Topological and Algebraic Aspects of Algorithmic Learning Theory

Matthew de Brecht

Graduate School of Informatics
Kyoto University

Supervisor: Akihiro Yamamoto

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Abstract

Algorithmic learning theory is a subfield of machine learning that attempts to give a mathematical formalization of the notion of "learning" and investigate the learnability and complexity of various learning problems. One well known model is the "identification in the limit" paradigm, which was introduced by E. M. Gold. In this paradigm, a learner receives information about some concept, and the learner guesses hypotheses to explain the concept. If the learner converges to a correct hypothesis after making only a finite number of incorrect guesses, then we say that the learner has identified the concept in the limit. If every concept in a collection of concepts, called a concept space, can be identified in the limit, then the concept space is said to be identifiable in the limit. In this thesis, we add algebraic and topological structure to concept spaces and investigate the relationship between identifiability in the limit and various algebraic and topological properties.

As for algebraic properties, we use algebraic closure operators and lattice theory to characterize sufficient conditions for identifiability in the limit. In particular, we introduce an algebraic closure operator for any given concept space. Intuitively, such an operator determines all the information that can be extracted from partially given information about some unknown concept. This operator embeds the concept space into an algebraic closure system, and the lattice theoretical aspects of this closure system are shown to be closely related to the identifiability and complexity of identifying the concept space in the limit. As an application, we use closure operators to construct a learning algorithm for the class of unbounded unions of restricted pattern languages, and characterize the complexity of this class using ordinals.

By naturally interpreting a concept space as a topological space, we were able to give topological characterizations of several necessary and sufficient conditions for identifiability. In addition, we clarified which structural properties of concept spaces are topologically invariant. In particular, identifiability in the limit is a topologically invariant property, so when an abstract topological space is interpreted as a concept space, whether or not the space is identifiable in the limit is independent of how the topological space is interpreted as a concept space. We have also given complete characterizations of reductions between learning problems using continuous functions. In some cases, it is possible to construct more intuitive learning algorithms for a concept space by continuously reducing it into a simpler space.

Finally, we analyzed the types of representations of concept spaces. In particular, we introduce a hierarchy of representations called $\mathbf{\Sigma}^0_\alpha$-admissible representations, and characterize which functions are realizable with respect to different levels of representations. This hierarchy is a generalization of "admissible representations," which have become important to the field of computable analysis. We show that the stream of information provided about an unknown concept is often $\mathbf{\Sigma}^0_1$-admissible with respect to a suitable topology on the concept space, and that the learner's sequence of hypotheses is a $\mathbf{\Sigma}^0_2$-admissible representation of the concept space with respect to the discrete topology. Based on these results, we can provide a general framework that includes classification in the limit, identification in the limit, and other variations of these learning paradigms.

Contents

Chapter 1

Introduction

In this thesis we will analyze connections between algorithmic learning theory, topology, and algebra. We will mainly focus on the identification in the limit model of learning proposed by E. M. Gold [21] in 1967. This is an abstract model of the learning process, where a learner receives information about some concept and attempts to converge to an explicit hypothesis for the concept. From this standpoint, learning is interpreted as the process of constructing a complete description of a concept when only partial information is available. Although somewhat idealistic and oversimplified, the identification in the limit model is unique as being one of the few attempts to create a mathematically rigorous framework in which the learning process can be studied. Thus the model is important not only for clarifying what we mean by "learning", but also for better understanding the limitations of learning agents and intelligent agents in general.

The goal of this thesis is to better understand the identification in the limit model using tools from algebra (particularly universal algebra) and general topology. The purpose of our approach is twofold. First, several aspects of the identification in the limit model are clarified by using algebraic and topological concepts. This will be the goal of Chapters 2 and 3, where we show that notions like identifiability, mind-change complexity, and reducibility have very simple algebraic and topological characterizations. These characterizations not only make it easier to apply the model in practice, but also help bring less explicit aspects of the model into the foreground. For example, the role that partial orders and topology play in the identfication in the limit model show that the intuitive notion of information in this model is closely related to the way information is viewed in domain theory. Since learning theory and domain theory have very different origins and goals, these unexpected connections are useful for understanding how humans conceptualize abstract notions like "information".

The second purpose of our approach is to abstract away superficial aspects of the model and try to get to the core of what the identification in the limit is actually modeling. This is a long distance attack at the problem of obtaining a *theory* of learning, as opposed to simply a collection of *models* of learning. This is one of the underlying goals of Chapter 4, where we reduce the identification in the limit model to representations of topological spaces and functions between the spaces. From this perspective, the learner's goal is to "compute" a function between topological spaces, where our use of the word "compute" here is consistent with the Type 2 Theory of Effectivity [59]. By varying the representations and the functions, several variations of the identification in the limit model can be easily reproduced. Although we only focus on direct descendents of Gold's original model, our approach may provide useful insight for unifying other "Platonic" learning models that view concepts as existing independently of the learner.

In addition to generalizing the learning model framework, our abstraction has emphasized the importance of better understanding how concepts are represented in terms of streams of information. This notion of representation includes not only the stream of information a learner receives about a concept (i.e., the learner's input), but also in what sense a learner's sequence of hypotheses (i.e., the learner's output) provides a representation of the concept. If learning or identifying a concept is to be anything more than just personal intellectual gratification for the learner, then we must clarify in what sense the hypotheses produced by the learner have greater utility than the original input stream of data. The result of the learning process is a hypothesis which we expect the learner will use in the future to achieve further goals. Clearly, if a hypothesis is correct, then it provides a complete description of a concept in a compact form that is more suitable for future computations or decisions than an incomplete sequence of facts about the concept. On the other hand, in general there is no guarantee that the learner's current hypothesis is correct, and consequently no guarantee that any computations the learner might make using the current hypothesis will be correct. The abstraction of representations in Chapter 4 provides a topological perspective of this tradeoff in terms of the topology that a representation induces on a set of concepts and the level of discontinuity of the representation. Although we have not yet investigated other important aspects of representations, such as data compression, probabilistic error, and computational complexity, our approach provides a unique perspective on the price that must be paid when attempting to construct a complete hypothesis from partial information.

1.1 Outline and Background

Since the goal of this thesis is to analyze connections between topology, algebra, and learning theory, it is difficult to present the ideas in a linear order without reference to one another. We therefore make a compromise by first introducing learning theory and topology in the remainder of this Introduction, then discuss the algebraic aspects in Chapter 2 so that by the time we get to Chapter 3 the reader will be familiar with enough of the ideas that we can better show how they are interrelated.

The ideas presented in the remainder of this Introduction are for the most part unoriginal. Algorithmic learning theory was introduced by E. M. Gold [21] in 1967 and intensively studied by many others ever since. Topology goes back even further, and even the more recent interpretations of topology in terms of "observable" properties have been proposed by many domain theorists like Smyth [56], Abramsky [1], and Battenfeld, et al. [7]. The only part of our introduction that might be considered original is the meshing of these two fields together. For example, our definition of a "learner" differs from Gold's in that we use the topological notion of continuous functions instead of Turing machines. However, our formalization of learning theory is essentially equivalent to Gold's model, except that we will tend to ignore computability restraints.

The algebraic aspects of learning theory will be discussed in Chapter 2. The main contribution of this chapter is the introduction of an algebraic closure operator that embeds a concept space into an algebraic closure system. This allows one to view an arbitrary concept space as a subset of a larger algebraic system, and then one can use these algebraic properties to help solve the learning problem. In particular, our results give an abstract generalization of previous work by Stephan and Ventsov [57] on the learnability of algebraic structures. Chapter 2 will also investigate partial orders and their order types, and show how they are related to the complexity of learning. As an application of these ideas, we will give close bounds on the complexity of learning unbounded unions of a restricted family of pattern languages.

Then in Chapter 3 we will return to topology and its role in learning theory. Connections between topology and learning theory have been investigated by several researchers, such as Kelly [31], Luo and Schulte [34], and Martin et al. [36]. Whereas Kelly and Martin et al. are more focused on the "classification in the limit" paradigm in learning theory, we will focus on the "identification in the limit" paradigm. The former deals with determining whether or not a particular property holds for an infinite sequence of numbers. The latter can be thought of as trying to determine which *equivalence class* a given infinite sequence belongs. The two paradigms are thus closely related, but different mathematical structures play different roles. The relationship between these two paradigms will be discussed more in Chapter 4.

Luo and Schulte showed some applications of topology in the identification in the limit model. In particular, they gave a topological characterization of the mind-change complexity of learning a concept from positive data. Chapter 3 extends Luo and Schulte's work by giving topological characterizations of many other sufficient criteria for learnability. Luo and Schulte also briefly noted a relationship between continuous functions and reducibility between learning problems, and we will give a full characterization of this relationship. We also give characterizations of properties of concept spaces that are purely topological, and analyze the topological aspects of representations of concept spaces. These results are important because, if a characterization of some property is topologically invariant, then results concerning the property can be easily applied in other variations of the learning model. In particular, we can easily convert Luo and Schulte's characterization of mind-change complexity for learning from positive data into a characterization for learning from positive and negative data.

In addition to the technical results just mentioned, a major contribution of Chapter 3 is the way that we treat topology as fundamental in learning theory. In particular, our characterization of the learning problem allows us to treat the learner, concept space, representations of concepts, and outputted hypotheses as topological objects, which in several cases allows simpler and more intuitive proofs than the papers just cited.

Chapter 4 generalizes the identification in the limit model to the more general problem of computing between represented sets. This generalization allows us to include other learning models, such as the classification in the limit model mentioned above, within a single coherent framework. The work in this chapter is heavily influenced and compatible with the Type II model of effectivity and approach to computable analysis introduced by Weihrauch [59].

The major contribution of Chapter 4 is the introduction and analysis of $\mathbf{\Sigma}^0_\alpha$-admissible representations, which generalize the admissible representations introduced by Weihrauch [59] and Schröder [48] for computable analysis. This includes a large class of ways of representing sets of objects, as we give characterizations of the functions between such sets that are "computable". Chapter 4 generalizes work on the realizability of discontinuous functions and "limit-computable" functions by computable analysists such as V. Brattka [8], M. Ziegler [62], and Brattka and Makananise [9].

The last two sections of this chapter introduce algorithmic learning theory and topology. We have placed more effort in these two sections in intuitively explaining the motivation of our formalization of learning theory and the role that topology plays. The body of later chapters will be more technical, but we will precede each chapter with a brief intuitive explanation of the ideas and results.

1.2 Algorithmic learning theory

By the term "algorithmic learning theory" we are referring to the collection of formal models of learnability that were inspired by E. M. Gold's 1967 paper "Language identification in the limit" [21]. The goal of these models is to provide a formal definition for the word "learn", and then to understand when learning is possible and how complex such a learning task is. These models are characterized by the assumption that there is a learner (usually a Turing machine) that receives information about some unknown concept, and occasionally outputs hypotheses to explain the concept. The learner successfully learns the concept if it eventually converges to some hypothesis that is reasonably correct.

In this section, we will introduce the basic identification in the limit model. In particular, we give formal definitions to words like "concept" and "learner", and give some of the motivation for the way we formalize these terms. Our formalization of the identification in the limit model is slightly more abstract than Gold's original definition, although it results in an essentially equivalent model. The advantage of our formalization is in its generality and that it will make it easier to analyze the topological aspects of the model. Although it is typical in algorithmic learning theory to require learners to be computable functions, we will not do so here. We do, however, formalize the model in a way that computability requirements can be naturally introduced if desired.

1.2.1 Concept spaces

The abstract notion of a concept will be handled formally by using the extension of the concept, which is the set of all objects or properties belonging to the concept. For example, the concept "prime integer" can be interpreted as the set of all prime integers. The precise definition of a concept is then heavily dependent on the set of objects and properties, or *universe*, we are considering. Since a major goal of algorithmic learning theory is to analyze learnability with respect to Turing machines, it is convenient to assume that our universe is a countable set, which implies that all concepts are countable sets. This means that we cannot define the concept "irrational real" as the set of all irrational real numbers, but we *can* define "irrational real" as the set of all sentences defining irrational numbers within some formal (countable) language.

Since we have required the universe to be countable, we can encode it as (a subset of) the natural numbers, which we denote by ω. Therefore, we will not lose any generality by simply assuming that the universe is ω. We can now formally define a *concept* to be any subset of ω.

A set of concepts is called a *concept space*. If we specify some "rule" for interpreting some of our intuitive concepts formally as subsets of natural numbers, then the set of concepts that can be formalized in this way determines a concept space. For example, we can assign a number from 1 to 52 to each card in a deck of playing cards. Then for each of the properties in the categories suit (heart, club, spade, diamond), rank (ace, two, three, four, five, six, seven, eight, nine, ten, jack, queen, king), and color (red, black) we can assign the subset of numbers that encodes all of the cards having the given property. Taking all boolean combinations of suits, ranks, and colors determines a concept space of the basic properties of playing cards. This determines a straight forward way to formalize concepts like "either a heart or a diamond", "two of spades", or "not a red king". By defining concepts in this way, we can analyze the relationships between concepts by the structure of the concept space. For example, the observation that "every diamond is red" is captured by the fact that the concept "diamond" is a subset of the concept "red".

As in the example above with playing cards, it is convenient to give the concepts names or descriptions like "red" or "spade". If a concept is a recursive set, then the Gödel number of a function that computes the characteristic function of the concept is often used as a description. We generalize these notions by defining a *hypothesis space* for a concept space $\mathcal{L}$ to be a pair $\langle \mathcal{H}, h \rangle$ with $\mathcal{H} \subseteq \omega$ and $h\colon \mathcal{H} \to \mathcal{L}$ a surjective function. An element $H \in \mathcal{H}$ is called a *hypothesis* or a *name* for $h(H) \in \mathcal{L}$. The surjectiveness of h guarantees that every concept has a name. Note that if a concept space has a hypothesis space then it can only have countably many concepts. Returning to the playing cards example, we can define a hypothesis space by encoding meaningful descriptions like "either a heart or diamond" as natural numbers, and then define h to map the description to the appropriate set of cards.

1.2.2 Texts and informants

Once we have determined a concept space, we must next determine how information about the concepts is represented to the learner. Once again, since we are concerned with the applicability of our theory to Turing machines, we will assume that information comes in discrete packets that have been encoded as natural numbers. We can then imagine that our learner receives a sequence of natural numbers encoding pieces of information about the concept. Such a sequence will be called a *representation* of the concept. For formal reasons, it will be convenient to assume that the representation is infinite, but finite representations can be modelled by using some number as an "end of input" marker, and letting the remainder of the sequence be arbitrary.

Let ω^ω denote the set of all infinite sequences of natural numbers. A *representation* of a concept space $\mathcal{L}$ is a pair $\langle \mathcal{R}, \rho \rangle$ where $\mathcal{R} \subseteq \omega^\omega$ and $\rho\colon \mathcal{R} \to \mathcal{L}$ is a surjective function. An element $R \in \mathcal{R}$ is called a *representation* for $\rho(R) \in \mathcal{L}$.

Note that a hypothesis space can also be interpreted as a representation by replacing each $H \in \mathcal{H}$ by the sequence consisting of infinitely many copies of H.

We next introduce two types of representations that will play a major role in this thesis. We will use $\#$ as a special symbol ($\# \notin \omega$) which will be used to mean a "pause" or "no information". Given a set X let X^ω be the set of all countably infinite sequences of elements of X. For $\xi \in X^\omega$, $\xi(n)$ will denote the $(n+1)$-th element in the sequence ξ, and $\xi[i]$ will denote the initial subsequence of length i of ξ (i.e., $\xi[i] = \langle \xi(0), \xi(1), \ldots, \xi(i-1) \rangle$). In particular, $\xi[0]$ is the *empty sequence*, which we denote by ε.

Let $\mathcal{L}$ be a concept space and $L \in \mathcal{L}$ a concept. A *text* for L is a sequence $T \in (\omega \cup \{\#\})^\omega$ such that $L = \{i \in \omega \mid \exists n \in \omega : T(n) = i\}$. Intuitively, a text for L is a sequence enumerating exactly the elements of L, with occasional pauses denoted by $\#$. A text for L is also sometimes called a *positive presentation* of L. We define

$$\mathcal{T}(\mathcal{L}) = \{T \in (\omega \cup \{\#\})^\omega \mid \exists L \in \mathcal{L} : T \text{ is a text for } L\}$$

and define the function $\tau_\mathcal{L} : \mathcal{T}(\mathcal{L}) \to \mathcal{L}$ so that

$$\tau_\mathcal{L}(T) = \{i \in \omega \mid \exists n \in \omega : T(n) = i\}.$$

Thus, $\mathcal{T}(\mathcal{L})$ is the set of all texts for all concepts in $\mathcal{L}$, and $\tau_\mathcal{L}$ maps a text to the concept in $\mathcal{L}$ that it represents.

An *informant* for a concept $L \in \mathcal{L}$ is a sequence $I \in ((\omega \times \{0,1\}) \cup \{\#\})^\omega$ such that for every $i \in \omega$ either $\langle i, 0 \rangle$ or $\langle i, 1 \rangle$ (but not both) occurs in I and $L = \{i \in \omega \mid \exists n \in \omega : I(n) = \langle i, 1 \rangle\}$. We define $\mathcal{I}(\mathcal{L})$ to be the set of all informants for all concepts in $\mathcal{L}$, and define $\iota_\mathcal{L} : \mathcal{I}(\mathcal{L}) \to \mathcal{L}$ to be the function that maps each informant to the concept it represents. Intuitively, an informant for a concept is a sequence listing all of the natural numbers along with flags indicating whether or not each number is in the concept. Informants are also sometimes called *complete presentations*. Elements of the form $\langle i, 1 \rangle$ are usually called *positive data* and elements of the form $\langle i, 0 \rangle$ are called *negative data*.

To aid intuition we have defined texts and informants so that they are not elements of ω^ω, but it is trivial to encode them as such and the reader should imagine that such an encoding has been done. We can therefore view $\langle \mathcal{T}(\mathcal{L}), \tau_\mathcal{L} \rangle$ and $\langle \mathcal{I}(\mathcal{L}), \iota_\mathcal{L} \rangle$ as two different representations of $\mathcal{L}$.

There is a sense in which informants are more *informative* about the concept they represent than texts, because they tell us not only which elements are in the concept, but also which elements are *not* in the concept. Thus, given an informant for a concept, we can produce a text for the concept by simply filtering out the negative data. However, if we assume that the learner only has access to a finite initial segment of a representation at any time, then it is not always possible to produce an informant given a text.

The notion of a text, however, is more general than that of an informant. For example, given a concept space $\mathcal{L}$, we can re-encode each concept $L \in \mathcal{L}$ as

$$\hat{L} = \{\langle i, 1 \rangle \mid i \in L\} \cup \{\langle i, 0 \rangle \mid i \notin L\}$$

and define the concept space $\hat{\mathcal{L}} = \{\hat{L} \mid L \in \mathcal{L}\}$. Then the texts for $\hat{\mathcal{L}}$ are exactly the informants for $\mathcal{L}$. Therefore, any results concerning texts also apply to informants. For this reason, we will be mainly concerned with texts, and informants will play only a minor role.

1.2.3 Learners

We next give a formal definition of a learner. Within our framework, we think of a learner as some agent that receives a representation of a concept and outputs hypotheses to explain the concept. It is possible that the learner may output a hypothesis and, after receiving more information, realize that the original hypothesis was mistaken. The learner can then output a new hypothesis that better explains the information received. Therefore, a learner can be viewed as a particular function from infinite streams of information to infinite streams of hypotheses. Requiring that the streams be infinite is only a mathematical convenience, and, as we mentioned before, a finite stream can be modeled by marking the end of it with some special symbol and allowing the remainder of the stream to be arbitrary.

We will assume that the hypotheses that the learner can choose from and their interpretation as concepts will be predetermined. This means that the learner must choose hypotheses from some hypothesis space $\langle \mathcal{H}, h \rangle$ for some concept space $\mathcal{L}$. This requirement is to prevent the learner from outputing hypotheses like "the concept that is being presented to me now" which would have different interpretations depending on the situation and thus little practical value.

Since the learner's choice of hypotheses is restricted, we will be mainly concerned with the behaviour of the learner when it receives information about a concept that it can describe by some available hypothesis. Therefore, in addition to the hypothesis space $\langle \mathcal{H}, h \rangle$ for $\mathcal{L}$, we will assume some representation $\langle \mathcal{R}, \rho \rangle$ of $\mathcal{L}$ that determines the kind of information the learner will receive.

