C_p(X)$ *be pointwise bounded. The following are equivalent:*

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

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

Our goal now is to characterize relatively compact subsets of $B_1(X, Y)$ 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 $\pi_P : \mathbb{R}^{\mathcal{P}} \to \mathbb{R}$. 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. In fact, $\mathbb{R}$ and any other Polish space is embeddable as a closed subspace of $\mathbb{R}^{\mathcal{P}}$.

Lemma 3.6. *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 $\mathbb{R}$. Finally,

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

is an F_σ 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_σ. □

Below we consider $\mathcal{P}$ with the discrete topology. For each $f : X \to \mathbb{R}^{\mathcal{P}}$ denote $\hat{f}(P, x) := \pi_P \circ f(x)$ for all $(P, x) \in \mathcal{P} \times X$. Similarly, for each $g : \mathcal{P} \times X \to \mathbb{R}$ 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 $f \in A$.

The map $\left(\mathbb{R}^{\mathcal{P}}\right)^X \to \mathbb{R}^{\mathcal{P}\times X}$ given by $f \mapsto \hat{f}$ is a homeomorphism and its inverse is given by $g \mapsto \check{g}$.

Lemma 3.7. *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. ($\Rightarrow$) Given an open set of reals U, we have that for every $P \in \mathcal{P}$, $f^{-1}[\pi_P^{-1}[U]]$ is F_σ by Lemma 3.6. Given that $\mathcal{P}$ is a discrete countable space, we observe that

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

is also an F_σ set. ($\Leftarrow$) By lemma 3.6 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_σ. □

We now direct our attention to a notion of the NIP that is more general than the one from the introduction. It can be interpreted as a sort of continuous version of the one presented in the preceding section.

Definition 3.8. We say that $A \subseteq \mathbb{R}^X$ has the *Non-Independence Property* (NIP) if and only if for every $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ and for every $a < b$ there are finite disjoint sets $E, F \subseteq \mathbb{N}$ such that

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

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

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

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 3.5.

Theorem 3.9. *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)⇒(2). Let $P \in \mathcal{P}$. Fix $\{f_n\}_{n\in\mathbb{N}} \subseteq A$ and $a < b$. By (1) we have that $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^{\mathcal{P}})$. Applying the homeomorphism $f \mapsto \hat{f}$ and using lemma 3.7 we get $\overline{\hat{A}|_{\mathcal{P}\times K}} \subseteq B_1(\mathcal{P} \times K)$. By Theorem 3.5, 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 compactness, there are finite $E \subseteq I$ and $F \subseteq \mathbb{N}\backslash I$ 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)⇒(1) Fix $f \in \overline{A|_K}$. By lemma 3.6 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 3.5 we have $\overline{\pi_P \circ A|_K} \subseteq B_1(K)$. But then $\pi_P \circ f \in \overline{\pi_P \circ A|_K} \subseteq B_1(K)$. □

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 3.10. *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)⇒(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). □

4. NIP in the context of Compositional Computation Structures

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

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

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-shards. Notice that $\mathcal{L}_{sh}$ is not necessarily equal to $\mathcal{L} = \overline{L}$, unless $\mathcal{P}$ is countable (see [?]). A *transition* is a map $f : L \to L$, in particular, every element in the semigroup Γ is a transition (these are called *realized computations*). In practice, one would like to work with "definable" computations, i.e., ones that can be described by a computer. In this topological framework, being continuous is an expected requirement. However, as in the case of complete-types in model theory, we will work with "unrealized computations", i.e., maps $f : \mathcal{L}_{sh} \to \mathcal{L}_{sh}$. Note that continuity of a computation does not imply that it can be continuously extended to $\mathcal{L}_{sh}$.

Suppose that a transition map $f : L \to \mathcal{L}$ can be extended continuously to a map $\mathcal{L} \to \mathcal{L}$. 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 [?] 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 $\tilde{\gamma} : \mathcal{L}_{sh} \to \mathcal{L}_{sh}$ such that for every sizer $r_\bullet$ there is an $s_\bullet$ such that $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \to \mathcal{L}[s_\bullet]$ is continuous. For a deeper discussion about this axiom, we refer the reader to [?].