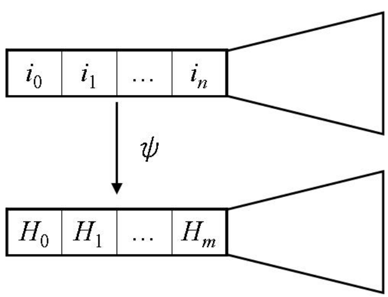


Figure 1.1: A finite portion of a learner ψ's output only depends on a finite portion of its input.

A learner is then a function from $\mathcal{R}$ (streams of information) to $\mathcal{H}^\omega$ (sequences of hypotheses). We do not, however, expect that every function from $\mathcal{R}$ to $\mathcal{H}^\omega$ is a learner. In particular, although a learner receives an infinite sequence of information, at any point in time it can only have access to some *finite* initial segment of the sequence. Therefore, if the learner outputs a sequence $H_0, H_1, \ldots, H_m$ of hypotheses in response to seeing some representation $R \in \mathcal{R}$ of a concept, then there must be some $n \in \omega$ such that the learner has only seen the initial finite sequence $\sigma = \langle i_0, i_1, \ldots, i_n \rangle$ of R when it outputted its mth hypothesis. This implies that given *any* representation of a concept that begins with σ the sequence of hypotheses outputted by the learner will always begin with the finite sequence $H_0, H_1, \dots, H_m$ (see Figure 1.1). In other words, every finite portion of the learner's output depends on only a finite portion of its input. We will call a function that satisfies this property a *continuous* function. This is the only property that we will require every learner to have. Formally, we define a *learner* to be any continuous function $\psi: \mathcal{R} \to \mathcal{H}^\omega$, where $\mathcal{R}$ and $\mathcal{H}$ are respectively a representation and a hypothesis space for some concept space $\mathcal{L}$.

1.2.4 Identification in the limit

We now give a criterion to define a *successful* learner. The sequence of hypotheses output by a successful learner should approach the concept that is being presented to the learner. This means that if the learner is seeing a representation R of a concept, its goal is to find a hypothesis H such that $h(H)$ (the concept H describes) is a good approximation of $\rho(R)$ (the concept being presented). In the *identification* model of learning, we require that the learner finds a hypothesis H such that $\rho(R) = h(H)$. This is a very strict criterion, but it results in very many interesting learning problems.

An infinite sequence $H_0, H_1, \dots$ of hypotheses *converges* to a hypothesis H if and only if there is $m \in \omega$ such that $H_n = H$ for all $n \geq m$. In other words, a sequence of hypotheses converges to H if and only if all but a finite number of hypotheses in the sequence are equal to H. In this case, we will also say that H is the *limit* of the sequence $H_0, H_1, \dots$.

Let $\mathcal{L}$ be a concept space and $\langle \mathcal{R}, \rho \rangle$ and $\langle \mathcal{H}, h \rangle$ be a representation and a hypothesis space of $\mathcal{L}$, respectively. If $\psi: \mathcal{R} \to \mathcal{H}^\omega$ is a learner and $L \in \mathcal{L}$ is a concept, then we say that ψ *identifies* L *in the limit* (with respect to $\langle \mathcal{R}, \rho \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if for every representation R of L, $\psi(R)$ converges to a hypothesis H such that $h(H) = \rho(R) = L$. In other words, ψ identifies L in the limit if and only if when ψ is given any representation of L, ψ outputs at most a finite number of hypotheses describing a concept different from L before converging to a correct hypothesis for L. We say that ψ *identifies* $\mathcal{L}$ *in the limit* (with respect to $\langle \mathcal{R}, \rho \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if ψ identifies every concept of $\mathcal{L}$ in the limit. In other words, ψ identifies $\mathcal{L}$ in the limit if and only if for every $R \in \mathcal{R}$, $\psi(R)$ converges to some $H \in \mathcal{H}$ such that $\rho(R) = h(H)$. A concept space $\mathcal{L}$ is *identifiable in the limit* or *learnable in the limit* (with respect to $\langle \mathcal{R}, \rho \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if there exists a learner ψ that identifies $\mathcal{L}$ in the limit.

Identification in the limit with respect to the representation $\langle \mathcal{T}(\mathcal{L}), \tau_{\mathcal{L}} \rangle$ of $\mathcal{L}$ is called *identification in the limit from positive data*. If we use $\langle \mathcal{I}(\mathcal{L}), \iota_{\mathcal{L}} \rangle$ as a representation then we say *indentification in the limit from positive and negative data*.

Finally, we introduce an important characterization due to Angluin [3] of concept spaces that are identifiable in the limit from positive data.

Definition 1.2.1 *Let $\mathcal{L}$ be a concept space and let $L \in \mathcal{L}$ be a concept. A* finite tell-tale *of L is a finite set $F \subseteq L$ such that for every $L' \in \mathcal{L}$, if $F \subseteq L'$ then L' is* not *a proper subset of L.* □

Theorem 1.2.2 (Angluin [3]) *Let $\mathcal{L}$ be a concept space and $\langle \mathcal{H}, h \rangle$ a hypothesis space for $\mathcal{L}$. Then $\mathcal{L}$ is identifiable in the limit from positive data with respect to $\langle \mathcal{H}, h \rangle$ if and only if every concept in $\mathcal{L}$ has a finite tell-tale.* □

Example 1.2.3 Let $L_n = \{m \in \omega \,|\, m \leq n\}$ and $\mathcal{L} = \{L_n \,|\, n \in \omega\}$. Then a text for L_n is just an infinite sequence composed of all of the natural numbers less than or equal to n (with occasional pauses denoted by "#"). Each $L_n \in \mathcal{L}$ has $\{n\}$ as a finite tell-tale, so $\mathcal{L}$ is identifiable in the limit from positive data. A learner that outputs L_m as hypothesis, where m is the largest natural number that has appeared so far in the text, will identify $\mathcal{L}$ in the limit.

Next consider $\mathcal{L}^+ = \mathcal{L} \cup \{\omega\}$. Then ω has no finite tell-tale in $\mathcal{L}^+$, because for any finite $F \subseteq \omega$, $F \subseteq L_{\max(F)}$ and $L_{\max(F)}$ is a proper subset of ω. Therefore, $\mathcal{L}^+$ is not identifiable in the limit from positive data. Assume for a contradiction that ψ identifies $\mathcal{L}^+$ in the limit. Start feeding ψ a text T_0 for ω. Since ψ identifies $\mathcal{L}^+$, at some point it must output a hypothesis for ω. However, since ψ's output only depends on a finite portion of its input, ψ can only have seen some initial segment σ_0 of the text T_0 when it outputs ω as hypothesis. If we let n_1 be the largest number that appears in σ_0, then σ_0 can be extended to a text T_1 for L_{n_1}. Thus, there must be some finite initial segment σ_1 of T_1 such that σ_1 extends σ_0 and ψ outputs L_{n_1} as hypothesis after reading in σ_1. Clearly, σ_1 can be extended to a text for ω, and we can repeat the above process of forcing ψ to change hypotheses between ω and some proper subset of ω indefinitely. Thus, we can construct a text for ω on which ψ never converges, which is a contradiction. $\square$

1.3 Observable properties and Topology

In this section we give some basic definitions from general topology. The main goal here is to give some intuitive explanation of the definitions, so that the reader can better understand how topology is related to learning theory as we have defined it in the previous section. In particular, we will view topology as a formalization of the notion of the "observable properties" of a system. This perspective is not new, and the relationship between topology and observable properties has been investigated by many domain theorists, in particular Smyth [56], Abramsky [1], and Battenfeld, et al. [7]. Our discussion of observable properties is heavily based on these references, except that our discussion will be from the perspective of learning theory. The most important reference is [7] because it extends the analogy between topology and observable sets to represented spaces, which is crucial in the learning theory framework we use.

1.3.1 Observable properties

Given a class of objects, a *property* of the objects is interpreted as the set of all objects that have the property. For example, if the class of objects is "animals", then the property "mammalian" will be interpreted as the set of all mammals. If the class of objects is "Turing machines", then "never halts on input i" is the set of all Turing machines that fail to halt when given i as input. Similar to concepts, the precise definition of a property will depend on the class, or *type*, of the objects to which it is being applied.

Of particular interest are the properties that we can actually observe in practice. For example, if we present a horse to a trained biologist, we expect that the biologist will be able to observe that the horse is a mammal within a finite amount of time. If we gave the biologist a frog instead of a horse, then the biologist will not mistakenly claim that the frog is a mammal. Since this distinction can be made, we consider the property mammalian to be observable.

On the other hand, the property "never halts on input i" of Turing machines is *not* observable. Even if a Turing machine with input i does not halt after one hundred years, there is still no guarantee that it will *never* halt. We can, however, observe when a Turing machine *does* halt.

We define an *observable property* to be a property which, when presented with any object that has the property we can determine in finite time that the object has the property, and when presented with an object that does not have the property we will not mistakenly claim that it has the property.

Note that the definition of an observable property is one-sided, because we do not require that we can always observe when a property is *not* present. So "halts" is observable while "does not halt" is not observable, even though they are complements of each other.

Our definition of observable property has been somewhat informal, but we can make it more formal using the learning theory framework introduced in the previous section. Given a concept space $\mathcal{L}$, a property of concepts in $\mathcal{L}$ is just any subset P of $\mathcal{L}$. If we are also given a representation $\langle \mathcal{R}, \rho \rangle$ of $\mathcal{L}$, then we can say a property P is observable (with respect to $\langle \mathcal{R}, \rho \rangle$) if and only if, given any representation $R \in \mathcal{R}$ of a concept, if $\rho(R) \in P$ then we can conclusively say so within a finite amount of time, and if $\rho(R) \notin P$ then we will never mistakenly say to the contrary.

It is important to notice that this definition of observability depends on the representation of a concept space. For example, if the representation is $\langle \mathcal{T}(\mathcal{L}), \tau_{\mathcal{L}} \rangle$, then the set of concepts containing the number i is observable with respect to this representation, but we cannot always observe when a concept does not contain i. However, it is possible to observe that a concept does not contain i with respect to the representation $\langle \mathcal{I}(\mathcal{L}), \iota_{\mathcal{L}} \rangle$. Therefore, different representations result in different notions of observability.

1.3.2 The information ordering of properties

The goal of a learner in the identification paradigm is to "observe" the identity of the concept being presented. However, the property "is equal to L" is in general not observable for an arbitrary concept $L \in \mathcal{L}$. Instead, the learner must rely on the properties of concepts that *are* observable and use these observations to produce better and better hypotheses.

Since the learner has a definite goal, namely to identify the concept being presented, it is clear that some observable properties of concepts are more useful or informative than others. For example, the property "is a concept in $\mathcal{L}$" (the property $\mathcal{L}$) is trivially observable because by assumption every representation is for a concept in $\mathcal{L}$. However, if the learner could observe the property "is equal to L" (the property $\{L\}$), then the learner would have successfully completed the learning problem. In general, observing a property that only applies to a few number of concepts is more helpful in identifying the concept. We can therefore assign a natural ordering, which we call the *information ordering*, to properties by saying property P_1 is *more informative* than P_2 if and only if P_1 is a strict subset of P_2. For example, in our earlier example with playing cards, observing that a card is a heart is more informative than observing that it is red.

The idea of an information ordering is common in domain theory, where partially ordered sets are used to model the semantics of computation [20]. Note that we only order properties according to their informativeness, and do not assign a measure of the information content as is done in Shannon's theory of information [53].

However, our notion of an information ordering is compatible with Shannon's theory in the following sense. For simplicity, we will assume that $\mathcal{L}$ is a finite set of concepts. Let p be any probability measure on $\mathcal{L}$, such that every singleton subset of $\mathcal{L}$ has non-zero probability. Let L_* be a random variable with values in $\mathcal{L}$. Intuitively, L_* is the result of randomly choosing a concept from $\mathcal{L}$ according to the probability measure p. The *entropy* of L_*, denoted $H(L_*)$, is defined as

$$H(L_*) = -\sum_{L\in\mathcal{L}} p(L_* = L)\log_2 p(L_* = L)$$

and has units in *bits*. Intuitively, if the entropy of L_* is high, then we have greater uncertainty about which concept will be chosen. A property $P \subseteq \mathcal{L}$ can be viewed as a random variable with values *yes* and *no*, depending on whether or not the property holds for a randomly chosen concept. Then the *conditional entropy* of L_* given that $P = yes$ is defined as

$$H(L_*|P = yes) = -\sum_{L\in\mathcal{L}} p(L_* = L|P = yes)\log_2 p(L_* = L|P = yes).$$

The *information gain* from observing that the property P holds, which we denote $I(L_*|P = yes)$, is then defined as $H(L_*) - H(L_*|P = yes)$. The information gain gives us a measurement in bits of how much our uncertainty about the identity of L_* decreases when we know that property P holds. We can then compare the informativeness of properties by how much information we gain by observing the property.

For example, let $\mathcal{L} = \{\{n\} \mid 1 \leq n \leq 52\}$ where we assume each concept $\{n\}$ in $\mathcal{L}$ corresponds to a particular playing card. Assuming p is the uniform distribution on $\mathcal{L}$, we have that

$$H(L_*) = -\sum_{L\in\mathcal{L}} \frac{1}{52}\log_2\frac{1}{52} = \log_2 52 \approx 5.7 \text{ bits}$$

Since there are 26 red cards and 13 hearts in a deck, it is straight forward to calculate that

$$H(L_*|red = yes) = \log_2 26 \approx 4.7 \text{ bits}$$

and

$$H(L_*|heart = yes) = \log_2 13 \approx 3.7 \text{ bits}.$$

So observing that a card is red gives us an information gain of one bit, and observing that a card is a heart gives us a gain of two bits. Therefore, observing that a card is a heart is more informative than only observing that it is red, which is consistent with our information ordering of properties.

In fact, the additivity of a probability measure guarantees that for *any* probability measure on $\mathcal{L}$ such that every concept has non-zero probability, if $P_1 \subset P_2$ are properties, then $H(L_*|P_1 = yes) > H(L_*|P_2 = yes)$. This implies that if P_1 is more informative than P_2 according to our information ordering, then according to Shannon's theory the information gain of observing P_1 to hold will be greater than the gain from observing P_2 to hold, *regardless* of the probability distribution on $\mathcal{L}$ (assuming every concept has non-zero probability).

Now, assuming the uniform distribution on a deck of playing cards, the information gain of observing a card is red is the same as observing that it is black, whereas red and black are incomparable under our information ordering. However, if we use a "stacked" deck, where black cards are drawn with greater frequency, then the information gain of red becomes strictly larger than that of black. In general, if P_1 and P_2 are incomparable under our information ordering, then there must be $L_1 \in P_1 \setminus P_2$ and $L_2 \in P_2 \setminus P_1$ (where $X \setminus Y$ is the set of elements in X and not in Y). We can then define two different probability measures p_1 and p_2 on $\mathcal{L}$ where in the first case L_1 occurs with very high probability and in the second case L_2 occurs with very high probability. If p_1 and p_2 are chosen carefully, then Shannon's theory will say that P_1 has strictly less information gain than P_2 with respect to p_1, whereas the opposite holds for p_2. Thus, the information gain of P_1 and P_2 will be incomparable under different probability measures.

Therefore, our information ordering is not only consistent with Shannon's theory, but it is arguably the *best* characterization of information content that one can do without assuming further structure, like probability measures, on the concept space.

1.3.3 Topological space

We have defined "observable property" and have ordered properties by their informativeness. Our next goal is to better understand the structure of observable properties and their relationship to topology.

Let X be a set of objects and let $\mathcal{O}(X)$ be the set of observable properties of objects in X. It is clear that both X and $\emptyset$, the empty set, are in $\mathcal{O}(X)$. In the first case, since every object is in X, we can immediately declare that the property X holds. In the second case, we can immediately declare that the property does not hold.

Now, if P_1 and P_2 are observable properties, then we can observe their intersection $P_1 \cap P_2$ by simply first observing P_1 and then observing P_2. This argument can be extended to show that the intersection of any *finite* collection of observable properties is also observable. We cannot, however, argue that the intersection of an *infinite* collection of observable properties is necessarily observable. This is because there is no way in general to bound the amount of time it takes to observe a property, so it may not be possible to observe all of the properties in an infinite collection within a finite amount of time.

However, arbitrary *unions* of observable properties are observable. If we are given a collection $\mathcal{U} \subseteq \mathcal{O}(X)$, then to observe $\bigcup \mathcal{U} = \{x \in X \mid \exists P \in \mathcal{U} : x \in P\}$, we only need to observe that any single property $P \in \mathcal{U}$ holds.

The above observations imply that the set of observable properties of X determines a *topology* on X.

Definition 1.3.1 *A* topological space *is a pair $\langle X, \mathcal{O}(X)\rangle$, where X is a set and $\mathcal{O}(X)$ is a set of subsets of X such that the following hold:*

1. *$\emptyset, X \in \mathcal{O}(X)$;*
2. *If $U_0, U_1, \ldots, U_n \in \mathcal{O}(X)$, then $U_0 \cap U_1 \cap \cdots \cap U_n \in \mathcal{O}(X)$;*

3. *If* $\mathcal{U} \subseteq \mathcal{O}(X)$ *then* $\bigcup \mathcal{U} \in \mathcal{O}(X)$.

$\mathcal{O}(X)$ *is called a* topology *on* X*, and the elements of* $\mathcal{O}(X)$ *are called* open sets. *A subset* $A \subseteq X$ *is a* closed set *if and only if* $A = X \setminus U$ *for some* $U \in \mathcal{O}(X)$. *A subset of* X *is a* clopen set *if and only if it is both open and closed.* □

We will often simply say that X is a topological space when $\mathcal{O}(X)$ can be omitted without causing any confusion.

It follows that the observable properties of a set correspond to open sets. If it can be observed when a property *does not* hold, then the property corresponds to a closed set. If it can both be observed when a property holds and it can also be observed when it does not hold, then the property corresponds to a clopen set.

As an example, we shall introduce a topology on ω^ω that is consistent with what we consider to be the observable properties of infinite sequences of natural numbers. We let $\omega^{<\omega}$ denote the set of all finite sequences of natural numbers, and for $\sigma \in \omega^{<\omega}$ and $\xi \in \omega^\omega$, we write $\sigma \prec \xi$ to mean that σ is a finite initial segment of ξ.

Now, for $\sigma \in \omega^{<\omega}$, consider the property

$$\uparrow\sigma = \{\xi \in \omega^\omega \mid \sigma \prec \xi\}$$

corresponding to all infinite sequences that extend σ. We imagine that infinite sequences are presented to us one element at a time. We can observe that a sequence ξ extends σ by just waiting long enough until we have seen the initial finite segment of ξ that has the same length as σ and then comparing the two finite sequences. Therefore, $\uparrow\sigma$ should be considered an observable property of ω^ω.

Next assume that $P \subseteq \omega^\omega$ is observable. If we are presented with some $\xi \in \omega^\omega$ and observe that ξ has property P, then since only a finite amount of time has passed, we could only have seen some finite initial segment σ of ξ when we declared that P holds. Therefore, since we never make a mistake when we say that an observable property holds, it follows that every other $\xi' \in \omega^\omega$ that extends σ must also have property P. This means that $\uparrow\sigma \subseteq P$. Now if for every $\xi \in P$ we let σ_ξ be the finite initial segment of ξ that we have seen when we declare P to hold for ξ, then we can conclude that $P = \bigcup_{\xi \in P} \uparrow\sigma_\xi$.

Thus, $\uparrow\sigma$ is observable for every $\sigma \in \omega^{<\omega}$ and every observable property is equal to a union of observable properties of the form $\uparrow\sigma$. The topology on ω^ω that is generated by taking the closure under finite intersection and arbitrary unions of sets of the form $\uparrow\sigma$ for $\sigma \in \omega^{<\omega}$ is called the *Baire topology* on ω^ω. Throughout this thesis, we will always view ω^ω as a topological space with the Baire topology.

Since we have formally defined representations $\langle \mathcal{R}, \rho \rangle$ so that $\mathcal{R} \subseteq \omega^\omega$, every representation naturally inherits a topology from ω^ω, called the *subspace topology*.

Definition 1.3.2 *Let* Y *be a topological space and* $X \subseteq Y$*. The* subspace topology *on* X *is defined so that* $U \subseteq X$ *is open if and only if* $U = X \cap V$ *for some open set* $V \subseteq Y$. □

Throughout this thesis, for any representation $\langle \mathcal{R}, \rho \rangle$, we will always view $\mathcal{R}$ as being a topological space with the subspace topology inherited from ω^ω.

Therefore, the set of texts $\mathcal{T}(\mathcal{L})$ and informants $\mathcal{I}(\mathcal{L})$ (properly encoded as subsets of ω^ω) are also topological spaces. It is easy to see that the topology on a representation $\mathcal{R}$ is generated by sets of the form $\mathcal{R} \cap (\uparrow \sigma)$ just like the Baire topology. If it is clear from context that we are dealing with a particular representation $\mathcal{R}$, then we will just write $\uparrow \sigma$ instead of $\mathcal{R} \cap (\uparrow \sigma)$.

Now that we have interpreted representations as topological spaces, it is straight forward to interpret a concept space $\mathcal{L}$ with a representation $\langle \mathcal{R}, \rho \rangle$ as a topological space. Recall that we have defined a property $P \subseteq \mathcal{L}$ to be observable if and only if we can declare within a finite amount of time that P holds given any representation of a concept in P, and if we will not declare P to hold if we are given a representation of a concept that is not in P. This means that $P \subseteq \mathcal{L}$ is observable if and only if $\rho^{-1}(P) = \{\xi \in \mathcal{R} \mid \rho(\xi) \in P\}$ is an observable property of $\mathcal{R}$.

Definition 1.3.3 *Let X be a topological space, Y a set, and $f\colon X \to Y$ a surjective function. The* quotient topology *on Y with respect to f is defined so that $U \subseteq Y$ is open if and only if $f^{-1}(U)$ is open in X.* □

For any concept space $\mathcal{L}$ with a representation $\langle \mathcal{R}, \rho \rangle$, we can then naturally view $\mathcal{L}$ as a topological space with the quotient topology determined by $\langle \mathcal{R}, \rho \rangle$. The quotient topology then gives a natural characterization of the properties of $\mathcal{L}$ that we consider to be observable with respect to the representation. It is important to note that the topology on $\mathcal{L}$ depends on the reprsentation $\langle \mathcal{R}, \rho \rangle$.

The topology gives us additional structure that we can use to analyze the properties of a represented concept space. This additional structure essentially comes for free, depending only on our intuitive notion of which properties of infinite sequences are observable in a finite amount of time.

1.3.4 Continuous functions

Consider a function $f\colon X \to Y$ between two sets X and Y of objects, and assume we have some way of physically implementing the function so that given an object $x \in X$, we can in some sense "compute" the value $f(x) \in Y$. Now, if P is an observable property of Y, then we can observe the property $f^{-1}(P)$ of X by simply computing f and observing that P holds for the result. This motivates the following definition.

Definition 1.3.4 *A function $f\colon X \to Y$ between topological spaces is* continuous *if and only if $f^{-1}(U) \in \mathcal{O}(X)$ for every $U \in \mathcal{O}(Y)$.* □

Recall that we have defined a learner to be any function $\psi\colon \mathcal{R} \to \mathcal{H}^\omega$ that is continuous. Since both $\mathcal{R}$ and $\mathcal{H}^\omega$ are viewed as subspaces of ω^ω, the intuitive notion of continuity that we used when defining learners is easily seen to be equivalent to the formal definition we have just given. For example, it is clear that "the nth hypothesis equals H" is an observable property of $\mathcal{H}^\omega$, so the subset of $\mathcal{R}$ of all representations for which ψ outputs H as the nth hypothesis must also be observable.

Continuity can be thought of as a *necessary* requirement for the learner to be "physically feasible" in the sense of G. Plotkin [42]. Whether or not continuity is sufficient to guarantee that a learner can be implemented by a physical system will depend on many factors, so to keep our discussion as general as possible we will only require learners to be continuous.

1.4 Notation

Throughout this thesis, it will be important to distinguish between a set X being a subset of a set Y and being a *proper* subset of Y. We will therefore write $X \subseteq Y$ to denote that X is a subset of Y, and $X \subset Y$ to denote that $X \subseteq Y$ and $X \neq Y$.

We will write $f\colon\subseteq X \to Y$ to mean that f is a *partial function* from the set X to the set Y. The domain of a partial function f (i.e., the set $\{x \in X \mid f(x) \text{ is defined}\}$) will be denoted $dom(f)$. The range of a (partial or total) function f (i.e., the set $\{y \in Y \mid \exists x \in X : f(x) = y\}$) will be denoted $range(f)$.

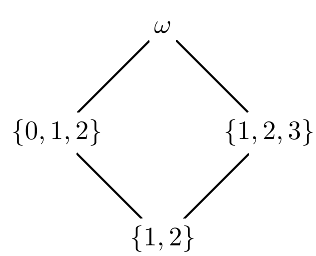
Other notations will be introduced as they become needed, and can be found in the index.

Chapter 2

Algebraic Properties of Concept Spaces

In this chapter we will introduce algebraic structure to a concept space and investigate the relationship between algebraic properties of the concept space and identification in the limit. In particular, we associate an algebraic closure operator to each concept space, which embeds the concept space into an algebraic lattice. Properties of the algebraic closure operator and the algebraic lattice provide sufficient criteria for guaranteeing that a concept space is identifiable in the limit, and are also useful for characterizing the difficulty of identifying the concept space in the limit.

Intuitively, if we observe that some unknown concept contains the set of elements F, then the *closure* of F, $C(F)$, is the largest set of elements that we can guarantee also belongs to the unknown concept. For example, if a concept space only contains the two concepts $\{0, 1, 2\}$ and $\{1, 2, 3\}$, and we observe that the unknown concept contains 3, then we can be sure that the concept also contains 1 and 2, so $C(\{3\}) = \{1, 2, 3\}$. On the other hand, if we only observe that the concept contains 1, then we can only conclude that it also contains 2, so $C(\{1\}) = \{1, 2\}$. Since $C(\{3\}) \supset C(\{1\})$, we can think of observing 3 to be more informative than observing 1. In fact, observing 1 or 2 is not informative at all, because we know from the beginning that every concept contains 1 and 2 (in other words, $C(\emptyset) = \{1, 2\}$). In this example, C produces the four closed sets $\{1, 2\}$, $\{0, 1, 2\}$, $\{1, 2, 3\}$, and ω, which form a lattice when ordered by set inclusion. Here, ω is added as a top element in order to form a lattice, and in this example ω can intuitively be understood as a "state of contradiction." For example, $C(\{0, 3\}) = \omega$ because no concept contains both 0 and 3.



We will show in this chapter under which circumstances the closure operator can be used as a learning algorithm by simply outputting as a hypothesis the closure of the elements seen. This type of learning strategy can be imagined as the learner climbing up the lattice of partial states of knowledge until it eventually converges to the correct concept. The "height" of this lattice can sometimes be measured using transfinite ordinals, and provides a measure of the difficulty of identifying the concept space in the limit.

In the next section we introduce algebraic closure operators and embed concept spaces into algebraic closure systems. In Section 2.2 we use transfinite ordinals to measure the complexity of identifying concept spaces in the limit, and give characterizations of the complexity in terms of order-theoretical properties of the concept spaces. In Section 2.3 we apply the results in this chapter to give upper and lower bounds on the complexity of identifying unbouned unions of restricted pattern languages in the limit. Most results from this chapter have been presented in [13] and [14].

2.1 Algebraic closure operators

2.1.1 Preliminaries

In this subsection we define algebraic closure operators and briefly discuss their properties. Everything presented here is well known, and we only include a couple of proofs to help the reader become more familiar with the concepts involved.

Closure operators and closure systems

Given a set X, we denote the power set of X by $\mathcal{P}(X)$.

Definition 2.1.1 *Let U be a set, and let $C\colon \mathcal{P}(U) \to \mathcal{P}(U)$ be a mapping on the powerset of U. C is called an* algebraic closure operator *on U if the following conditions hold for all subsets X and Y of U:*

1. $X \subseteq C(X)$ *(extensive)*
2. $C(X) = C(C(X))$ *(idempotent)*
3. $X \subseteq Y$ *implies* $C(X) \subseteq C(Y)$ *(monotone)*
4. $C(X) = \bigcup\{C(F) \mid F$ *is a finite subset of* $X\}$ *(finitary)*

An operator that only fulfills the first three conditions above is simply called a closure operator*. A* closed set *or a* fixed point *of C is a set X such that $X = C(X)$. A* finitely generated closed set *is a set X such that $X = C(F)$ for some finite subset F of U.* □

The set of all fixed points of an algebraic closure operator is called an *algebraic closure system*. Algebraic closure systems naturally occur in many areas of abstract algebra (see [11]). The set of all subgroups of a group G and the set of all ideals of a ring R are two well known examples of algebraic closure systems. Algebraic closure operators also play an important role in abstract algebraic logic [10], where they are used to model the "logical consequence" of a set of formulas. It is also easily seen that $\mathcal{P}(\omega)$, the power set of the natural numbers, is an algebraic closure system on ω.

Lattice structure of closure systems

Definition 2.1.2 *A* partially ordered set*, or* poset*, is a pair $\langle P, \leq \rangle$ consisting of a set P and a binary relation $\leq$ on P that is* reflexive *($x \leq x$),* transitive *($x \leq y$ and $y \leq z$ implies $x \leq z$), and* anti-symmetric *($x \leq y$ and $y \leq x$ implies $x = y$).* □

If $\langle P, \leq \rangle$ is a poset and $S \subseteq P$, then $p \in P$ is the *supremum* of S if and only if $s \leq p$ for all $s \in S$ (i.e., p is an *upper bound* of S), and if $q \in P$ is any other upper bound of S then $p \leq q$ (i.e., p is the *least upper bound* of S). Dually, we can define the *infimum*, or *greatest lower bound* of S. The supremum (infimum) of S, if it exists, is denoted $\bigvee S$ ($\bigwedge S$).

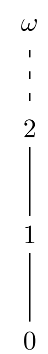
Definition 2.1.3 *A poset $\langle L, \leq \rangle$ is a* complete lattice *if and only if every subset S of L has both a supremum ($\bigvee S$) and an infimum ($\bigwedge S$) in L.* □

An algebraic closure system $\mathcal{C}$ forms a complete lattice (ordered by subset inclusion) where:

$$\bigvee_{i \in I} C(X_i) = C\left(\bigcup_{i \in I} X_i\right) \qquad \text{and} \qquad \bigwedge_{i \in I} C(X_i) = \bigcap_{i \in I} C(X_i).$$

Given a poset P, a non-empty subset D of P is called a *directed set* if for all $x, y \in D$, there exists $z \in D$ such that $x \leq z$ and $y \leq z$.

Definition 2.1.4 *Let L be a lattice and let x be an element of L. We say that x is* compact *if and only if for every directed set $D \subseteq L$ such that $x \leq \bigvee D$, there exists an element $d \in D$ such that $x \leq d$.* □



The above diagram shows all ordinals up to and including ω. The ordinals 0, 1, 2, ... are compact, but ω is *not* compact. Compact elements are sometimes called *finite* or *finitary* elements. The compact elements of an algebraic closure system are precisely the finitely generated closed sets. To show that every compact element $X \in \mathcal{C}$ is finitely generated, just note that $\mathcal{D} = \{C(F) \mid F \subseteq X \text{ is finite}\}$ is directed with $X \subseteq \bigvee \mathcal{D}$, so we must have $X \subseteq C(F)$ for some finite $F \subseteq X$, which of course implies that $X = C(F)$. The converse can be shown by using the following proposition, which is important in its own right.

Proposition 2.1.5 *If $\mathcal{C}$ is an algebraic closure system and $\mathcal{D} \subseteq \mathcal{C}$ is directed, then $\bigvee \mathcal{D} = \bigcup \mathcal{D}$. In particular, the union of any directed subset of $\mathcal{C}$ is an element of $\mathcal{C}$.*

Proof: By definition, $\bigvee \mathcal{D} = C(\bigcup \mathcal{D})$, so $\bigcup \mathcal{D} \subseteq \bigvee \mathcal{D}$ follows because C is extensive. To show that $\bigvee \mathcal{D} \subseteq \bigcup \mathcal{D}$, let x be any element of $\bigvee \mathcal{D}$. Since C is finitary, there must be finite $F \subseteq \bigcup \mathcal{D}$ such that $x \in C(F)$. Since F is finite, there must be a finite sequence $D_0, D_1, \ldots, D_n$ of closed sets in $\mathcal{D}$ such that $F \subseteq D_0 \cup D_1 \cup \cdots \cup D_n$. Recursively using the fact that $\mathcal{D}$ is directed, there exists some $D \in \mathcal{D}$ such that $D_i \subseteq D$ for $0 \leq i \leq n$. Therefore, $F \subseteq D$, so $x \in C(F) \subseteq C(D) = D$. Since $D \in \mathcal{D}$, it follows that $x \in \bigcup \mathcal{D}$, hence $\bigvee \mathcal{D} \subseteq \bigcup \mathcal{D}$. □

Now if $X \in \mathcal{C}$ is such that $X = C(F)$ for some finite $F \subseteq X$, and $X \subseteq \bigvee \mathcal{D}$ for some directed $\mathcal{D} \subseteq \mathcal{L}$, then since $\bigvee \mathcal{D} = \bigcup \mathcal{D}$ we must have that $F \subseteq \bigcup \mathcal{D}$. Using the fact that F is finite and $\mathcal{D}$ is directed, there must be $D \in \mathcal{D}$ that contains F, so the monotonicity of C implies that $X = C(F) \subseteq C(D) = D$. Thus, $X \subseteq D$, and since $\mathcal{D}$ was arbitrary, X is compact.

2.1.2 Embedding concept spaces into closure systems

We now introduce an algebraic closure operator on ω with respect to a given concept space $\mathcal{L}$.

Definition 2.1.6 *Let $\mathcal{L}$ be a concept space. For any subset S of ω, let*

$$C_{\mathcal{L}}(S) = \bigcup \{ \bigcap \{ L \in \mathcal{L} \mid F \subseteq L \} \mid F \textit{ is a finite subset of } S \}.$$

Define $\mathcal{A}(\mathcal{L})$ to be the set of fixed points of $C_{\mathcal{L}}(\cdot)$. □

In the above definition, we follow the convention that $\bigcap \emptyset = \omega$. It is clear from the definition that $C_{\mathcal{L}}(L) = L$ for all $L \in \mathcal{L}$, so $\mathcal{L} \subseteq \mathcal{A}(\mathcal{L})$. The algebraic closure system $\mathcal{A}(\mathcal{L})$ is special because it is the *smallest* algebraic closure system on ω that contains $\mathcal{L}$.

Example 2.1.7 Let $L_n = \{m \in \omega \mid m \leq n\}$ and $\mathcal{L} = \{L_n \mid n \in \omega\}$. Then $C_{\mathcal{L}}(F) = L_{\max(F)}$ for any non-empty finite $F \subseteq \omega$, $C_{\mathcal{L}}(\emptyset) = L_0$, and $C_{\mathcal{L}}(S) = \omega$ for any infinite $S \subseteq \omega$. Therefore, $\mathcal{A}(\mathcal{L}) = \mathcal{L} \cup \{\omega\}$. □

Theorem 2.1.8 *If $\mathcal{C}$ is an algebraic closure system on ω such that $\mathcal{L} \subseteq \mathcal{C}$, then $\mathcal{A}(\mathcal{L}) \subseteq \mathcal{C}$. In particular, $\mathcal{A}(\mathcal{A}(\mathcal{L})) = \mathcal{A}(\mathcal{L})$.*

Proof: Assume $C\colon \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ is the algebraic closure operator that generates $\mathcal{C}$. Since $\mathcal{L} \subseteq \mathcal{C}$ by assumption, we must have that $C(L) = L$ for all $L \in \mathcal{L}$. We want to show that any $X \subseteq \omega$ satisfying $C_{\mathcal{L}}(X) = X$ also satisfies $C(X) = X$, which would imply that $\mathcal{A}(\mathcal{L}) \subseteq \mathcal{C}$. So assume, for a contradiction, that $C_{\mathcal{L}}(X) = X$ but $C(X) \neq X$, which, because C is extensive, implies that X is a proper subset of $C(X)$. Let x be any element in $C(X) \setminus X$. Since C is finitary, there must be a finite $F \subseteq X$ such that $x \in C(F)$. But since $x \notin X$, $x \notin C_{\mathcal{L}}(F)$ because $C_{\mathcal{L}}(F) \subseteq C_{\mathcal{L}}(X) = X$. By definition of $C_{\mathcal{L}}$, this means that x is not in $\bigcap \{L \in \mathcal{L} \mid F \subseteq L\}$. If no $L \in \mathcal{L}$ contained F, then $\{L \in \mathcal{L} \mid F \subseteq L\}$ would be the empty set and so x would be in $\omega = \bigcap \emptyset$, a contradiction. So there must be at least some $L \in \mathcal{L}$ containing F, and since x is not in $\bigcap \{L \in \mathcal{L} \mid F \subseteq L\}$, there must be some $L \in \mathcal{L}$ that contains F but does not contain x. But then $F \subseteq L$ and $x \in C(F)$ and $x \notin L$, so $C(F) \not\subseteq L = C(L)$, contradicting the monotonicity of C. Therefore, we must conclude that $C(X) = X$, which completes the proof that $\mathcal{A}(\mathcal{L}) \subseteq \mathcal{C}$.

Now, since $\mathcal{A}(\mathcal{L})$ is clearly an algebraic closure system containing $\mathcal{A}(\mathcal{L})$, it follows from what we have just proved that $\mathcal{A}(\mathcal{A}(\mathcal{L})) \subseteq \mathcal{A}(\mathcal{L})$. Therefore, $\mathcal{A}(\mathcal{A}(\mathcal{L})) = \mathcal{A}(\mathcal{L})$. □

Intuitively, given a set $S \subseteq \omega$, $C_{\mathcal{L}}(S)$ is the largest subset of ω that we can be sure is included in each concept of $\mathcal{L}$ that contains S, given that we can only inspect finite subsets of S at a given time. Thus, given two finite subsets F and G of some unknown concept $L \in \mathcal{L}$, we can think of F as containing "more information" about L than G if $C_{\mathcal{L}}(G) \subset C_{\mathcal{L}}(F)$.

2.1.3 Compact elements and characteristic sets

We next show an application in the identification in the limit from positive data model. If $T \in \mathcal{T}(\mathcal{L})$ is a text for some concept in L, then for each $i \in \omega$, $X_i = C_{\mathcal{L}}(\{n \in \omega \mid \exists j \leq i : T(j) = n\})$ is the *largest* subset of ω that we can guarantee to be a subset of L by only inspecting the first i elements of the text T. We also have that $\{X_i\}_{i\in\omega}$ is an ascending sequence of closed sets in $\mathcal{A}(\mathcal{L})$ such that $L = \bigcup_{i\in\omega} X_i$. Now, if L is compact in $\mathcal{A}(\mathcal{L})$, then since $\{X_i\}_{i\in\omega}$ is a directed family of subsets of L, there must be $i \in \omega$ such that $L \subseteq X_i$. Since $\{X_i\}_{i\in\omega}$ is an ascending sequence of subsets of L, it follows that $L = X_i$, and in fact $L = X_j$ for all $j \geq i$.

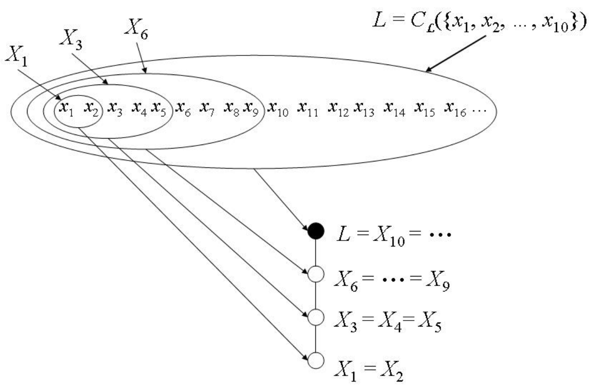


Figure 2.1: Example of taking the closure of elements appearing in a text for a compact concept $L \in \mathcal{L}$.

Based on the above observation, we can easily prove the following lemma.