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 $\gamma|_{L[r_\bullet]} : L[r_\bullet] \to L[r_\bullet]$ for every $r_\bullet \in R$ and $\gamma \in \Delta$. Elements in Δ are called *real-valued computations* (in this article we will refer to them simply as *computations*) and elements in $\overline{\Delta} \subseteq \mathcal{L}_{sh}^L$ 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 [?].

4.1. **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 4.1. *Let $(\mathsf{L}, \mathcal{P}, \Gamma)$ 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) $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ *for all* $r_\bullet \in \mathsf{R}$.

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

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

Moreover, if any (hence all) of the preceding conditions hold, then every deep computation $f \in \overline{\Delta}$ *can be extended to a Baire-1 function on shards, i.e., there is* $\tilde{f} : \mathcal{L}_{sh} \to \mathcal{L}_{sh}$ *such that* $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ *for all* $r_\bullet \in \mathsf{R}$*. 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 3.9 and Lemma 3.10 prove the equivalence of (1) and (2). If (1) holds and $f \in \overline{\Delta}$, then write $f = \mathcal{U}\lim_i \gamma_i$ as an ultralimit. Define $\tilde{f} := \mathcal{U}\lim_i \tilde{\gamma}_i$. Hence, for all $r_\bullet \in \mathsf{R}$ 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 [?]). □

Given a countable set Δ of computations satisfying the NIP on features and shards (condition (2) of Theorem 4.1) 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 ([?]) culminates in a trichotomy theorem for separable Rosenthal Compacta. Inspired by the work of Glasner and Megrelishvili ([?]), 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 [?]). This suggests the following definition:

Definition 4.2. Let $(\mathsf{L}, \mathcal{P}, \Gamma)$ 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 4.1). We say that Δ is:

(i) NIP_1 if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ is first countable for every $r_\bullet \in \mathsf{R}$.

(ii) NIP_2 if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ is hereditarily separable for every $r_\bullet \in \mathsf{R}$.

(iii) NIP_3 if $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$ is metrizable for every $r_\bullet \in \mathsf{R}$.

Observe that $\mathsf{NIP}_3 \Rightarrow \mathsf{NIP}_2 \Rightarrow \mathsf{NIP}_1 \Rightarrow \mathsf{NIP}$. 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 $\delta_a : 2^{\mathbb{N}} \to \mathbb{R}$ given by $\delta_a(x) = 1$ if $x = a$ and $\delta_a(x) = 0$ otherwise. Let $A(2^{\mathbb{N}}) = \{\delta_a : a \in 2^{\mathbb{N}}\} \cup \{0\}$, where 0 is the zero map. Notice that $A(2^{\mathbb{N}})$ is a compact subset of $B_1(2^{\mathbb{N}})$, in fact $\{\delta_a : a \in 2^{\mathbb{N}}\}$ is a discrete subspace of $B_1(2^{\mathbb{N}})$ and its pointwise closure is precisely $A(2^{\mathbb{N}})$. Hence, this is a Rosenthal compactum which is not first countable. Notice that this space is also not separable.

(2) *Extended Alexandroff compactification.* For each finite binary sequence $s \in 2^{<\mathbb{N}}$, let $\nu_s : 2^{\mathbb{N}} \to \mathbb{R}$ be given by $\nu_s(x) = 1$ if x extends s and $\nu_s(x) = 0$ otherwise. Let $\hat{A}(2^{\mathbb{N}})$ be the pointwise closure of $\{\nu_s : s \in 2^{<\mathbb{N}}\}$, i.e., $\hat{A}(2^{\mathbb{N}}) = A(2^{\mathbb{N}}) \cup \{\nu_s : s \in 2^{<\mathbb{N}}\}$. 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 $f_a^- : 2^{\mathbb{N}} \to \mathbb{R}$ be given by $f_a^-(x) = 1$ if $x < a$ and $f_a^-(x) = 0$ otherwise. Let $f_a^+ : 2^{\mathbb{N}} \to \mathbb{R}$ be given by $f_a^+(x) = 1$ if $x \leq a$ and $f_a^+(x) = 0$ otherwise. The split Cantor is the space $S(2^{\mathbb{N}}) = \{f_a^- : a \in 2^{\mathbb{N}}\} \cup \{f_a^+ : a \in 2^{\mathbb{N}}\}$. This is a separable Rosenthal compactum. One example of a countable dense subset is the set of all f_a^+ and f_a^- where a is an infinite binary sequence that is eventually constant. Moreover, it is hereditarily separable but it is not metrizable.