Lemma 2.1.9 *If every $L \in \mathcal{L}$ is compact in $\mathcal{A}(\mathcal{L})$ and $\langle \mathcal{H}, h\rangle$ is a hypothesis space for $\mathcal{L}$, then there is a continuous function $f\colon \mathcal{T}(\mathcal{L}) \to \mathcal{H}^\omega$ such that for any $T \in \mathcal{T}$:*

1. *For all $i \in \omega$, $f(T)(i) \neq f(T)(i+1)$ implies $h(f(T)(i)) \subset h(f(T)(i+1))$,*
2. *There exists $i \in \omega$ such that $h(f(T)(j)) = \tau_{\mathcal{L}}(T)$ for all $j \geq i$.*

Proof: For each $L \in \mathcal{L}$ choose some $H_L \in \mathcal{H}$ such that $h(H_L) = L$. For each $T \in \mathcal{T}(\mathcal{L})$, let $\{X_i\}_{i\in\omega}$ be the ascending chain of closed sets in $\mathcal{A}(\mathcal{L})$ defined as $X_i = C_{\mathcal{L}}(\{n \in \omega \mid \exists j \le i : T(j) = n\})$. Since $\tau_{\mathcal{L}}(T) = \bigcup_{i\in\omega} X_i$, and $\tau_{\mathcal{L}}(T) \in \mathcal{L}$ is compact, it follows that there is $i \in \omega$ such that $X_j = \tau_{\mathcal{L}}(T)$ for all $i \ge j$. Define $i_0 = \min(\{i \in \omega \mid X_i \in \mathcal{L}\})$ and $i_{n+1} = \min(\{i \in \omega \mid i > i_n \text{ and } X_i \in \mathcal{L}\})$. Then $i_0, i_1, i_2, \ldots$ is an infinite sequence such that $X_{i_n} \in \mathcal{L}$ and $X_{i_n} \subseteq X_{i_{n+1}}$ for all $n \in \omega$, and there exists $m \in \omega$ such that $X_{i_n} = \tau_{\mathcal{L}}(T)$ for all $n \ge m$. Finally, for $n \in \omega$ define $f(T)(n) = H_L$, where $L = X_{i_n}$.

It is immediate from the construction of f that properties (1) and (2) hold, and f is continuous because every finite initial segment of the output of f depends only on a finite initial segment of the input. □

Clearly, f in the above proof identifies $\mathcal{L}$ in the limit, so we obtain the following.

Theorem 2.1.10 *If every $L \in \mathcal{L}$ is compact in $\mathcal{A}(\mathcal{L})$, then $\mathcal{L}$ is identifiable in the limit from positive data.* □

The notion of a compact element is closely related to the following well known structural property of concept spaces.

Definition 2.1.11 (Angluin [4], Kobayashi [32]) *Let L be a concept in $\mathcal{L}$. A* characteristic set *for L is a finite set F such that $F \subseteq L$ and for all $L' \in \mathcal{L}$, if $F \subseteq L'$ then $L \subseteq L'$.* □

It trivially follows that if F is a characteristic set for $L \in \mathcal{L}$, then F is also a finite tell-tale for L.

Theorem 2.1.12 *For any concept space $\mathcal{L}$ and concept $L \in \mathcal{L}$, the following are equivalent:*

1. *L is compact in $\mathcal{A}(\mathcal{L})$,*
2. *There exists a finite subset F of L such that $C_{\mathcal{L}}(F) = L$,*
3. *L has a characteristic set.*

Proof: The equivalence of compactness and being finitely generated has already been shown, so it suffices to show that (2) and (3) are equivalent.

If $F \subseteq L$ is finite and $C_{\mathcal{L}}(F) = L$, then for any $L' \in \mathcal{L}$ containing F, $L = C_{\mathcal{L}}(F) \subseteq C_{\mathcal{L}}(L') = L'$ follows from the monotonicity of $C_{\mathcal{L}}$. Therefore, F is a characteristic set for L.

For the converse, if F is a characteristic set of L, then every $L' \in \mathcal{L}$ containing F contains L, hence $L \subseteq \bigcap\{L' \in \mathcal{L} \mid F \subseteq L'\} = C_{\mathcal{L}}(F)$. Since $L = C_{\mathcal{L}}(L)$ and $F \subseteq L$, it follows that $L = C_{\mathcal{L}}(F)$. □

Corollary 2.1.13 (Angluin [4], Kobayashi [32]) *If every concept $L \in \mathcal{L}$ has a characteristic set, then $\mathcal{L}$ is identifiable in the limit from positive data.* □

2.1.4 Noetherian closure systems and finite unions

An algebraic closure system is *Noetherian* if and only if it does not contain an infinite strictly ascending chain of closed sets. This means that if $\mathcal{C}$ is a Noetherian algebraic closure system and $X_0 \subseteq X_1 \subseteq X_2 \subseteq \cdots$ is an infinite chain of closed sets in $\mathcal{C}$, then there must be some i such that $X_i = X_j$ for all $j \geq i$. The Noetherian propety is of fundamental importance in algebraic geometry [12], and in this section we show that it also plays an important role in learning theory. In particular, the Noetherian property plays a role in identifying "unions" of concept spaces.

Definition 2.1.14 (Wright [61], Motoki et al.[37]) *A concept space $\mathcal{L}$ has* infinite elasticity *if and only if there exists an infinite sequence of concepts $L_1, L_2, L_3, \ldots$ in $\mathcal{L}$ and elements $x_0, x_1, x_2, \ldots$ such that $\{x_0, \ldots, x_{n-1}\} \subseteq L_n$ but $x_n \notin L_n$. $\mathcal{L}$ has* finite elasticity *if and only if it does not have infinite elasticity.* □

Finite elasticity was introduced by Wright [61] because it is preserved under "unions" of concept spaces (we use quotations and the symbol $\tilde{\cup}$ to distinguish from the set theoretical definition of union):

Theorem 2.1.15 (Wright [61]) *If $\mathcal{K}$ and $\mathcal{L}$ have finite elasticity, then their "union" $\mathcal{K} \,\tilde{\cup}\, \mathcal{L} = \{K \cup L \mid K \in \mathcal{K} \text{ and } L \in \mathcal{L}\}$ has finite elasticity.* □

The following theorem shows that finite elasticity and the Noetherian property are closely related.

Theorem 2.1.16 *The following are equivalent for any concept space $\mathcal{L}$.*

1. *$\mathcal{L}$ has finite elasticity.*
2. *$\mathcal{A}(\mathcal{L})$ has finite elasticity.*
3. *$\mathcal{A}(\mathcal{L})$ is Noetherian.*
4. *Every element in $\mathcal{A}(\mathcal{L})$ is compact.*
5. *$\mathcal{A}(\mathcal{L})$ is identifiable in the limit from positive data.*

□

Proof: $(1 \Rightarrow 2)$: Assume an infinitely increasing chain of closed sets $X_0 \subset X_1 \subset \cdots$ exists in $\mathcal{A}(\mathcal{L})$. Let x_0 be any element of X_0, and for each $i \geq 0$ choose $x_i \in X_{i+1} \setminus X_i$. By the definition of $C_{\mathcal{L}}$ and the assumption that each X_i is closed, $x_i \notin \bigcap\{L \in \mathcal{L} \mid F_i \subseteq L\}$ for $F_i = \{x_0, \ldots, x_{i-1}\} \subseteq X_i$. However, there is a finite $G \subseteq X_{i+1}$ such that $x_i \in \bigcap\{L \in \mathcal{L} \mid G \subseteq L\}$, and we can choose G large enough that $F_i \cup \{x_i\} \subseteq G$. This implies that there must be some $L_i \in \mathcal{L}$ such that $F_i \subseteq L_i$ but $G \not\subseteq L_i$, hence in particular $x_i \notin L_i$. Thus, $\{x_0, \ldots, x_{i-1}\} \subseteq L_i$ and $x_i \notin L_i$, and since i was arbitrary, this shows that $\mathcal{L}$ has infinite elasticity, a contradiction.

$(2 \Rightarrow 3)$: Assume $X_1 \subset X_2 \subset X_3 \subset \cdots$ is an infinite strictly ascending chain of closed sets in $\mathcal{A}(\mathcal{L})$. Let x_0 be any element of X_1, and for each $i \geq 0$ choose $x_i \in X_{i+1} \setminus X_i$. Then the elements $x_0, x_1, x_2, \ldots$ and sets $X_1, X_2, X_3, \ldots$ demonstrate that $\mathcal{A}(\mathcal{L})$ has infinite elasticity, a contradiction.

$(3 \Rightarrow 1)$: Assume that the languages $L_1, L_2, L_3, \ldots$ and elements $x_0, x_1, x_2, \ldots$ show the infinite elasticity of $\mathcal{L}$. From the definition of infinite elasticity, $\{x_0, \ldots, x_{n-1}\} \subseteq C_{\mathcal{L}}(\{x_0, \ldots, x_{n-1}\}) \subseteq L_n$, but $x_n \notin L_n$ so $C_{\mathcal{L}}(\{x_0, \ldots, x_{n-1}\}) \subset C_{\mathcal{L}}(\{x_0, \ldots, x_n\})$. Since n was arbitrary, $C_{\mathcal{L}}(\{x_0, \ldots, x_i\})$ $(i \geq 0)$ is an infinite strictly ascending chain, contadicting that $\mathcal{A}(\mathcal{L})$ is Noetherian.

$(3 \Leftrightarrow 4)$: This equivalence holds for all algebraic closure systems and is well known (see [11]), but we will give the proof for completeness. First assume $\mathcal{A}(\mathcal{L})$ is Noetherian and let $X \in \mathcal{A}(\mathcal{L})$ be given. Let $F_0 \subseteq F_1 \subseteq \cdots$ be an increasing sequence of finite subsets of ω such that $X = \bigcup_{i \in \omega} F_i$. Then $C_{\mathcal{L}}(F_0) \subseteq C_{\mathcal{L}}(F_1) \subseteq \cdots$ is an increasing sequence of closed sets in $\mathcal{A}(\mathcal{L})$ such that $X = \bigcup_{i \in \omega} C_{\mathcal{L}}(F_i)$. Since $\mathcal{A}(\mathcal{L})$ is Noetherian, there must be $i \in \omega$ such that $C_{\mathcal{L}}(F_i) = C_{\mathcal{L}}(F_j)$ for all $j \geq i$, which implies that $X = C_{\mathcal{L}}(F_i)$, thus X is finitely generated, hence X is compact.

For the converse, if every element of $\mathcal{A}(\mathcal{L})$ is compact and $X_0 \subseteq X_1 \subseteq \cdots$ is an increasing sequence of closed sets, then $\{X_i\}_{i \in \omega}$ is directed so $X = \bigcup_{i \in \omega} X_i$ is an element of $\mathcal{A}(\mathcal{L})$ by Proposition 2.1.5. Since X is compact by assumption, there is $i \in \omega$ such that $X \subseteq X_i$, which implies that $X = X_j$ for all $j \geq i$. Therefore, $\mathcal{A}(\mathcal{L})$ is Noetherian.

$(4 \Rightarrow 5)$: This follows from Theorem 2.1.10 because $\mathcal{A}(\mathcal{A}(\mathcal{L})) = \mathcal{A}(\mathcal{L})$.

$(5 \Rightarrow 4)$: Assume $X \in \mathcal{A}(\mathcal{L})$ is not compact. For any finite $F \subseteq X$, $C_{\mathcal{L}}(F) \subset X$ because X is not finitely generated. Therefore, X does not have a finite tell-tale, so by Theorem 1.2.2 $\mathcal{A}(\mathcal{L})$ is not identifiable in the limit from positive data. □

As corollary, we obtain a characterization of the algebraic closure systems that are identifiable in the limit from positive data.

Corollary 2.1.17 *If $\mathcal{C}$ is an algebraic closure system, then $\mathcal{C}$ is identifiable in the limit from positive data if and only if $\mathcal{C}$ is Noetherian.*

Proof: Simply note that $\mathcal{A}(\mathcal{C}) = \mathcal{C}$ by Theorem 2.1.8. □

The above corollary generalizes a result by Stephan and Ventsov [57] showing that the class of all ideals of a countable ring is identifiable in the limit from positive data if and only if the ring is Noetherian. In addition, we obtain the following corollaries, which are already well known. In particular, we get a sufficient criterion for guaranteeing that the "union" of two concept spaces is identifiable from positive data.

Corollary 2.1.18 (Kobayashi [32]) *If $\mathcal{L}$ has finite elasticity then every $L \in \mathcal{L}$ has a characteristic set.* □

Corollary 2.1.19 (Wright [61]) *If $\mathcal{L}$ has finite elasticity then $\mathcal{L}$ is identifiable in the limit from positive data.* □

Corollary 2.1.20 (Wright [61]) *If $\mathcal{K}$ and $\mathcal{L}$ have finite elasticity, then $\mathcal{K} \tilde{\cup} \mathcal{L}$ is identifiable in the limit from positive data.* □

2.1.5 Well-partial-orders and unbounded unions

In this subsection we investigate well-partial-orderings, which is even stronger than the Noetherian property just discussed. We saw that the Noetherian property plays a role in identifying "unions" of concepts, and we will see in this subsection that well-partial-orders are useful in guaranteeing that *unbounded unions* of concepts are identifiable.

The notion of "unions" of concept spaces was expanded to unbounded unions by Shinohara and Arimura [54]. Given a concept space $\mathcal{L}$, define the space of unbounded unions $\mathcal{L}^{<\omega}$ to be the set of all finite unions of concepts of $\mathcal{L}$. Formally, $\mathcal{L}^{<\omega}$ is defined as:

$$\mathcal{L}^{<\omega} = \{\bigcup_{i \in I} L_i \mid L_i \in \mathcal{L}, I \text{ is a non-empty finite subset of } \omega\}.$$

Shinohara and Arimura gave a sufficient criterion on $\mathcal{L}$ to guarantee that $\mathcal{L}^{<\omega}$ is identifiable in the limit from positive data. We will reprove their result below, but we will use rather different methods.

Let $\langle P, \leq_P \rangle$ be a partial order. An *anti-chain* of P is a subset $A \subseteq P$ of elements that are mutually incomparable with respect to $\leq_P$. That is, for all $x, y \in A$ such that $x \neq y$, neither $x \leq_P y$ nor $y \leq_P x$ holds.

Lemma 2.1.21 *Let $\leq_P$ and $\sqsubseteq_P$ be two partial orders on P such that for $x, y \in P$, $x \leq_P y$ implies that $x \sqsubseteq_P y$. If there are no infinite anti-chains in P with respect to $\leq_P$, then there are no infinite anti-chains in P with respect to $\sqsubseteq_P$.*

Proof: Obviously, if x and y are incomparable with respect to $\sqsubseteq_P$ then they are incomparable with respect to $\leq_P$, so any infinite anti-chain with respect to $\sqsubseteq_P$ is an infinite anti-chain with respect to $\leq_P$. $\square$

A partial order $\langle P, \leq_P \rangle$ is a *well-partial-order* if and only if P contains no infinite strictly descending chains and no infinite anti-chains with respect to $\leq_P$. A finite or infinite sequence $x_0, x_1, \ldots$ of elements of P is a *bad sequence* if for all i and j such that $i < j$, $x_i \not\leq_P x_j$. Note that P is well-partially-ordered if and only if it does not contain any infinite bad sequences. We will define $Bad(\langle P, \leq_P \rangle)$ to be the set of all finite bad sequences of P.

Given a partial order $\langle P, \leq_P \rangle$, let P^* be the set of all finite ordered sequences of elements of P. We will write $s\langle x \rangle$ to represent the concatenation of an element $x \in P$ to the end of a sequence $s \in P^*$. The *Higman embedding*, $\preceq_H$, is a partial ordering on P^* defined such that for $s, t \in P^*$, $s \preceq_H t$ if and only if $s = \langle x_0, \ldots, x_n \rangle$ and $t = \langle y_0, \ldots, y_m \rangle$ and there exists $j_0 < \cdots < j_n \leq m$ such that $x_0 \leq_P y_{j_0}, \ldots, x_n \leq_P y_{j_n}$. The *subsequence relation*, $\preceq_S$, on P^* is defined similarly, with the stronger requirement that $s \preceq_S t$ if and only if there exists $j_0 < \cdots < j_n \leq m$ such that $x_0 = y_{j_0}, \ldots, x_n = y_{j_n}$. Note that $\preceq_S$ is equivalent to $\preceq_H$ if P is ordered by equality (i.e., $x \leq_P y \iff x = y$ for all $x, y \in P$).

Lemma 2.1.22 (Higman [23]. See also [19]) *For any well-partial-ordering $\leq_P$ on P, the Higman embedding $\preceq_H$ is a well-partial-ordering on P^*.* $\square$

We next show that by placing some assumptions on a concept space $\mathcal{L}$, we can show that if $\langle \mathcal{L}, \supseteq \rangle$ (i.e., $\mathcal{L}$ ordered by reverse subset inclusion) is a well-partial-order, then $\langle \mathcal{A}(\mathcal{L}), \supseteq \rangle$ is also a well-partial order. This will have applications when we discuss the identifiability of *unbounded unions* of concepts later.

Definition 2.1.23 (Angluin [3]) *A concept space $\mathcal{L}$ has* finite thickness *if and only if $\emptyset \notin \mathcal{L}$ and each $x \in \omega$ is contained in at most a finite number of concepts in $\mathcal{L}$.* □

Finite thickness was introduced by Angluin, and is well known to be a property held by pattern languages [3]. It is easy to see that if $\mathcal{L}$ has finite thickness then $\mathcal{L}$ has finite elasticity, but the converse does not hold in general.

Lemma 2.1.24 *If $\mathcal{L}$ is a concept space with finite elasticity, then $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order if and only if $\mathcal{L}$ does not contain any infinite anti-chains.*

Proof: If $\mathcal{L}$ contains an infinite anti-chain, then it is not a well-partial-order by definition. For the converse, if we assume that $\mathcal{L}$ does not contain any infinite anti-chains, then it only remains to show that $\langle \mathcal{L}, \supseteq \rangle$ does not have any infinite strictly decreasing chains. Now, a decreasing chain in $\langle \mathcal{L}, \supseteq \rangle$ is just a chain of concepts increasing with respect to subset inclusion. Since $\mathcal{L} \subseteq \mathcal{A}(\mathcal{L})$ and $\mathcal{A}(\mathcal{L})$ is Noetherian by the assumption that $\mathcal{L}$ has finite elasticity, $\mathcal{L}$ does not have any infinite strictly ascending chains with respect to $\subseteq$. Therefore, $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order. □

Theorem 2.1.25 *If $\mathcal{L}$ is a concept space with finite thickness, then $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order if and only if $\langle \mathcal{A}(\mathcal{L}), \supseteq \rangle$ is a well-partial-order.*

Proof: It is immediate that if $\langle \mathcal{L}, \supseteq \rangle$ is *not* a well-partial-order, then neither is $\langle \mathcal{A}(\mathcal{L}), \supseteq \rangle$. For the converse, from the previous lemma it suffices to show that $\mathcal{A}(\mathcal{L})$ does not contain any infinite anti-chains. Since $C_{\mathcal{L}}(\emptyset)$ is a subset of every other closed set in $\mathcal{A}(\mathcal{L})$, it cannot be in any infinite anti-chain. Therefore, we only need to show that $\mathcal{A}(\mathcal{L}) \setminus \{C_{\mathcal{L}}(\emptyset)\}$ contains no infinite anti-chains.

Since $\mathcal{L}$ has finite thickness, it has finite elasticity, so for every $X_i \in \mathcal{A}(\mathcal{L}) \setminus \{C_{\mathcal{L}}(\emptyset)\}$ there is a non-empty finite set $F_i \subseteq \omega$ such that $X_i = C_{\mathcal{L}}(F_i)$. From the definition of finite thickness, any non-empty set of elements is contained in at most a finite number of concepts in $\mathcal{L}$. Therefore, every $X_i \in \mathcal{A}(\mathcal{L}) \setminus \{C_{\mathcal{L}}(\emptyset)\}$ can be written as $X_i = L^i_0 \cap \cdots \cap L^i_{n_i}$ $(n_i \geq 0)$, where L^i_j $(0 \leq j \leq n_i)$ is the (possibly empty) finite sequence of all concepts containing F_i.

Define $f : (\mathcal{A}(\mathcal{L}) \setminus \{C_{\mathcal{L}}(\emptyset)\}) \to \mathcal{L}^*$ so that $f(X_i) = \langle L^i_0, \ldots, L^i_{n_i} \rangle$. We order $\mathcal{L}^*$ by the Higman embedding $\preceq_H$ based on the ordering $\supseteq$ on $\mathcal{L}$. Since $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order by assumption, Lemma 2.1.22 guarantees that $\langle \mathcal{L}^*, \preceq_H \rangle$ is a well-partial-order, and therefore contains no infinite anti-chains.