(4) *Alexandroff Duplicate.* Let K be any compact metric space and consider the Polish space $X = C(K) \sqcup K$, i.e., the disjoint union of $C(K)$ (with its supremum norm topology) and K. For each $a \in K$ define $g_a^0, g_a^1 : X \to \mathbb{R}$ as follows:
$$g_a^0(x) = \begin{cases} x(a), & x \in C(K) \\ 0, & x \in K \end{cases}$$
$$g_a^1(x) = \begin{cases} x(a), & x \in C(K) \\ \delta_a(x), & x \in K \end{cases}$$
Let $D(K) = \{g_a^0 : a \in K\} \cup \{g_a^1 : a \in K\}$. Notice that $D(K)$ is a first countable Rosenthal compactum. It is not separable if K is uncountable. The interesting case will be when $K = 2^{\mathbb{N}}$.

(5) *Extended Alexandroff Duplicate of the split Cantor.* For each finite binary sequence $t \in 2^{<\mathbb{N}}$ let $a_t \in 2^{\mathbb{N}}$ be the sequence starting with t and ending with 0's and let $b_t \in 2^{\mathbb{N}}$ be the sequence starting with t and ending with 1's. Define $h_t : 2^{\mathbb{N}} \to \mathbb{R}$ by
$$h_t(x) = \begin{cases} 0, & x < a_t \\ 1/2, & a_t \leq x \leq b_t \\ 1, & b_t < x \end{cases}$$
Let $\hat{D}(S(2^{\mathbb{N}}))$ be the pointwise closure of the set $\{h_t : t \in 2^{<\mathbb{N}}\}$. Hence, $\hat{D}(S(2^{\mathbb{N}}))$ is a separable first countable Rosenthal compactum which is not hereditarily separable. In fact, it contains an uncountable discrete subspace (see Theorem 5 in [Tod99]).

Theorem 4.3 (Todorčević's Trichotomy, [Tod99], Theorem 3 in [?])**.** *Let K be a separable Rosenthal Compactum.*

(i) If K *is hereditarily separable but non-metrizable, then* $\mathsf{S}(2^{\mathbb{N}})$ *embeds into* K.

(ii) If K *is first countable but not hereditarily separable, then either* $\mathsf{D}(2^{\mathbb{N}})$ *or* $\hat{\mathsf{D}}(\mathsf{S}(2^{\mathbb{N}}))$ *embeds into* K.

(iii) If K *is not first countable, then* $\mathsf{A}(2^{\mathbb{N}})$ *embeds into* K.

In other words, we have the following classification:

$$\begin{tikzcd}[column sep=tiny]
 & \mathsf{K}\text{ is separable Rosenthal compactum} \arrow[dl] \arrow[dr] & & & \\
\mathsf{K}\text{ is metrizable} & & \mathsf{K}\text{ is not metrizable} \arrow[dl] \arrow[dr] & & \\
 & \begin{array}{c}\mathsf{K}\text{ is hereditarily separable}\\ \text{(copy of }\mathsf{S}(\mathrm{I}))\end{array} & & \mathsf{K}\text{ is not hereditarily separable} \arrow[dl] \arrow[dr] & \\
 & & \begin{array}{c}\mathsf{K}\text{ is first countable}\\ \text{(copy of }\mathsf{D}(2^{\mathbb{N}})\text{ or }\hat{\mathsf{D}}(\mathsf{S}(\mathrm{I})))\end{array} & & \begin{array}{c}\mathsf{K}\text{ is not first countable}\\ \text{(copy of }\mathsf{A}(2^{\mathbb{N}}))\end{array}
\end{tikzcd}$$

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

Question 4.4. Is there a non-topological characterization for NIP_i, $i = 1, 2, 3$?

More can be said about the nature of the embeddings in Todorčević's Trichotomy. 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. 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:

Definition 4.5. Given a Polish space X, a countable set I and two pointwise bounded families $\{f_i : i \in \mathrm{I}\} \subseteq \mathbb{R}^{\mathsf{X}}$, $\{g_i : i \in \mathrm{I}\} \subseteq \mathbb{R}^{\mathsf{X}}$ indexed by I. We say that $\{f_i : i \in \mathrm{I}\}$ and $\{g_i : i \in \mathrm{I}\}$ are *equivalent* if and only if the map $f_i \mapsto g_i$ is extended to a homeomorphism from $\overline{\{f_i : i \in \mathrm{I}\}}$ to $\overline{\{g_i : i \in \mathrm{I}\}}$.

Notice that in the separable examples discussed before ($\hat{\mathsf{A}}(2^{\mathbb{N}})$, $\mathsf{S}(2^{\mathbb{N}})$ and $\hat{\mathsf{D}}(\mathsf{S}(2^{\mathbb{N}}))$) 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 [**?**].

Definition 4.6. 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 $\{f_t : t \in 2^{<\mathbb{N}}\}$.

One of the main results in [**?**] is that there are (up to equivalence) seven minimal families of Rosenthal compacta and that for every relatively compact $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq \mathsf{B}_1(\mathsf{X})$ 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 [**?**]. 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 $\mathsf{R} = \{s_t : t \in 2^{<\mathbb{N}}\}$ 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 ν_t be the characteristic function of the set $\{x \in 2^{\mathbb{N}} : x \text{ extends } t\}$. Let $<$ be the lexicographic order in $2^{\mathbb{N}}$. Given $a \in 2^{\mathbb{N}}$, let $f_a^+ : 2^{\mathbb{N}} \to \{0, 1\}$ be the characteristic function of $\{x \in 2^{\mathbb{N}} : a \leq x\}$ and let $f_a^- : 2^{\mathbb{N}} \to \{0, 1\}$ be the characteristic function of $\{x \in 2^{\mathbb{N}} : a < x\}$. Given two maps $f, g : 2^{\mathbb{N}} \to \mathbb{R}$ we denote by $(f, g) : 2^{\mathbb{N}} \sqcup 2^{\mathbb{N}} \to \mathbb{R}$ the function which is f on the first copy of $2^{\mathbb{N}}$ and g on the second copy of $2^{\mathbb{N}}$.

(1) $D_1 = \{\frac{1}{|t|+1}\nu_t : t \in 2^{<\mathbb{N}}\}$. 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) $D_5 = \{\nu_t : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_5} = \hat{A}(2^{\mathbb{N}})$.
(6) $D_6 = \{(\nu_{s_t}, s_t^\frown 0^\infty) : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_6} = \hat{D}(2^{\mathbb{N}})$.
(7) $D_7 = \{(\nu_{s_t}, x^+_{s_t^\frown 0^\infty}) : t \in 2^{<\mathbb{N}}\}$. This is discrete in $\overline{D_7} = \hat{D}(S(2^{\mathbb{N}}))$

Theorem 4.7 (Heptacotomy of minimal families, Theorem 2 in [**?**]). *Let X be Polish. For every relatively compact $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$, there exists $i = 1, 2, \ldots, 7$ and a regular dyadic subtree $\{s_t : t \in 2^{<\mathbb{N}}\}$ of $2^{<\mathbb{N}}$ such that $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$ is equivalent to D_i. Moreover, all D_i are minimal and mutually non-equivalent.*