We now show that $f(X_i) \preceq_H f(X_j)$ implies that $X_i \supseteq X_j$. Assume that $f(X_i) \preceq_H f(X_j)$, then there exists $k_0 < \cdots < k_{n_i} \leq n_j$ such that $L^i_0 \supseteq L^j_{k_0}, \ldots, L^i_{n_i} \supseteq L^j_{k_{n_i}}$. This implies that $L^i_0 \cap \cdots \cap L^i_{n_i} \supseteq L^j_{k_0} \cap \cdots \cap L^j_{k_{n_i}}$. Since $L^j_{k_0} \cap \cdots \cap L^j_{k_{n_i}} \supseteq L^j_0 \cap \cdots \cap L^j_{n_j}$, it follows that $X_i \supseteq X_j$. From Lemma 2.1.21, it follows that there are no infinite anti-chains in $\langle \mathcal{A}(\mathcal{L}), \supseteq \rangle$. □

We can now give an alternative proof of the result by Shinohara and Arimura [54] showing that if $\mathcal{L}$ has finite thickness and no infinite anti-chains, then $\mathcal{L}^{<\omega}$ is identifiable in the limit from positive data.

Theorem 2.1.26 *If $\mathcal{L}$ is a concept space with finite thickness and $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order, then $\mathcal{L}^{<\omega}$ has finite-elasticity. In particular, $\mathcal{L}^{<\omega}$ is identifiable in the limit from positive data.*

Proof: Assume for a contradiction that there are concepts $U_1, U_2, U_3, \ldots$ in $\mathcal{L}^{<\omega}$ and elements $x_0, x_1, x_2, \ldots$ such that $\{x_0, \ldots, x_{n-1}\} \subseteq U_n$ but $x_n \notin U_n$. If $m < n$ and $x_n \in C_{\mathcal{L}}(\{x_m\})$, then every concept in $\mathcal{L}$ containing x_m contains x_n, but since $x_m \in U_n$ there must be $L \in \mathcal{L}$ such that $x_m \in L \subseteq U_n$, hence $x_n \in L \subseteq U_n$, a contradiction. Thus, for every $m < n$, $C_{\mathcal{L}}(\{x_m\}) \not\supseteq C_{\mathcal{L}}(\{x_n\})$, contradicting the fact that $\langle \mathcal{A}(\mathcal{L}), \supseteq \rangle$ contains no infinite bad sequences. □

It was left open by Shinohara and Arimura [54] as to whether or not a similar statement could be made for finite elasticity. However, the next proposition shows that "finite thickness" cannot be weakened to "finite elasticity" in Theorem 2.1.25 and Theorem 2.1.26.

Proposition 2.1.27 *There exists a concept space $\mathcal{L}$ with finite elasticity such that*

1. *$\mathcal{L}$ contains no infinite anti-chains with respect to set inclusion,*

2. *$\mathcal{A}(\mathcal{L})$ contains an infinite anti-chain,*

3. *$\mathcal{L}^{<\omega}$ is not inferable from positive data.*

Proof: Define

$$\mathcal{SINGLE} = \{\{i\} \mid i \in \omega\} \quad \text{and} \quad \mathcal{COINIT} = \{\{k \in \omega \mid k \geq j\} \mid j \in \omega\}.$$

Consider the concept space $\mathcal{L} = \mathcal{SINGLE} \,\tilde{\cup}\, \mathcal{COINIT}$. $\mathcal{L}$ has finite elasticity because both $\mathcal{SINGLE}$ and $\mathcal{COINIT}$ have finite elasticity. It will be convenient to use ordered pairs of natural numbers to represent $\mathcal{L}$ as the set $\{\langle i, j \rangle_{\mathcal{L}} \mid i < j\}$, where $\langle i, j \rangle_{\mathcal{L}}$ is defined to be $\{i\} \cup \{k \in \omega \mid k \geq j\}$ (note that if $i \geq j$, then $\langle i, j \rangle_{\mathcal{L}} = \langle j, j+1 \rangle_{\mathcal{L}}$, which is why we can restrict our attention to only the cases where $i < j$). Then it is easily shown that $\langle i, j \rangle_{\mathcal{L}} \supseteq \langle i', j' \rangle_{\mathcal{L}} \iff i' \geq j$ or ($i' = i$ and $j' \geq j$).

(1). Assume that $\mathcal{L}$ contains an infinite anti-chain $\{\langle i_k, j_k \rangle_{\mathcal{L}}\}_{k \geq 0}$. Then since $\langle i_0, j_0 \rangle_{\mathcal{L}} \not\supseteq \langle i_k, j_k \rangle_{\mathcal{L}}$ for all $k > 0$, it follows that $i_k < j_0$ for all $k > 0$. Since there are only finitely many natural numbers less than j_0, there must be some $i < j_0$ such that $i_k = i$ for infinitely many $k > 0$. Thus, there is an infinite subsequence $\{\langle i, j_{f(k)} \rangle_{\mathcal{L}}\}_{k \geq 0}$ of $\{\langle i_k, j_k \rangle_{\mathcal{L}}\}_{k \geq 0}$ which is also an infinite anti-chain (here we are using the strictly monotonic function $f : \omega \to \omega$ to specify the indices of the subsequence). For any $k \geq 0$, since $\langle i, j_{f(k)} \rangle_{\mathcal{L}} \not\supseteq \langle i, j_{f(k+1)} \rangle_{\mathcal{L}}$, it must hold that $j_{f(k+1)} < j_{f(k)}$, which contradicts the well-orderedness of the natural numbers. Therefore, $\mathcal{L}$ does not contain any infinite anti-chains.

(2). For any $i \in \omega$, note that $C_{\mathcal{L}}(\{i\}) = \{i\}$, because $\{i\} = \bigcap_{j > i} \langle i, j \rangle_{\mathcal{L}}$. Therefore, $\{\{i\} \mid i \in \omega\}$ is an infinite anti-chain in $\mathcal{A}(\mathcal{L})$.

(3). Note that $\omega = \langle 0, 1 \rangle_{\mathcal{L}}$ is in $\mathcal{L}^{<\omega}$. Let $F = \{i_1, i_2, \ldots, i_n\}$ be any finite subset of ω, and let $j = \max(F) + 2$, where $\max(F)$ denotes the largest natural number in F. Then $L = \langle i_1, j \rangle_{\mathcal{L}} \cup \langle i_2, j \rangle_{\mathcal{L}} \cup \cdots \cup \langle i_n, j \rangle_{\mathcal{L}}$ is in $\mathcal{L}^{<\omega}$ and contains F, and since it does not contain $\max(F) + 1$ it is a proper subset of ω. Thus, ω has no finite tell-tale and so $\mathcal{L}^{<\omega}$ is not inferable from positive data. □

2.2 Order-types and mind-change complexity

In this section, we analyze the mind-change complexity of concept spaces. Mind-change complexity was proposed by Freivalds and Smith [18] as a method of characterizing the complexity of identification in the limit using ordinals. After reviewing the definition of ordinals and mind-change complexity, we will give some characterizations of the mind-change complexity of some concept spaces based on the order-type of a partial order.

2.2.1 Ordinals

For the purpose of defining ordinals, we will assume that we are working within Zermelo-Fraenkel set theory with the axiom of choice. We will use [33] as a reference on ordinals.

A partial order $\langle P, \leq \rangle$ is *well-ordered* if and only if it is *totally ordered* (for every $x, y \in P$, either $x \leq y$ or $y \leq x$) and *well-founded* (every non-empty subset of P has a least element with respect to $\leq$).

A *strict well-order* is a pair $\langle P, < \rangle$ consisting of a set P and a relation $<$ on P, such that $<$ is *irreflexive* ($\neg x < x$), *transitive* ($x < y$ and $y < z$ implies $x < z$), *total* (for every $x, y \in P$, either $x = y$ or $x < y$ or $y < x$), and *well-founded* (for every non-empty subset S of P, there is $x \in P$ such that $x < y$ for all $y \in S$ such that $y \neq x$).

Note that a strict well-ordering $\langle P, < \rangle$ can be turned into a well-order $\langle P, \leq \rangle$ by defining $x \leq y$ if and only if $x = y$ or $x < y$. Similarly, a well-order $\langle P, \leq \rangle$ is turned into a strict well-order $\langle P, < \rangle$ by defining $x < y$ if and only if $x \leq y$ and $x \neq y$.

A bijective function $f\colon P \to Q$ between partially-ordered sets $\langle P, \leq_P \rangle$ and $\langle Q, \leq_Q \rangle$ is an *order-isomorphism* if and only if $x \leq_P y \iff f(x) \leq_Q f(y)$ for all $x, y \in P$. In such a case we say that $\langle P, \leq_P \rangle$ and $\langle Q, \leq_Q \rangle$ are *order-isomorphic* and write $\langle P, \leq_P \rangle \cong \langle Q, \leq_Q \rangle$. A similar definition can be given to strict well-orders.

Definition 2.2.1 *A set X is an* ordinal *if and only if X is strictly well-ordered by $\in$ and every element of X is a subset of X.* □

As an example, $0 = \emptyset, 1 = \{0\}, 2 = \{0, 1\}, \ldots$ and $\omega = \{0, 1, 2, \ldots\}$, are all ordinals. On the other hand, $\{\{\emptyset\}\}$ is not an ordinal because $\emptyset \notin \{\{\emptyset\}\}$ thus $\{\emptyset\} \not\subseteq \{\{\emptyset\}\}$. If X is an ordinal and is non-empty, then there is a $\in$-least element $Y \in X$ which must also then satisfy $Y \subseteq X$. Thus, if there is $Z \in Y$, then $Z \in X$, contradicting the claim that Y is the $\in$-least element of X. Therefore, $Y = \emptyset$.

Proposition 2.2.2 (see Kunen [33])

1. *If X is an ordinal and $Y \in X$, then Y is an ordinal and* $Y = \{Z \in X \mid Z \in Y\}$,
2. *If X and Y are ordinals and $X \cong Y$, then $X = Y$,*
3. *If X and Y are ordinals then exactly one of the following hold: $X = Y$, $X \in Y$, or $Y \in X$,*

4. If X, Y, and Z are ordinals, $X \in Y$ and $Y \in Z$, then $X \in Z$. □

We will usually use α, β, γ and λ as variables ranging over ordinals, and write $\alpha \leq \beta$ to denote $\alpha = \beta$ or $\alpha \in \beta$. From the above proposition, it follows that for any ordinal α, $\langle \alpha, \leq \rangle$ is the well-ordered set of all ordinals less than α.

Proposition 2.2.3 (see Kunen [33]) *If $\langle P, \leq_P \rangle$ is a well-ordered set, then there exists a unique ordinal α such that $\langle P, \leq_P \rangle \cong \langle \alpha, \leq \rangle$.* □

The ordinal α from the above proposition will be denoted $Ord(\langle P, \leq_P \rangle)$, and is called the *order-type* of the well-order $\langle P, \leq_P \rangle$.

If S is a set of ordinals, then $\bigcup S$ is an ordinal and is the supremum of all ordinals in S. We will usually write $\bigvee S$ instead of $\bigcup S$ to emphasize the order-theoretic aspsect of the ordinal. If S is a non-empty set of ordinals, then $\bigcap S$ is an ordinal and is the infimum of all ordinals in S (denoted $\bigwedge S$).

If α is an ordinal, then $\alpha \cup \{\alpha\}$ is also an ordinal, called the *successor* of α, and is denoted $\alpha + 1$. An ordinal α is a *successor ordinal* if and only if there is an ordinal β such that $\alpha = \beta + 1$. An ordinal that is not the empty set and not a successor ordinal is called a *limit ordinal*.

A partial order $\langle P, \leq_P \rangle$ is *well-founded* if and only if $\langle P, \leq_P \rangle$ contains no infinitely descending chains. We can assign an ordinal to every well-founded order $\langle P, \leq_P \rangle$, which we will also call the *order type* of $\langle P, \leq_P \rangle$ and denote $Ord(\langle P, \leq_P \rangle)$. For $x \in P$, define

$$ord_P(x) = \bigvee \{ord_P(y) + 1 \mid y \in P \text{ and } y < x\}.$$

In particular, if there is no $y \in P$ less than x, then $ord_P(x) = 0$. Now we define the order type of $\langle P, \leq_P \rangle$ to be

$$Ord(\langle P, \leq_P \rangle) = \bigvee \{ord_P(x) + 1 \mid x \in P\}.$$

It is easy to see that this definition is consistent with our earlier one in the case that $\langle P, \leq_P \rangle$ is a well-order.

Given ordinals α and β, we use transfinite recursion to define the following operations on ordinals (see [33], [29]).

1. The *sum* $\alpha + \beta$:
   (a) $\alpha + 0 = \alpha$,
   (b) $\alpha + 1 =$ the successor of α,
   (c) $\alpha + (\beta + 1) = (\alpha + \beta) + 1$,
   (d) $\alpha + \lambda = \bigvee_{\beta < \lambda} \alpha + \beta$ for limit ordinal λ.
2. The *product* $\alpha \cdot \beta$ (or $\alpha\beta$):
   (a) $\alpha \cdot 0 = 0$,
   (b) $\alpha \cdot (\beta + 1) = \alpha \cdot \beta + \alpha$,
   (c) $\alpha \cdot \lambda = \bigvee_{\beta < \lambda} (\alpha \cdot \beta)$ for limit ordinal λ.
3. The *exponential* α^β:

(a) $\alpha^0 = 1$,

(b) $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$,

(c) $\alpha^\lambda = \bigvee_{\beta<\lambda} \alpha^\beta$ for limit ordinal λ..

Associativity holds in general: $\alpha+(\beta+\gamma) = (\alpha+\beta)+\gamma$ and $\alpha\cdot(\beta\cdot\gamma) = (\alpha\cdot\beta)\cdot\gamma$, and multiplication is left-distributive: $\alpha\cdot(\beta+\gamma) = \alpha\cdot\beta+\alpha\cdot\gamma$. However, addition and multiplication are *not* commutative. For example, $1 + \omega = \omega \neq \omega + 1$ and $2 \cdot \omega = \omega \neq \omega \cdot 2$. Furthermore, $(1 + 1) \cdot \omega = \omega \neq 1 \cdot \omega + 1 \cdot \omega$ shows that distributivity for multiplication on the right fails. However, the definitions agree with the usual definition of addition and multiplication when $\alpha, \beta < \omega$.

Example 2.2.4 (see Kunen [33]) Let α and β be ordinals. Define $P = \alpha \times \{0\} \cup \beta \times \{1\}$ and for $\langle \xi, i\rangle, \langle \eta, j\rangle \in P$, define $\langle \xi, i\rangle \leq_P \langle \eta, j\rangle$ if and only if $i < j$ or ($i = j$ and $\xi \leq \eta$). Then $\langle P, \leq_P\rangle$ is a well-order and $Ord(\langle P, \leq_P\rangle) = \alpha + \beta$. □

Example 2.2.5 (see Kunen [33]) Let α and β be ordinals. Define $P = \beta\times\alpha$ and for $\langle \xi, \eta\rangle, \langle \xi', \eta'\rangle \in P$, define $\langle \xi, \eta\rangle \leq_P \langle \xi', \eta'\rangle$ if and only if $\xi < \xi'$ or ($\xi = \xi'$ and $\eta \leq \eta'$). Then $\langle P, \leq_P\rangle$ is a well-order and $Ord(\langle P, \leq_P\rangle) = \alpha \cdot \beta$. □

We have already seen that each natural number and ω are all ordinals. Clearly, ω is the least infinite ordinal (infinite in the sense that ω is a set with infinite cardinality). Another important ordinal is ε_0 which is defined as the least ordinal α such that $\alpha = \omega^\alpha$. From the definition of exponentiation, we can see that this implies that ε_0 is the supremum of the sequence $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$.

Finally, we let ω_1 denote the least uncountable ordinal, which implies that ω_1 is the set of all countable ordinals. Since the countable union of countable sets is countable, the supremum of any countable sequence of countable ordinals is a countable ordinal. Thus, in particular, ε_0 is countable. In this thesis, we will usually only work with ordinals that are countable.

2.2.2 Mind-change complexity

Mind-change complexity was proposed by Freivalds and Smith [18] as a method of characterizing the complexity of identification in the limit using ordinals. Basically, mind-change complexity measures the number of times that a learner must change hypotheses before converging to a correct hypothesis. Intuitively, a learner starts with an ordinal α, and every time the learner changes its hypothesis it must output a smaller ordinal. Since every ordinal is well-founded, this limits the number of times the learner can change hypotheses.

Consider the problem of identifying $\mathcal{COINIT} = \{\{k \in \omega \mid k \geq j\} \mid j \in \omega\}$ from positive data. If the learner sees k in the text of some concept in $\mathcal{COINIT}$, then since there are only $k+1$ concepts in $\mathcal{COINIT}$ that contain k, the learner can put a bound on the number of times that it will change its hypothesis in the future. In particular, the learner will never need to make more than ω mind-changes to identify any concept in $\mathcal{COINIT}$. However, it is clear that for any $n < \omega$, we can force the learner to change its hypothesis more than n times. Therefore, it takes at least ω mind-changes to identify every concept in $\mathcal{COINIT}$. We therefore say that $\mathcal{COINIT}$ has mind-change complexity ω.

When identifying a concept in $\mathcal{SINGLE} = \{\{i\} \mid i \in \omega\}$, the learner can simply wait until anything other than # appears in the text, and then the learner immediately knows which concept is being presented. Therefore, the learner only needs to output a hypothesis once to succesfully identify any concept in $\mathcal{SINGLE}$, so we say that $\mathcal{SINGLE}$ has mind-change complexity of one.

We should note that some authors do not count the learner's first outputted hypothesis as a mind-change, so they would say that $\mathcal{SINGLE}$ has mind-change complexity of zero. We will count the first hypothesis as a mind-change, because it results in a more mathematically natural theory. The reader can think of the learner's first hypothesis as a mind-change from the empty hypothesis "do not know at this time" to a more concrete hypothesis.

Given sets X and Y we define $X \times Y = \{\langle x, y\rangle \mid x \in X \text{ and } y \in Y\}$. If X and Y are subsets of ω, then we can also assume that $X \times Y$ is appropriately encoded as a subset of ω. We define the *projections* $\pi_1 \colon X \times Y \to X$ and $\pi_2 \colon X \times Y \to Y$ as $\pi_1(\langle x, y\rangle) = x$ and $\pi_2(\langle x, y\rangle) = y$. We will also define projections of sequences $\pi_1^\omega \colon (X \times Y)^\omega \to X^\omega$ and $\pi_2^\omega \colon (X \times Y)^\omega \to Y^\omega$ as $\pi_1^\omega(\xi)(i) = \pi_1(\xi(i))$ and $\pi_2^\omega(\xi)(i) = \pi_2(\xi(i))$ for each $\xi \in (X \times Y)^\omega$. Intuitively, given a sequence ξ of pairs of elements of X and Y, $\pi_1^\omega(\xi)$ is the infinite sequence of elements of X obtained by just ignoring the second element of each pair (and similarly for $\pi_2^\omega(\xi)$).

We now give a formal definition of mind-change complexity. Although the original definition is due to Freivalds and Smith [18], we present a modified version that is suited for our interpretation of learners as continuous functions.

Definition 2.2.6 *Let $\mathcal{L}$ be a concept space and $\langle \mathcal{H}, h\rangle$ a hypothesis space for $\mathcal{L}$, and let α be a countable ordinal. We define a new hypothesis space $\langle \mathcal{H}_\alpha, h_\alpha\rangle$ for $\mathcal{L}$ by*

$$\begin{aligned} \mathcal{H}_\alpha &= \mathcal{H} \times \alpha = \{\langle H, \beta\rangle \mid H \in \mathcal{H} \textit{ and } \beta < \alpha\}, \\ h_\alpha(\langle H, \beta\rangle) &= h(H). \end{aligned}$$

Given a representation $\langle \mathcal{R}, \rho\rangle$ of $\mathcal{L}$ and a learner (a continuous function) $\psi \colon \mathcal{R} \to \mathcal{H}_\alpha^\omega$, we say that ψ identifies $\mathcal{L}$ in the limit with α mind-changes *(with respect to $\langle \mathcal{R}, \rho\rangle$) if and only if for every $R \in \mathcal{R}$:*

1. *$\pi_1^\omega(\psi(R))$ converges to a hypothesis for $\rho(R)$,*
2. *$\pi_2^\omega(\psi(R))$ is non-increasing,*
3. *$i < j$ and $\pi_1(\psi(R)(i)) \neq \pi_1(\psi(R)(j))$ implies $\pi_2(\psi(R)(i)) > \pi_2(\psi(R)(j))$.*

$\mathcal{L}$ is identifiable in the limit with α mind-changes *(with respect to $\langle \mathcal{R}, \rho\rangle$) if and only if there is a learner that identifies $\mathcal{L}$ in the limit with α mind-changes (with respect to $\langle \mathcal{R}, \rho\rangle$). We say that $\mathcal{L}$ has* mind-change complexity α *(with respect to $\langle \mathcal{R}, \rho\rangle$) if and only if $\mathcal{L}$ is identifiable in the limit with α mind-changes (with respect to $\langle \mathcal{R}, \rho\rangle$) and there is no $\beta < \alpha$ such that $\mathcal{L}$ is identifiable in the limit with β mind-changes (with respect to $\langle \mathcal{R}, \rho\rangle$).* □

By the definition of $\mathcal{H}_\alpha$, it is clear that the value of ψ's mind-change counter is always greater than 0 and *strictly* less than α. Assume that $\psi \colon \mathcal{R} \to \mathcal{H}_\alpha^\omega$ is a learner and R is a representation of some concept. For $i \in \omega$, if $\psi(R)(i) = \langle H_i, \beta_i\rangle$ then we can refer to $H_i = \pi_1(\psi(R)(i))$ as ψ's ith hypothesis, and $\beta_i = \pi_2(\psi(R)(i))$ as the value of ψ's *mind-change counter* at time i. Then the first criterion on ψ in the above definition means that ψ's sequence of hypotheses

$H_0, H_1, \ldots$ converges to a correct hypothesis for $\rho(R)$. The second criterion simply states that ψ's mind-change counter never increases (i.e., $\beta_0 \geq \beta_1 \geq \cdots$). The third criterion states that if ψ changes its hypothesis between time i and time j, then ψ's mind-change counter *strictly* decreases.

Note that the function $\pi_1^\omega : \mathcal{H}_\alpha^\omega \to \mathcal{H}^\omega$ is continuous, so if $\psi : \mathcal{R} \to \mathcal{H}_\alpha^\omega$ identifies $\mathcal{L}$ with α mind-changes, then $\pi_1^\omega \circ \psi : \mathcal{R} \to \mathcal{H}^\omega$ identifies $\mathcal{L}$ in the limit according to our original definition (without mind-changes). It is also clear that if $\alpha < \beta$ then $\mathcal{H}_\alpha^\omega \subseteq \mathcal{H}_\beta^\omega$, so if $\mathcal{L}$ is identifiable with α mind-changes it is identifiable with β mind-changes.

2.2.3 Characterizations

In this subsection we will give characterizations of the mind-change complexity of identifying concept spaces with finite elasticity from positive data. In the next chapter, we will give a more general topological characterization due to Luo and Schulte [34], but the characterization we give here is often easier to use in practice when finite-elasticity can be assumed.

Theorem 2.2.7 *Assume every concept in $\mathcal{L}$ is compact in $\mathcal{A}(\mathcal{L})$ and $\langle \mathcal{L}, \supseteq \rangle$ is well-founded. Then the mind-change complexity of identifying $\mathcal{L}$ in the limit from positive data is $Ord(\langle \mathcal{L}, \supseteq \rangle)$.*

Proof: To simplify notation, let $\alpha = Ord(\langle \mathcal{L}, \supseteq \rangle)$. Also, note that the definition of ord_P when applied to the poset $\langle \mathcal{L}, \supseteq \rangle$ yields the definition $ord_{\mathcal{L}}(L) = \bigvee \{ ord_{\mathcal{L}}(L') + 1 \mid L' \in \mathcal{L} \text{ and } L' \supset L \}$ for $L \in \mathcal{L}$.

We will first show that if $\beta < \alpha$ then $\mathcal{L}$ is not identifiable with β mind-changes. Assume for a contradiction that $\psi : \mathcal{T}(\mathcal{L}) \to \mathcal{H}_\beta^\omega$ identifies $\mathcal{L}$. We show by transfinite induction that if T is a text for $L \in \mathcal{L}$ and $ord_{\mathcal{L}}(L) = \gamma$, then the sequence $\pi_2^\omega(\psi(T))$ contains no ordinals strictly less than γ. This would contradict the claim that ψ identifies $\mathcal{L}$, because by definition of $Ord(\langle \mathcal{L}, \supseteq \rangle)$ there must be a concept $L \in \mathcal{L}$ such that $ord_{\mathcal{L}}(L) \geq \beta$, but ψ's mind-change counter must always be strictly less than β.

First, it is trivial if $ord_{\mathcal{L}}(L) = 0$ then ψ's mind-change counter never falls below 0. So assume for all $\gamma' < \gamma$ that $\pi_2^\omega(\psi(T'))$ never falls below γ' for all texts T' for all concepts L' with $ord_{\mathcal{L}}(L') = \gamma'$, and we will show the same holds for γ. Assume for a contradiction that $ord_{\mathcal{L}}(L) = \gamma$ and T is a text for L such that $\pi_2(\psi(T)(i)) = \gamma' < \gamma$ for some $i \in \omega$. Since $\pi_1^\omega(\psi(T))$ must converge to a hypothesis for L, we can assume without loss of generality that i is chosen large enough that $H = \pi_1(\psi(T)(i))$ is a hypothesis for L. Since ψ is continuous, there must be some finite sequence $\sigma \prec T$ such that $\psi(T)[i+1] \prec \psi(T')$ whenever $\sigma \prec T' \in \mathcal{T}(\mathcal{L})$. Now, since $ord_{\mathcal{L}}(L) = \gamma$ and $\gamma' < \gamma$, there must be $L' \in \mathcal{L}$ such that $L' \supset L$ and $ord_{\mathcal{L}}(L') \geq \gamma'$. But then every element in σ is an element of L', so there must be a text T' for L' extending σ. This implies that $\psi(T')(i) = \langle H, \gamma' \rangle$, where H is a hypothesis for L. Since $\pi_1^\omega(\psi(T'))$ converges to a hypothesis for L', there must be $j > i$ such that $\psi(T') = \langle H', \gamma'' \rangle$ where H' is some hypothesis for L', thus $H \neq H'$ hence $\gamma'' < \gamma'$. But then $\pi_2^\omega(\psi(T'))$ has fallen below $ord_{\mathcal{L}}(L')$, contradicting the induction hypothesis. This completes the proof that $\mathcal{L}$ is not identifiable with $\beta < \alpha$ mind-changes.

Next we prove that $\mathcal{L}$ is identifiable with α mind-changes. Let $f : \mathcal{T}(\mathcal{L}) \to \mathcal{H}^\omega$ be a continuous function as in Lemma 2.1.9. Define $\psi : \mathcal{T}(\mathcal{L}) \to \mathcal{H}_\alpha^\omega$ so that

$\psi(T)(i) = \langle f(T)(i), ord_{\mathcal{L}}(h(f(T)(i)))\rangle$ for all $i \in \omega$. Since $ord_{\mathcal{L}}(L) < \alpha$ for all $L \in \mathcal{L}$, ψ is well defined and easily seen to be continuous. Furthermore, by our choice of f, $\pi_1^\omega(\psi(T))$ converges to a hypothesis for $\tau_{\mathcal{L}}(T)$, and $f(T)(i) \neq f(T)(i+1)$ implies $h(f(T)(i)) \subset h(f(T)(i+1))$ hence $ord_{\mathcal{L}}(h(f(T)(i))) > ord_{\mathcal{L}}(h(f(T)(i+1)))$ for all $i \in \omega$. Therefore, ψ identifies $\mathcal{L}$ in the limit with α mind-changes. □

Corollary 2.2.8 *Assume $\mathcal{L}$ has finite elasticity. Then the mind-change complexity of identifying $\mathcal{L}$ in the limit from positive data is $Ord(\langle\mathcal{L}, \supseteq\rangle)$.*

Proof: Since $\mathcal{L}$ has finite elasticity, Theorem 2.1.16 guarantees that every concept in $\mathcal{L}$ is compact and that $\langle\mathcal{L}, \supseteq\rangle$ is well-founded. □

Corollary 2.2.9 *Assume $\mathcal{L}$ has finite thickness. Then the mind-change complexity of identifying $\mathcal{L}$ in the limit from positive data is at most ω.*

Proof: Since $\mathcal{L}$ has finite thickness, for any $L \in \mathcal{L}$ there are at most finitely many $L' \in \mathcal{L}$ containing L. Therefore, it can be shown by induction that $ord_{\mathcal{L}}(L) < \omega$ for each $L \in \mathcal{L}$, hence $Ord(\langle\mathcal{L}, \supseteq\rangle) \leq \omega$. □

The following example shows that requiring each $L \in \mathcal{L}$ to have a characteristic set is required for Theorem 2.2.7 to hold.

Example 2.2.10 Let $\mathcal{L} = \{\omega\} \cup \{\omega \setminus \{n\} \mid n \in \omega\}$. Since $\omega \setminus \{m\}$ and $\omega \setminus \{n\}$ are incomparable whenever $n \neq m$, it is clear that $\langle\mathcal{L}, \supseteq\rangle)$ is well-founded. In fact, $ord_{\mathcal{L}}(\omega \setminus \{n\}) = 0$ and $ord_{\mathcal{L}}(\omega) = 1$, thus $Ord(\langle\mathcal{L}, \supseteq\rangle) = 2$. However, if $F \subseteq \omega$ is finite, then there is some $n \in \omega \setminus F$, hence $F \subseteq \omega \setminus \{n\}$. Therefore, ω has no finite tell-tale, so $\mathcal{L}$ is not identifiable in the limit from positive data. □

Theorem 2.1.26 showed that if $\mathcal{L}$ has finite thickness and $\langle\mathcal{L}, \supseteq\rangle$ is a well-partial-order, then $\mathcal{L}^{<\omega}$ has finite elasticity. We can therefore use Corollary 2.2.8 to determine the mind-change complexity of $\mathcal{L}^{<\omega}$. However, in practice it is often difficult to compute $Ord(\langle\mathcal{L}^{<\omega}, \supseteq\rangle)$ exactly. We will conclude this section by giving upper and lower bounds on $Ord(\langle\mathcal{L}^{<\omega}, \supseteq\rangle)$ which are sometimes easier to compute. But first we must introduce some more basic facts about well-partial-orders.

Lemma 2.2.11 *If $\langle P, \leq_P\rangle$ and $\langle Q, \leq_Q\rangle$ are well-founded and $f\colon\subseteq P \to Q$ is a surjective partial function such that for all $p \in dom(f)$ and $q \in Q$:*

$$q \leq_Q f(p) \Longrightarrow (\exists p' \in dom(f))[p' \leq_P p \;\&\; f(p') = q\,],$$

then $Ord(\langle P, \leq_P\rangle) \geq Ord(\langle Q, \leq_Q\rangle)$.

Proof: We show by induction that $ord_P(p) \geq ord_Q(f(p))$ for all $p \in dom(f)$, and the lemma follows because f is a surjection. Assume $p \in dom(f)$ and the claim holds for all $p' \in dom(f)$ with $ord_P(p') < ord_P(p)$. If $q \in Q$ and $q < f(p)$, then there is $p' \leq_P p$ such that $f(p') = q$. Clearly, $p' < p$, so $ord_P(p') < ord_P(p)$. By induction hypothesis, $ord_P(p') \geq ord_Q(q)$, hence $ord_P(p) > ord_Q(q)$ for all $q < f(p)$. It follows that $ord_P(p) \geq ord_Q(f(p))$. □

Corollary 2.2.12 *If $\langle P, \leq_P\rangle$ and $\langle Q, \leq_Q\rangle$ are well-founded and $f\colon P \to Q$ is an injective function that is monotonic (i.e., $p \leq_P p' \Longrightarrow f(p) \leq_Q f(p')$ for all $p, p' \in P$) then $Ord(\langle P, \leq_P\rangle) \leq Ord(\langle Q, \leq_Q\rangle)$.*

Proof: Define $g\colon\subseteq Q \to P$ by $dom(g) = range(f)$ and $g(q) = p \iff q = f(p)$, which is well-defined because f is injective. It is immediate that g is a surjection. If $p \leq_P g(q)$, then $f(p) \leq_Q f(g(q)) = q$ by the monotonicity of f and definition of g. Since $f(p) \leq_Q q$ and $g(f(p)) = p$, Lemma 2.2.11 implies $Ord(\langle P, \leq_P\rangle) \leq Ord(\langle Q, \leq_Q\rangle)$. □

Recall that $Bad(\langle P, \leq_P\rangle)$ is the set of all finite bad sequences of elements of P. We will partially order $s, t \in Bad(\langle P, \leq_P\rangle)$ by $s \preceq t$ if and only if s is an initial prefix of t. The following proposition is well known (for example, see [22]).

Proposition 2.2.13 *$\langle Bad(\langle P, \leq_P\rangle), \succeq\rangle$ is well-founded if and only if $\langle P, \leq_P\rangle$ is a well-partial order.*

Proof: If $\langle Bad(\langle P, \leq_P\rangle), \succeq\rangle$ is not well-founded, then there is a sequence $s_0 \prec s_1 \prec s_2 \prec \cdots$ of bad sequences of elements of P. We can assume without loss of generality that each $s_i \neq \varepsilon$ (the empty sequence). For each $i \in \omega$, s_i is a finite non-empty sequence, so let x_i be the last element of s_i. If $i < j$, then $s_i \prec s_j$, so x_i occurs before x_j in s_j. Since s_j is a bad sequence, $x_i \not\leq_P x_j$ holds. Therefore, $x_0, x_1, x_2, \ldots$ is an infinite bad sequence, so $\langle P, \leq_P\rangle$ is not a well-partial-order.

For the converse, if $x_0, x_1, x_2, \ldots$ is an infinite bad sequence of elements of P, then $\langle x_0\rangle \prec \langle x_0, x_1\rangle \prec \langle x_0, x_1, x_2\rangle \prec \cdots$ is an infinite "descending" chain in $\langle Bad(\langle P, \leq_P\rangle), \succeq\rangle$. □

We will write $Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)$ for the subset of $Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)$ of non-empty sequences. We now define $Ord_B(\langle P, \leq_P\rangle) = \langle Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle), \succeq\rangle$ for any well-partial-order $\langle P, \leq_P\rangle$. Thus, abbreviating $ord(s)$ for $ord_{Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)}(s)$, for each $s \in Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)$ we have:

$$ord(s) = \bigvee\{ord(t) + 1 \mid s \prec t \in Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)\},$$

and

$$Ord_B(\langle P, \leq_P\rangle) = \bigvee\{ord(s) + 1 \mid s \in Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)\}.$$

Since the empty sequence ε is a strict subsequence of every $s \in Bad^{\neq\varepsilon}(\langle P, \leq_P\rangle)$, we could alternatively define $Ord_B(\langle P, \leq_P\rangle) = ord(\varepsilon)$, where $ord(\varepsilon)$ is given the obvious meaning.

The next theorem shows that $Ord_B(\langle \mathcal{L}, \supseteq\rangle)$ is an upper bound on the mind-change complexity of identifying $\mathcal{L}^{<\omega}$ in the limit from positive data.

Theorem 2.2.14 *If $\mathcal{L}$ has finite thickness and $\langle \mathcal{L}, \supseteq\rangle$ is a well-partial-order, then $Ord(\langle \mathcal{L}^{<\omega}, \supseteq\rangle) \leq Ord_B(\langle \mathcal{L}, \supseteq\rangle)$.*

Proof: Define $f\colon Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq\rangle) \to \mathcal{L}^{<\omega}$ so that $f(s) = L_0 \cup \cdots \cup L_n$ for each $s = \langle L_0, \ldots, L_n\rangle \in Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq\rangle)$. Clearly, f is a surjection because every $U \in \mathcal{L}^{<\omega}$ can be written as a union $L_0 \cup \cdots \cup L_n$ with $L_i \not\supseteq L_j$ for $i < j$.

Assume $s = \langle L_0, \ldots, L_n\rangle \in Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq\rangle)$, $U \in \mathcal{L}^{<\omega}$, and $f(s) \subseteq U$. If $f(s) \neq U$, then there is $x_0 \in U \setminus f(s)$ and some $L_0' \in \mathcal{L}$ such that $x_0 \in L_0' \subseteq U$.

Since $x_0 \notin f(s)$, $L_0' \not\subseteq L_i$ for $0 \leq i \leq n$. Thus, $t_0 = \langle L_0, \ldots, L_n, L_0' \rangle \in Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq \rangle)$ and $s \prec t_0$. If $f(t_n) \neq U$, then we can find $x_{n+1} \in U \setminus f(t_n)$ and $L_{n+1}' \subseteq U$ containing x_{n+1} to construct a bad sequence $t_{n+1} \succ t_n$. Since $Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq \rangle)$ is well-founded, we can only repeat this process finitely many times until we have obtained a bad sequence $t \succ s$ satisfying $f(t) = U$.

It follows from Lemma 2.2.11 that $Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle) \leq Ord_B(\langle \mathcal{L}, \supseteq \rangle)$. □

Given a concept space $\mathcal{L}$, we can define a *preorder* (i.e., a reflexive, transitive relation) on ω by $x \preceq_{\mathcal{L}} y \iff x \in C_{\mathcal{L}}(\{y\})$ for all $x, y \in \omega$. Note that this is equivalent to defining $x \preceq_{\mathcal{L}} y \iff C_{\mathcal{L}}(\{x\}) \subseteq C_{\mathcal{L}}(\{y\})$. Define $x \equiv_{\mathcal{L}} y$ if and only if $x \preceq_{\mathcal{L}} y$ and $y \preceq_{\mathcal{L}} x$. We will let $\omega/\equiv_{\mathcal{L}}$ denote the set of equivalence classes of $\equiv_{\mathcal{L}}$, and view $\preceq_{\mathcal{L}}$ as a partial order on $\omega/\equiv_{\mathcal{L}}$. We can think of $\omega/\equiv_{\mathcal{L}}$ as the subset of $\mathcal{A}(\mathcal{L})$ of all closed sets generated by a singleton, in which case $\preceq_{\mathcal{L}}$ is like the "restriction" of the subset relation on $\mathcal{A}(\mathcal{L})$ to $\omega/\equiv_{\mathcal{L}}$. It follows from Theorem 2.1.25 that if $\mathcal{L}$ has finite thickness and $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order, then $\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle$ is a well-partial-order. This gives us another method of computing an upper bound on the mind-change complexity of $\mathcal{L}^{<\omega}$.

Theorem 2.2.15 *If $\mathcal{L}$ has finite thickness and $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order, then $Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle) \leq Ord_B(\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle)$.*

Proof: Define

$$f(\langle x_0, \ldots, x_n \rangle) = C_{\mathcal{L}}(\{x_0\}) \cup \cdots \cup C_{\mathcal{L}}(\{x_n\})$$

for each $\langle x_0, \ldots, x_n \rangle \in Bad^{\neq\varepsilon}(\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle)$.

First we show that for each $U \in \mathcal{L}^{<\omega}$, there is $s \in Bad^{\neq\varepsilon}(\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle)$ such that $f(s) = U$. Let $U \in \mathcal{L}^{<\omega}$ be given, choose any $x_0 \in U$ and define $s_0 = \langle x_0 \rangle$. Working recursively, if $f(s_n) \neq U$, then choose any $x_{n+1} \in U \setminus f(s_n)$. Since $x_{n+1} \notin C_{\mathcal{L}}(\{x_i\})$ for $0 \leq i \leq n$, $s_{n+1} =_{def} s_n\langle x_{n+1} \rangle \in Bad^{\neq\varepsilon}(\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle)$. Clearly $C_{\mathcal{L}}(\{x_{n+1}\}) \subseteq U$, so $f(s_{n+1}) \subseteq U$. Since $Bad^{\neq\varepsilon}(\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle)$ is well-founded, this process terminates after a finite number of steps resulting in a bad sequence s satisfying $f(s) = U$.

Assume $s = \langle L_0, \ldots, L_n \rangle \in Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq \rangle)$, $U \in \mathcal{L}^{<\omega}$, and $f(s) \subseteq U$. If $f(s) \neq U$, then we can choose any $x_0 \in U \setminus f(s)$ and $s\langle x_0 \rangle$ will be a bad sequence with $f(s\langle x_0 \rangle) \subseteq U$. Repeating the argument in the previous paragraph, we obtain a bad sequence $t \succ s$ such that $f(t) = U$.