4.2. **NIP and definability by universally measurable functions.** We now turn to the question: what happens when $\mathcal{P}$ is uncountable? Notice that the countability assumption is crucial in the proof of Theorem 3.9 essentially because it makes $\mathbb{R}^{\mathcal{P}}$ a Polish space. For the uncountable case, we may lose Baire-1 definability so we shall replace $B_1(X)$ by a bigger class. Recall that the purpose of studying the class of Baire-1 functions is that a pointwise limit of continuous functions is not necessarily continuous. In [BFT78], J. Bourgain, D.H. Fremlin and M. Talagrand characterized the Non-Independence Property of a set of continuous functions with various notions of compactness in function spaces containing $C(X)$, such as $B_1(X)$. In this section we will replace $B_1(X)$ with the larger space $M_r(X)$ of universally measurable functions. The development of this section is based on Theorem 2F in [BFT78]. We now give the relevant definitions. Readers with little familiarity with measure theory can review the appendix for standard definitions appearing in this subsection.

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

Lemma 4.8. *Let X be a Hausdorff space and $Y = \prod_{i \in I} Y_i$ be any product of measurable spaces (Y_i, Σ_i) for $i \in I$. Let Σ_Y be the cylinder σ-algebra generated by the measurable spaces (Y_i, Σ_i). Let $f : X \to Y$. The following are equivalent:*

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

Proof. (i)⇒(ii) is clear since the projection maps π_i are measurable and the composition of measurable functions is measurable. To prove (ii)⇒(i), suppose that $C = \prod_{i \in I} C_i$ is a measurable cylinder and let J be the finite set of $i \in I$ such that $C_i \neq Y_i$. 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 measurable set by assumption. □

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

Definition 4.9. Let $(L, \mathcal{P}, \Gamma)$ be a CCS. We say that a transition $f : L \to L$ is *universally measurable shard-definable* if and only if there exists $\tilde{f} : \mathcal{L}_{sh} \to \mathcal{L}_{sh}$ extending f such that for every sizer $r_\bullet$ there is a sizer $s_\bullet$ such that the restriction $\tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \to \mathcal{L}[s_\bullet]$ is universally measurable, i.e. $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \to [-s_P, s_P]$ is μ-measurable for every Radon probability measure μ on $\mathcal{L}[r_\bullet]$.

We will need the following result about NIP and universally measurable functions:

Theorem 4.10 (Bourgain-Fremlin-Talagrand, Theorem 2F in [BFT78])**.** *Let X be a Hausdorff space and $A \subseteq C(X)$ be pointwise bounded. The following are equivalent:*

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

Theorem 3.5 immediately yields the following.

Theorem 4.11. *Let $(L, \mathcal{P}, \Gamma)$ be a CCS satisfying the Extendibility Axiom. Let R be an exhaustive collection of sizers. Let $\Delta \subseteq \Gamma$ be R-confined. If $\pi_P \circ \Delta|_{L[r_\bullet]}$ has the NIP for all $P \in \mathcal{P}$ and all $r_\bullet \in R$, then every deep computation is universally measurable shard-definable.*

Proof. By the Extendibility Axiom, Theorem 3.5 and lemma 3.10 we have that $\overline{\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq M_r(\mathcal{L}[r_\bullet])$ for all $r_\bullet \in R$ and $P \in \mathcal{P}$. Let $f \in \overline{\Delta}$ be a deep computation. Write $f = \mathcal{U}\lim_i \gamma_i$ as an ultralimit of computations in Δ. Define $\tilde{f} := \mathcal{U}\lim_i \tilde{\gamma}_i$. Then, 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 $\pi_P \circ f|_{\mathcal{L}[r_\bullet]} \in \overline{\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}} \subseteq M_r(\mathcal{L}[r_\bullet])$. □

Question 4.12. Under the same assumptions of the previous Theorem, suppose that every deep computation of Δ is universally measurable shard-definable. Must $\pi_P \circ \Delta|_{L[r_\bullet]}$ have the NIP for all $P \in \mathcal{P}$ and all $r_\bullet \in R$?