By restricting the domain of f to $dom(f) = \{s \in Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq \rangle) \mid f(s) \in \mathcal{L}^{<\omega}\}$, we see that $f\!: \subseteq Bad^{\neq\varepsilon}(\langle \mathcal{L}, \supseteq \rangle) \to \mathcal{L}^{<\omega}$ satisfies the requirements of Lemma 2.2.11, hence $Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle) \leq Ord_B(\langle \omega/\equiv_{\mathcal{L}}, \succeq_{\mathcal{L}} \rangle)$. □

Next we give a method for computing a lower bound for the mind-change complexity of $\mathcal{L}^{<\omega}$.

Theorem 2.2.16 *Assume $\mathcal{L}$ has finite thickness and $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order. If $f\!: \alpha \to \mathcal{L}$ is such that for every finite sequence $\alpha > \beta_0 > \cdots > \beta_n > \gamma$, $f(\gamma) \not\subseteq \bigcup_{0 \leq i \leq n} f(\beta_i)$, then $\alpha \leq Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle)$.*

Proof: Assume for a contradiction there is $\beta_0 < \alpha$ such that $ord_{\mathcal{L}^{<\omega}}(f(\beta_0)) < \beta_0$. Let $L_0 = f(\beta_0)$, and note that $ord_{\mathcal{L}^{<\omega}}(L_0) < \beta_0$. Working recursively, for $n \geq 0$ we can assume that $ord_{\mathcal{L}^{<\omega}}(L_n) < \beta_n$. Now let $\beta_{n+1} = ord_{\mathcal{L}^{<\omega}}(L_n)$ and $L_{n+1} = f(\beta_{n+1}) \cup L_n$. Since $\alpha > \beta_0 > \cdots > \beta_n > \beta_{n+1}$, $f(\beta_{n+1}) \not\subseteq L_n =$

$\bigcup_{0 \le i \le n} f(\beta_i)$, hence $L_{n+1} \supset L_n$. Therefore, $ord_{\mathcal{L}^{<\omega}}(L_{n+1}) < \beta_{n+1}$. Repeating this proccess results in an infinite sequence $\alpha > \beta_0 > \beta_1 > \cdots$ of strictly descending ordinals, a contradiction. Thus, $ord_{\mathcal{L}^{<\omega}}(f(\beta)) \ge \beta$ for all $\beta < \alpha$. It follows that $\alpha \le Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle)$. □

To conclude this section, we give a simple example of how to apply the above theorems. We first give a proposition that is useful for computing upper bounds on $Ord_B(\langle P, \le_P \rangle)$. We will omit the easy proof (see Simpson [55], Hasegawa [22]).

Proposition 2.2.17 *Let α be an ordinal and $\langle P, \le_P \rangle$ a partial-order. If there is a function $f\colon Bad(\langle P, \le_P \rangle) \to \alpha+1$ such that $f(s) > f(s\langle x \rangle)$ for every $s, s\langle x \rangle \in Bad(\langle P, \le_P \rangle)$, then $\langle P, \le_P \rangle$ is a well-partial-order and $Ord_B(\langle P, \le_P \rangle) \le \alpha$.* □

The function f in the above proposition is called a *reification*. We will construct a reification in the following example (a similar reification was constructed in [55]).

Example 2.2.18 Let $L_{\langle x,y \rangle} = \{\langle x', y' \rangle \in \omega \times \omega \mid x \le x' \text{ and } y \le y'\}$, and $\mathcal{L} = \{L_{\langle x,y \rangle} \mid \langle x', y' \rangle \in \omega \times \omega\}$. Clearly, $L_{\langle x,y \rangle} = C_{\mathcal{L}}(\{\langle x, y \rangle\})$, and $L_{\langle x,y \rangle} \supseteq L_{\langle x',y' \rangle} \iff \langle x', y' \rangle \in C_{\mathcal{L}}(\{\langle x, y \rangle\}) \iff \langle x', y' \rangle \preceq_{\mathcal{L}} \langle x, y \rangle$. It is immediate that $\mathcal{L}$ has finite thickness, because for any $\langle x, y \rangle$, there are only finitely many $\langle x', y' \rangle$ such that $x' < x$ and $y' < y$. It will follow from the results below on the upper bound of $Ord_B(\langle \mathcal{L}, \supseteq \rangle)$ that $\langle \mathcal{L}, \supseteq \rangle$ is a well-partial-order.

We first show how to apply Theorem 2.2.16 to give a lower bound on $Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle)$. Consider the ordinal $\omega^2 = \omega \cdot \omega$. Every ordinal strictly less than ω^2 can be written as $\omega \cdot x + y$ with $x, y < \omega$, and $\omega \cdot x + y \le \omega \cdot x' + y'$ if and only if $x < x'$ or ($x = x'$ and $y \le y'$) (see Example 2.2.5). Define a function $f\colon \omega^2 \to \mathcal{L}$ so that $f(\omega \cdot x + y) = L_{\langle x,y \rangle}$. Now, if $\omega^2 > \omega \cdot x_0 + y_0 > \cdots > \omega \cdot x_n + y_n > \omega \cdot x + y$, then for each $i \le n$, either $x < x_i$ or ($x = x_i$ and $y < y_i$), so $\langle x, y \rangle \notin L_{\langle x_i, y_i \rangle}$. Therefore, $L_{\langle x,y \rangle} \not\subseteq \bigcup_{0 \le i \le n} L_{\langle x_i, y_i \rangle}$. Thus, $f\colon \omega^2 \to \mathcal{L}$ satisfies the requirements of Theorem 2.2.16, so $\omega^2 \le Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle)$.

Next we give an upper bound on $Ord(\langle \mathcal{L}^{<\omega}, \supseteq \rangle)$ by constructing a reification $g\colon Bad(\langle \mathcal{L}, \supseteq \rangle) \to \omega^2 + 1$. First, let $g(\varepsilon) = \omega^2$, where ε is the empty sequence.

For every non-empty bad sequence s, we will assign finite sequences A_s and B_s of ordinals as follows. For every single element sequence $\langle L_{\langle x,y \rangle} \rangle$, define A_s to be the sequence $\langle \omega, \omega, \cdots, \omega \rangle$ of length x (thus, if $x = 0$ then A_s is the empty sequence), and define B_s similarly but of length y. Next, assume s is non-empty, and $s\langle L_{\langle x,y \rangle} \rangle$ is a bad sequence. If $L_{\langle x',y' \rangle}$ is the first element of s, then since $L_{\langle x',y' \rangle} \not\supseteq L_{\langle x,y \rangle}$, either $x < x'$ or $y < y'$. If $x < x'$, then define $A_{s\langle L_{\langle x,y \rangle} \rangle}$ to be the sequence of length x' such that $A_{s\langle L_{\langle x,y \rangle} \rangle}(x) = y$ and $A_{s\langle L_{\langle x,y \rangle} \rangle}(i) = A_s(i)$ for $i \ne x$, and define $B_{s\langle L_{\langle x,y \rangle} \rangle} = B_s$. If $x \ge x'$, then $y < y'$ so define $B_{s\langle L_{\langle x,y \rangle} \rangle}$ to be the sequence of length y' such that $B_{s\langle L_{\langle x,y \rangle} \rangle}(y) = x$ and $B_{s\langle L_{\langle x,y \rangle} \rangle}(i) = B_s(i)$ for $i \ne y$, and define $A_{s\langle L_{\langle x,y \rangle} \rangle} = A_s$. So, for example:

$$
\begin{aligned}
A_{\langle L_{\langle 3,2 \rangle} \rangle} &= \langle \omega, \omega, \omega \rangle \\
B_{\langle L_{\langle 3,2 \rangle} \rangle} &= \langle \omega, \omega \rangle \\
A_{\langle L_{\langle 3,2 \rangle}, L_{\langle 2,4 \rangle} \rangle} &= \langle \omega, \omega, 4 \rangle \\
B_{\langle L_{\langle 3,2 \rangle}, L_{\langle 2,4 \rangle} \rangle} &= \langle \omega, \omega \rangle \\
A_{\langle L_{\langle 3,2 \rangle}, L_{\langle 2,4 \rangle}, L_{\langle 5,1 \rangle} \rangle} &= \langle \omega, \omega, 4 \rangle
\end{aligned}
$$

$$\begin{aligned} B_{\langle L_{\langle 3,2\rangle}, L_{\langle 2,4\rangle}, L_{\langle 5,1\rangle}\rangle} &= \langle \omega, 5\rangle \\ A_{\langle L_{\langle 3,2\rangle}, L_{\langle 2,4\rangle}, L_{\langle 5,1\rangle}, L_{\langle 2,1\rangle}\rangle} &= \langle \omega, \omega, 1\rangle \\ B_{\langle L_{\langle 3,2\rangle}, L_{\langle 2,4\rangle}, L_{\langle 5,1\rangle}, L_{\langle 2,1\rangle}\rangle} &= \langle \omega, 5\rangle \end{aligned}$$

It is clear that for every non-empty s, either $A_{s\langle L_{\langle x,y\rangle}\rangle} \neq A_s$ and $B_{s\langle L_{\langle x,y\rangle}\rangle} = B_s$, or else $A_{s\langle L_{\langle x,y\rangle}\rangle} = A_s$ and $B_{s\langle L_{\langle x,y\rangle}\rangle} \neq B_s$. In the first case, if $A_{s\langle L_{\langle x,y\rangle}\rangle} \neq A_s$, then they are only different in the $(x+1)$th element. By definition, $A_{s\langle L_{\langle x,y\rangle}\rangle}(x)$ equals y. Let $y' = A_s(x)$. If $y' \neq \omega$, then by construction of A_s, $L_{\langle x,y'\rangle}$ must occur in s. Since $s\langle L_{\langle x,y\rangle}\rangle$ is a bad sequence, $L_{\langle x,y'\rangle} \not\supseteq L_{\langle x,y\rangle}$, so it must be the case that $y < y'$. Thus, $A_{s\langle L_{\langle x,y\rangle}\rangle}(x) < A_s(x)$ and $A_{s\langle L_{\langle x,y\rangle}\rangle}(i) = A_s(i)$ for all $i \neq x$. We can similarly show that $B_{s\langle L_{\langle x,y\rangle}\rangle}(y) < B_s(y)$ and $B_{s\langle L_{\langle x,y\rangle}\rangle}(i) = B_s(i)$ for all $i \neq y$ whenever $B_{s\langle L_{\langle x,y\rangle}\rangle} \neq B_s$.

For each non-empty sequence s, let S_s be the finite sequence of ordinals formed by sorting the concatenation of A_s and B_s into descending order. For example, if $A_s = \langle 3, \omega, 6\rangle$ and $B_s = \langle \omega, 8, \omega, 2\rangle$, then $S_s = \langle \omega, \omega, \omega, 8, 6, 3, 2\rangle$. We define $g(s) = 0$ if S_s is empty, and otherwise define $g(s) = S_s(0) + S_s(1) + \cdots + S_s(n-1)$ where $n > 0$ is the length of S_s. For example:

$$\begin{aligned} g(\langle L_{\langle 3,2\rangle}\rangle) &= \omega+\omega+\omega+\omega+\omega \\ g(\langle L_{\langle 3,2\rangle}, L_{\langle 2,4\rangle}\rangle) &= \omega+\omega+\omega+\omega+4 \\ g(\langle L_{\langle 3,2\rangle}, L_{\langle 2,4\rangle}, L_{\langle 5,1\rangle}\rangle) &= \omega+\omega+\omega+5+4 \\ g(\langle L_{\langle 3,2\rangle}, L_{\langle 2,4\rangle}, L_{\langle 5,1\rangle}, L_{\langle 2,1\rangle}\rangle) &= \omega+\omega+\omega+5+1 \end{aligned}$$

Since $A_{s\langle L_{\langle x,y\rangle}\rangle}(i) \leq A_s(i)$ and $B_{s\langle L_{\langle x,y\rangle}\rangle}(i) \leq B_s(i)$ for all coordinates i, and at least one coordinate is strictly smaller, it is easy to see that our construction guarantees that $g(s) > g(s\langle L_{\langle x,y\rangle}\rangle)$.

Thus, g satisfies Proposition 2.2.17, so $Ord_B(\langle \mathcal{L}, \supseteq\rangle) \leq \omega^2$. By Theorem 2.2.14, it follows that $Ord(\langle \mathcal{L}^{<\omega}, \supseteq\rangle) \leq \omega^2$. It follows that the the mind-change complexity of identifying $\mathcal{L}^{<\omega}$ from positive data is ω^2. □

2.3 Mind-change complexity of $L(\mathbf{RP}^l)^{<\omega}$

In this section, we apply the results of the previous sections to compute upper and lower bounds for the mind-change complexity of identifying unbounded unions of restricted pattern languages from positive data. The mind-change complexity of identifying n unions of pattern languages was shown by Jain and Sharma [26] to be equal to ω^n. Furthermore, Stephan and Ventsov [57] have shown that the mind-change complexity of identifying ideals over the polynomial ring with rational coefficients and n variables is ω^n. Here we show that the mind-change complexity of identifying unbounded unions of a restricted set of pattern languages is of the form $\omega^{\omega^n} + m$, with $m, n < \omega$. This is the first time to the author's knowledge that a natural class of languages has been shown to have mind-change complexity greater than ω^ω.

2.3.1 Pattern languages

Given an alphabet Σ, we use $\Sigma^{<l}$, $\Sigma^{\leq l}$, $\Sigma^{=l}$, Σ^*, to denote the set of all strings of Σ of length less than l, less than or equal to l, exactly equal to l, or of finite

length, respectively. Σ^+ is the subset of Σ^* of non-empty strings.

Pattern languages were originally introduced into inductive inference by Angluin [3] and later became a rich field of research. Let Σ be a finite alphabet and let $V = x_0, x_1, \ldots$ be a countably infinite set of symbols disjoint from Σ. A finite string of elements of Σ is called a *constant segment* and elements of V are called *variables*. A *pattern* is a non-empty finite string over $\Sigma \cup V$. A pattern p is said to be *regular* if every variable x_i appearing in p occurs only once. Let $\mathbf{RP}$ be the set of regular patterns, and let $\mathbf{RP}^l$ be the set of regular patterns which contain constant segments of length no longer than l.

The language of a pattern p, denoted $L(p)$, is the subset of Σ^* that can be obtained by substituting a non-empty constant segment $s_i \in \Sigma^+$ for each occurrence of the variable x_i in p for all $i \geq 0$. For example, if $\Sigma = \{a, b\}$ and $p = ax_1b$, then $L(p)$ is the subset of Σ^* of strings beginning with "a", ending with "b", and of length greater than or equal to three. For a set of patterns P, we define $L(P) = \{L(p) \,|\, p \in P\}$. We can assume that pattern languages are properly encoded as sets of natural numbers, so $L(P)$ can be interpreted as a concept space.

The next two theorems will be useful for showing the mind-change complexity of $L(\mathbf{RP}^l)^{<\omega}$, the class of unbounded unions of languages of regular patterns with constant segment length bound l.

Theorem 2.3.1 (Shinohara & Arimura [54]) *For any $l \geq 1$, $L(\mathbf{RP}^l)$ has finite thickness and contains no infinite anti-chains with respect to set inclusion.*

Theorem 2.3.2 (Shinohara & Arimura [54]) *For any $l \geq 1$, $L(\mathbf{RP}^l)^{<\omega}$ is inferable from positive data.*

2.3.2 Lower bound for $L(\mathbf{RP}^l)^{<\omega}$

We first give a lower bound on the mind-change complexity of identifying $L(\mathbf{RP}^l)^{<\omega}$ from positive data. For the following lemma, recall that given a set A, A^* is the set of finite strings of elements of A, and $\preceq_S$ is the subsequence relation.

Lemma 2.3.3 (Simpson [55]) *Let A be a finite set containing exactly k elements. Then there exists a function $g : \omega^{\omega^{k-1}} \to A^*$ with the property that $\alpha \not\leq \beta$ implies $g(\alpha) \not\preceq_S g(\beta)$.*

Let Σ be a finite alphabet containing at least two elements. We fix an element $c \in \Sigma$ and let $\Sigma_{-c} = \Sigma - \{c\}$. We will abbreviate $(\Sigma_{-c})^{=l}$, the set of strings of elements of Σ_{-c} with length l, as $\Sigma_{-c}^{=l}$. We define $x(\mathbf{RP}_{-c}^{=l})y$ to be the subset of $\mathbf{RP}^l$ of patterns that begin and end with a variable, do not contain any occurrences of the constant element c, and only have constant segments of length exactly equal to l. Note that although no $p \in x(\mathbf{RP}_{-c}^{=l})y$ contains the element c, $L(p)$ is defined over Σ, so c may occur in some elements of the language $L(p)$.

Next we define $P : (\Sigma_{-c}^{=l})^* \to x(\mathbf{RP}_{-c}^{=l})y$ so that

$$P(\langle w_1, \ldots, w_n \rangle) = x_1 w_1 \cdots x_n w_n x_{n+1}.$$

Let $\preceq'_S$ be the subsequence relation on $(\Sigma_{-c}^{=l})^*$. The following lemma is related to a theorem proved by Mukouchi [39].

Lemma 2.3.4 *Let $\sigma, \tau_1, \ldots, \tau_n \in (\Sigma_{-c}^{=l})^*$ for $n \geq 1$. If for all i $(1 \leq i \leq n)$, $\tau_i \not\preceq'_S \sigma$, then $L(P(\sigma)) \not\subseteq \bigcup_{1 \leq i \leq n} L(P(\tau_i))$.*

Proof: Let $\sigma = \langle w_1, \ldots, w_m \rangle$, and let $s = cw_1c \cdots cw_mc$. Obviously $s \in L(P(\sigma))$. Assume $s \in \bigcup_{1 \leq i \leq n} L(P(\tau_i))$, then for some j, $s \in L(P(\tau_j))$. Assume $\tau_j = \langle u_1, \ldots, u_{m'} \rangle$, so $P(\tau_j) = x_1u_1 \cdots x_{m'}u_{m'}x_{m'+1}$. Each constant segment $u_{i'}$ $(1 \leq i' \leq m')$ in $P(\tau_j)$ must map to a segment in s, but since $u_{i'}$ does not contain c, $u_{i'}$ must appear within some $w_{k_{i'}}$ $(1 \leq k_{i'} \leq m)$. Since $|u_{i'}| = |w_{k_{i'}}| = l$, it follows that $u_{i'} = w_{k_{i'}}$. Furthermore, the ordering of the mapping must be preserved, so $k_{i'} < k_{i'+1}$ for $i' < m'$. But this shows that $\tau_j \preceq'_S \sigma$, which contradicts the hypothesis. □

We now give a lower bound for the mind-change complexity of $L(\mathbf{RP}^l)^{<\omega}$.

Theorem 2.3.5 *$L(\mathbf{RP}^l)^{<\omega}$ is not identifiable from positive data with mind-change bound less than $\omega^{\omega^{(|\Sigma|-1)^l-1}}$ for $l \geq 1$ and for finite Σ containing at least two elements.*

Proof: Since $|\Sigma_{-c}^{=l}| = |\Sigma_{-c}|^l = (|\Sigma|-1)^l$, we can use Lemma 2.3.3 to define a mapping $g : \omega^{\omega^{(|\Sigma|-1)^l-1}} \to (\Sigma_{-c}^{=l})^*$ with the property that $\alpha > \beta$ implies $g(\alpha) \not\preceq'_S g(\beta)$. We now define $f : \omega^{\omega^{(|\Sigma|-1)^l-1}} \to L(x(\mathbf{RP}_{-c}^{=l})y)$ to be $f(\alpha) = L(P(g(\alpha)))$. It follows from Lemma 2.3.4 that if $\omega^{\omega^{(|\Sigma|-1)^l-1}} > \alpha_0 > \cdots > \alpha_n > \beta$ for finite n, then $f(\beta) \not\subseteq \bigcup_{0 \leq i \leq n} f(\alpha_i)$.

It follows from Theorem 2.2.16 that $\omega^{\omega^{(|\Sigma|-1)^l-1}} \leq Ord(\langle L(\mathbf{RP}^l)^{<\omega}, \supseteq \rangle)$. Therefore, by Corollary 2.2.8, $L(\mathbf{RP}^l)^{<\omega}$ is not inferable from positive data with mind-change bound less than $\omega^{\omega^{(|\Sigma|-1)^l-1}}$. □

2.3.3 Upper bound for $L(\mathbf{RP}^l)^{<\omega}$

We next give an upper bound on the mind-change complexity of identifying $L(\mathbf{RP}^l)^{<\omega}$ from positive data.