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

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

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

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

Lemma 4.13. *If* A *is Talagrand* μ*-stable, then* $\overline{A}$ *is also Talagrand* μ*-stable and* $\overline{A} \subseteq \mathcal{L}^0(X, \mu)$.

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

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

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

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

Lemma 4.15 (Fremlin's Dichotomy, 463K in [?])**.** *If* (X, Σ, μ) *is a perfect* σ*-finite measure space (in particular, for* X *compact and* μ *a Radon probability measure on* X*) and* $\{f_n : n \in \mathbb{N}\}$ *be a sequence of real-valued measurable functions on* X*, then either:*

(i) $\{f_n : n \in \mathbb{N}\}$ *has a subsequence that converges* μ*-almost everywhere, or*

(ii) $\{f_n : n \in \mathbb{N}\}$ *has a subsequence with no* μ*-measurable accumulation point in* $\mathbb{R}^X$.

The preceding lemma can be considered as the measure theoretic version of Rosenthal's Dichotomy. Combining this dichotomy with the Theorem 4.10 we get the following result:

Theorem 4.16. *Let* X *be a Hausdorff space and* $A \subseteq C(X)$ *be pointwise bounded. The following are equivalent:*

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

Proof. Notice that the equivalence (i)-(iii) is Theorem 4.10. Notice that the equivalence of (iii) and (iv) is Fremlin's Dichotomy Theorem. □

Lemma 4.17. *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.*

Proof. By Theorem 4.10, it suffices to show that A is relatively countably compact in $\mathcal{L}^0(X, \mu)$ for all Radon probability measure μ on X. Since A is Talagrand μ-stable for any such μ, then $\overline{A} \subseteq \mathcal{L}^0(X, \mu)$. In particular, A is relatively countably compact in $\mathcal{L}^0(X, \mu)$. □

Question 4.18. Is the converse true?

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

Theorem 4.19 (Talagrand, Theorem 9-3-1(a) in [**?**]). *Let* X *be a compact Hausdorff space and* $A \subseteq M_r(X)$ *be countable and pointwise bounded. Assume that* $[0, 1]$ *is not the union of* $< \mathfrak{c}$ *closed measure zero sets. If* A *has the NIP, then* A *is universally Talagrand stable.*

Theorem 4.20 (Fremlin, Shelah, [**?**]). *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.*

Appendix: Measure Theory

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

$f : X \to \mathbb{R}$ 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, Σ), a σ-*additive measure* is a non-negative function $\mu : \Sigma \to \mathbb{R}$ 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, Σ, μ) a *measure space*. A σ-additive measure is called a *probability measure* if $\mu(X) = 1$. A measure μ 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 μ, have measure zero as well). A measure μ is σ-*finite* if $X = \cup_{n=1}^{\infty} X_n$ 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 μ is *perfect* if for every measurable $f : X \to \mathbb{R}$ 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 μ-*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 μ 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 [**?**]).

While not immediately obvious, sets can be measurable according to one measure, but non-measurable according to another. Given a measure space (X, Σ, μ) we say that a set $E \subseteq X$ is μ-*measurable* if there are $A, B \in \Sigma$ such that $A \subseteq E \subseteq B$ and $\mu(B \backslash A) = 0$. The set of all μ-measurable sets is a σ-algebra containing Σ and it is denoted by Σ_μ. A set $E \subseteq X$ is *universally measurable* if it is μ-measurable for every Radon probability measure on X. It follows that Borel sets are universally measurable. We say that $f : X \to \mathbb{R}$ is μ-*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 μ-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 σ-algebra Σ_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* σ-*algebra*, as used in Lemma 4.8. Namely, let Σ be the σ-algebra generated by sets of the form

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

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

References

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

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

[CRBD] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. Preprint. https://arxiv.org/abs/1806.07366.

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

[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.

[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.

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

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

[VC15] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. pages 11–30, 2015. Reprint of Theor. Probability Appl. **16** (1971), 264–280.