Lemma 2.3.6 (Simpson [55], Hasegawa [22]) *Let A be a finite set containing exactly k elements. Then there exists a function $f : Bad(\langle A^*, \preceq_S \rangle) \to \omega^{\omega^{k-1}} + 1$ with the property that $f(s\langle a \rangle) < f(s)$ for all $s, s\langle a \rangle \in Bad(\langle A^*, \preceq_S \rangle)$.*

Let C_{RP_l} be the algebraic closure operator on $L(\mathbf{RP}^l)$ as in Definition 2.1.6, and let $\mathcal{A}(L(\mathbf{RP}^l))$ be the set of all fixed points of C_{RP_l}. Thus, in particular, $C_{RP_l}(F) = \bigcap \{L \in L(\mathbf{RP}^l) \mid F \subseteq L\}$ for any finite subset F of Σ^*. Theorem 2.3.1 implies that $\langle L(\mathbf{RP}^l), \supseteq \rangle$ is a well-partial order, and since $L(\mathbf{RP}^l)$ has finite thickness, $\langle \mathcal{A}(L(\mathbf{RP}^l)), \supseteq \rangle$ is also a well-partial-order by Theorem 2.1.25.

Let Σ be a finite alphabet containing at least two elements, and let # be a new symbol not in Σ. Define $\Sigma_{\#}^{=l}$ to be the set of elements of $\Sigma^{=l}$ with the symbol # appended to the beginning or end. Define a mapping $h : \Sigma^{>l} \to (\Sigma_{\#}^{=l})^*$ such that for $s = a_1 \cdots a_n$ $(n > l)$,

$$h(s) = \langle \#a_1 \cdots a_l, a_2 \cdots a_{l+1}\#, \ldots, \#a_{n-l+1} \cdots a_n \rangle,$$

where # appears on the left side of the initial and final segments, and # appears on the right of all other segments.

Lemma 2.3.7 *For any $s, t \in \Sigma^*$, if $|s| \leq l$ and $s = t$, or if $h(s) \preceq_S h(t)$, then $t \in C_{RP_l}(\{s\})$.*

Proof: The case where $s = t$ is obvious, so assume $s = a_1 \cdots a_n$, $t = b_1 \cdots b_{n'}$, and that $h(s)$ is a subsequence of $h(t)$. It follows that there is a strict monotonic function $f : \{1, \ldots, n-l+1\} \to \{1, \ldots, n'\}$ such that for all i $(1 \leq i \leq n-l+1)$, $a_i \cdots a_{i+l-1} = b_{f(i)} \cdots b_{f(i)+l-1}$. We now define $f' : \{1, \ldots, n\} \to \{1, \ldots, n'\}$ so that $f'(i) = f(i)$ when $i \leq n - l + 1$ and $f'(i) = n' - n + i$ otherwise. By definition f' is strict monotonic, implying that $f'(j) - f'(i) \geq j - i$ for all $j \geq i$, where j and i are in the domain of f'. Also, note that the placement of the # symbols guarantee that first and last l elements of s and t are the same. Therefore, for all i $(1 \leq i \leq n)$ and all k $(1 \leq k \leq l)$, if $i + k - 1 \leq n$ then $a_i \cdots a_{i+k-1} = b_{f'(i)} \cdots b_{f'(i)+k-1}$.

Let $p = w_1 x_1 \cdots w_m x_m w_{m+1}$ be a pattern in $\mathbf{RP}^l$ such that $s \in L(p)$, where the x_i's are variables and the w_i's are in $\Sigma^{\leq l}$. For each non-empty w_i $(1 \leq i \leq m+1)$ in p, w_i is mapped to a segment in s, so let k_i be the position in s where the first element of w_i is mapped. If w_i is empty for $i \leq m$ then let k_i be the position of the first element of s that x_i matches. We can ignore the case that w_{m+1} is empty. Note that $k_{i+1} \geq k_i + |w_i| + 1$ for $i \leq m$, since x_i must be mapped to at least one element of s.

For $i \leq m+1$ and non-empty w_i, it is clear that $w_i = a_{k_i} \cdots a_{k_i + |w_i| - 1} = b_{f'(k_i)} \cdots b_{f'(k_i) + |w_i| - 1}$, so each w_i matches a segment of t in the proper order. The strict monotonicity of f' guarantees that there is at least as much space between $f'(k_i)$ and $f'(k_i + 1)$ as there is between k_i and $k_i + 1$, so there is enough room for each x_i to match at least one element of t. If w_i is empty then x_i would match a segment of t starting at $f'(k_i)$, otherwise x_i would start matching at $f'(k_i) + |w_i|$. In either case, x_i would match the segment of t up to the position $f'(k_{i+1}) - 1$, or to the end of the string if w_{m+1} is empty. Also, the first and last l elements of s and t are associated by f', so w_1 and w_{m+1} match to the head and tail segments of t. Therefore, we can see that $t \in L(p)$. Since p was arbitrary, we can conclude that $t \in C_{RP_l}(\{s\})$. $\square$

Theorem 2.3.8 *$L(\mathbf{RP}^l)^{<\omega}$ is identifiable from positive data with mind-change bound $\omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$ for any $l \geq 1$ and finite Σ containing at least two elements.*

Proof: Define a preorder $\preceq_{RP_l}$ on elements of Σ^+ by $s \preceq_{RP_l} t \iff t \in C_{RP_l}(\{s\})$. Since every $L \in L(\mathbf{RP}^l)$ is a subset of Σ^+, by Theorem 2.2.15 it suffices to show that $Ord_B(\langle \Sigma^+ / \equiv_{RP_l}, \succeq_{RP_l} \rangle) \leq \omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$.

Since $\Sigma_{\#}^{=l}$ contains $2|\Sigma|^l$ elements, Lemma 2.3.6 implies that there is a function $f : Bad(\langle (\Sigma_{\#}^{=l})^*, \preceq_S \rangle) \to \omega^{\omega^{2|\Sigma|^l - 1}} + 1$ such that $f(s\langle a \rangle) < f(s)$ for all $s, s\langle a \rangle \in Bad(\langle A^*, \preceq_S \rangle)$.

We define a function $g : Bad(\langle \Sigma^+ / \equiv_{RP_l}, \succeq_{RP_l} \rangle) \to \omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}| + 1$ such that $g(s\langle x \rangle) < g(s)$ for all $s, s\langle x \rangle \in Bad(\langle \Sigma^+ / \equiv_{RP_l}, \succeq_{RP_l} \rangle)$. It will then follow from Proposition 2.2.17 that $Ord_B(\langle \Sigma^+ / \equiv_{RP_l}, \succeq_{RP_l} \rangle) \leq \omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$.

Define $g(\varepsilon) = \omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$. Assume that $g(s)$ has been defined and $s\langle x \rangle \in Bad(\langle \Sigma^+ / \equiv_{RP_l}, \succeq_{RP_l} \rangle)$. Let t be the subsequence of $s\langle x \rangle$ of exactly the elements with length greater than l, and let r be the number of elements in $s\langle x \rangle$

with length less than or equal to l. Note that $r < |\Sigma^{\leq l}|$, because $\varepsilon \in \Sigma^{\leq l}$ but $\varepsilon \notin \Sigma^+$. Let h be the function from Lemma 2.3.7. Define $h_t = \varepsilon$ if t is empty, and otherwise define $h_t = \langle h(t(0)), h(t(1)), \ldots, h(t(n-1))\rangle$ where $n > 0$ is the length of t. If t is non-empty, then since t is a bad sequence, $t(i) \not\preceq_{RP_l} t(j)$ hence $t(j) \notin C_{RP_l}(\{t(i)\})$ for all $i < j$, thus $h_t(i) \not\preceq_S h_t(j)$ by Lemma 2.3.7. Therefore, h_t in $Bad(\langle(\Sigma_{\#}^{=l})^*, \preceq_S\rangle)$. We define $g(s\langle x\rangle) = f(t) + (|\Sigma^{\leq l}| - r)$. Note that either t is non-empty or else $r > 0$, so either $f(t) < \omega^{\omega^{2|\Sigma|^l - 1}}$ or else $(|\Sigma^{\leq l}| - r) < |\Sigma^{\leq l}|$, thus $g(s\langle x\rangle) < \omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$.

Finally, we show that $g(s\langle x\rangle) < g(s)$ for every bad sequence $s\langle x\rangle$. Let t be the subsequence of elements in s that are longer than l, and r the number of elements in s that have length less than or equal to l. Define t', r' similarly, but with respect to $s\langle x\rangle$. Then t is a prefix of t' and $r \geq r'$. Furthermore, it must be the case that either t' is strictly longer that t, or else r' is strictly larger than r. Therefore, by the properties of f and the definition of g, $g(s\langle x\rangle) \leq g(s)$. $\square$

The following algorithm identifies $L(\mathbf{RP}^l)^{<\omega}$ from positive data with mind-change bound $\omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$. The function $h\colon \Sigma^{>l} \to (\Sigma_{\#}^{=l})^*$ in line 12 is the function from Lemma 2.3.7, and the function $f\colon Bad(\langle(\Sigma_{\#}^{=l})^*, \preceq_S\rangle) \to \omega^{\omega^{2|\Sigma|^l - 1}} + 1$ in line 13 is the function f in the proof of Theorem 2.3.8.

Algorithm 1 Algorithm to identify $L(\mathbf{RP}^l)^{<\omega}$ from positive data with mind-change bound $\omega^{\omega^{2|\Sigma|^l - 1}} + |\Sigma^{\leq l}|$. Hypothesis H corresponds to the language $L(H) = \bigcup_{w \in H} C_{RP_l}(w)$.

1: $H := \emptyset$; // initialize as empty set
2: $BadSeq := \varepsilon$; // initialize as empty sequence
3: $counter_A := \omega^{\omega^{2|\Sigma|^l - 1}} + 1$;
4: $counter_B := |\Sigma^{\leq l}| - 1$;
5: $counter := counter_A + counter_B$;
6: **loop**
7:   $s :=$ **Input**(); // receive next element in the presentation
8:   **if** $(\forall w \in H)[s \notin C_{RP_l}(w)]$ **then**
9:     **if** $|s| \leq l$ **then**
10:       $counter_B := counter_B - 1$;
11:     **else**
12:       $BadSeq := BadSeq\langle h(s)\rangle$; // append $h(s)$ to the end of $BadSeq$
13:       $counter_A := f(BadSeq)$;
14:     **end if**
15:     $H := H \cup \{s\}$;
16:     $counter := counter_A + counter_B$;
17:     **Output**($\langle H, counter\rangle$); // output new hypothesis and counter value
18:   **end if**
19: **end loop**

Note that the lower bound shown by Theorem 2.3.5 only involved a subset of $L(\mathbf{RP}^l)^{<\omega}$, so we expect that this bound can be raised somewhat. Furthermore, the reification in Theorem 2.3.8 does not take advantage of all of the properties of pattern languages, so there is still much room for improvement.

Chapter 3

Topological Properties of Concept Spaces

In this chapter we will interpret concept spaces as a topological space and compare topological and other structural properties of concept spaces. We will see later that this topological approach is related to the algebraic closure operator approach via the Scott-topology. Most of the results in this chapter are with respect to identification in the limit from positive data, but the basic concepts introduced are important when we generalize identification in the limit to more abstract forms of information presentation in the following chapter.

From the topological perspective, each concept in a concept space is viewed as a point in an abstract space. The relationship between points (concepts) in this space is determined by the topology, or family of open sets, on the space. The open sets can be interpreted as the "observable" properties of the concepts as we discussed in the introduction.

One philosophical advantage of viewing a concept space as a topological space is that the formal definition of a concept is no longer relevant. To *define* a concept to be a set of natural numbers is rather artificial, and places unnecessary restrictions on which objects can be viewed as concepts, or at least creates the extra burden of determining how to encode a set of concepts as sets of natural numbers. Since a learner can only access a stream of information about a particular concept, the precise mathematical definition of a concept is somewhat arbitrary. Since the structure of a topological space is defined externally as sets of points, by viewing a concept space as a topological space we can avoid giving a precise definition of "concept" and therefore generalize the identification in the limit paradigm to more abstract mathematical objects.

However, the internal structure of a concept (as a set of natural numbers) certainly provides important information about the structure of the concept space, so it is imporant to clarify which properties of concept spaces can be viewed from a completely topological perspective and which properties cannot. In this chapter we will show that several important properties of concept spaces, in particular the property of being identifiable in the limit from positive data, are topological. On the other hand, we will also show that some well known properties, like finite elasticity, are not topological. These results are important for understanding which structural properties we can retain if we were to take a completely topological interpretation of the identification in the limit model.

Another important result in this chapter is that we give completely topological characterizations of reductions between concept spaces. A reduction between concept spaces essentially reduces one learning problem into a different learning problem. Luo and Schulte [34] observed that strong reductions determine a continuous injective function between concept spaces. We extend this observation to prove the converse, that every injective continuous function determines a strong reduction. In addition, we also characterize weak reductions in terms of lower semicontinuous multivalued functions.

We will discuss basic topological properties of concept spaces in the next section. In Section 3.2 we investigate which structural properties of concept spaces are topological. In Section 3.3 we give topological characterizations of reductions between concept spaces. Most results from this chapter have been presented in [15].

3.1 The Positive information topology

In this section, we investigate a topology on concept spaces that is relevant to identification in the limit from positive data.

Given a concept space $\mathcal{L}$ and any subset S of ω, we define

$$\uparrow_{\mathcal{L}} S = \{L \in \mathcal{L} \mid S \subseteq L\} \text{ and } \downarrow_{\mathcal{L}} S = \{L \in \mathcal{L} \mid L \subseteq S\}.$$

Definition 3.1.1 (Luo and Schulte [34]) *Let $\mathcal{L}$ be a concept space. A subset of $\mathcal{L}$ is called a* Π-basic open set *if and only if it is equal to $\uparrow_{\mathcal{L}} F$ for some finite subset F of ω. An arbitrary union (including the empty union) of Π-basic open sets is called a* Π-open set. Π-closed sets *and* Π-clopen *sets are defined as usual. The resulting topology is called the* positive information topology (Π-topology) *on the concept space $\mathcal{L}$. A mapping between concept spaces that is continuous with respect to the Π-topologies is said to be* Π-continuous. □

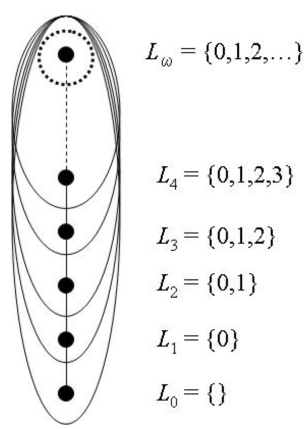


Figure 3.1: The solid lines show Π-open subsets of the concept space $\mathcal{L} = \{L_i\}_{i \leq \omega}$. The dotted circle consisting only of $\{L_\omega\}$ is a subset of $\mathcal{L}$ that is *not* Π-open.

The Π-topology was introduced to the learning community by Luo and Schulte [34] to characterize mind-complexity. The Π-topology is closely related to the *Scott-topology* on partially ordered sets (posets), which is important in domain theory (see [20]). A subset U of a poset $\langle P, \leq \rangle$ is *Scott-open* if and only if the following two conditions hold:

1. U is an *upper set*, i.e. $q \in U$ and $q \leq p$ implies $p \in U$;

2. For every directed $D \subseteq P$ with $\bigvee D$ defined, $\bigvee D \in U$ implies $D \cap U \neq \emptyset$.

The Scott-topology has a particularly simple form for algebraic closure systems. In this case, the Scott-topology on an algebraic closure system $\mathcal{C}$ is the smallest topology on $\mathcal{C}$ containing all sets of the form $\uparrow X = \{Y \in \mathcal{C} \mid X \subseteq Y\}$ for X a compact element of $\mathcal{C}$. Based on this observation, it is easy to see that the Scott-topology and Π-topology on $\mathcal{A}(\mathcal{L})$ coincide for every concept space $\mathcal{L}$.

If Y is a topological space, and X is a subset of Y, the *subspace topology* on X inherited from Y is defined in such a way that $U \subseteq X$ is open if and only if there is open $V \subseteq Y$ such that $U = V \cap X$.

The following proposition, which easily follows from the definitions, gives an alternative characterization of the Π-topology in terms of the Scott-topology.

Proposition 3.1.2 *The* Π*-topology on* $\mathcal{L}$ *coincides with the subspace topology inherited from* $\mathcal{A}(\mathcal{L})$ *with the Scott-topology.*

In this chapter we will usually assume the Π-topology on a concept space $\mathcal{L}$, unless specifically stated otherwise.

3.1.1 Countably based T_0-spaces

Two topological spaces X and Y are *homeomorphic* if and only if there is a continuous bijection $f: X \to Y$ such that the inverse $f^{-1}: Y \to X$ is also continuous. In such a case, f is called a *homeomorphism* from X to Y.

A natural question to ask is which topological spaces are homeomorphic to concept spaces with the Π-topology. To answer this question, we first need some definitions.

Definition 3.1.3 *A topological space X satisfies the T_0 separation axiom (or X is a T_0-space) if and only if for every pair of distinct elements $x, y \in X$, there is an open subset of X that contains exactly one of x and y.* □

Clearly every concept space with the Π-topology is a T_0-space.

Definition 3.1.4 *A set B of open subsets of a topological space X is a* basis *for the topology on X if and only if every open subset of X is equal to the union of a subset of B. A topological space is* countably based *if and only if it has a countable basis.* □

For example, the Π-basic open subsets of a concept space form a countable basis for the Π-topology. Thus, every concept space with the Π-topology is a countably based T_0-space. We next show that the converse of this holds up to homeomorphism.

Proposition 3.1.5 *Every countably based T_0-space is homeomorphic to some concept space with the Π-topology.*

Proof: The basic idea of the proof is well known (for example, see the proof of Lemma II-3.4 (ii) in [20]). Let X be a countably based T_0-space and let $\{B_i\}_{i\in I}$ be a countable base for X, where $I \subseteq \omega$. For $x \in X$, define $\eta(x) = \{i \in I \mid x \in B_i\}$. Define $P(X) = \{\eta(x) \mid x \in X\}$. It is clear that $P(X)$ is a concept space, so it only remains to show that $\eta: X \to P(X)$ is a homeomorphism (i.e., η is a continuous bijection and η^{-1} is also continuous).

We first show that η is a bijection. If $x, y \in X$ are such that $x \neq y$ then the T_0 property guarantees that there is an open set U that contains either x or y, but not both (without loss of generality, assume $x \in U$ and $y \notin U$). Since $\{B_i\}_{i\in I}$ is a basis for X, it follows that there is some $i \in I$ such that $x \in B_i \subseteq U$. Hence, $i \in \eta(x)$ but $i \notin \eta(y)$, so $\eta(x) \neq \eta(y)$. Therefore, η is injective, and since it is surjective by definition, η is a bijection.

To show that a function is continuous, it suffices to show that the preimage of every basic open set is open. Thus, to prove that η is continuous, it suffices to show that $\eta^{-1}(\uparrow_{P(X)} F)$ is open in X for any finite set F of natural numbers. If $F \not\subseteq I$ then $\uparrow_{P(X)} F$ is empty, hence $\eta^{-1}(\uparrow_{P(X)} F)$ is empty, which is open by definition. Otherwise, it is easily seen that $\eta^{-1}(\uparrow_{P(X)} F) = \bigcap_{i\in F} B_i$ which is the intersection of a finite number of open sets and therefore open.

Finally, η^{-1} is continuous because for each $i \in I$, $\eta(B_i) = \uparrow_{P(X)} \{i\}$ which is clearly Π-open in $P(X)$. □

Since every concept space is a countably based T_0-space, Proposition 3.1.5 implies that the category of concept spaces and Π-continuous maps is equivalent to the category of countably based T_0-spaces and continuous maps (see Mac Lane [35] for more on category theory). A simple consequence of this observation is that the category of concept spaces and Π-continuous maps is not cartesian closed, because of the well known fact that the category of countably based T_0-spaces is not cartesian closed (see [20] for characterizations of the topological spaces which are exponentiable).

3.1.2 Π-continuous and Scott-continuous functions

As we noted earlier, the Π-topology on $\mathcal{A}(\mathcal{L})$ is precisely the Scott-topology on $\mathcal{A}(\mathcal{L})$ when viewed as a lattice. It follows that a function $f: \mathcal{A}(\mathcal{K}) \to \mathcal{A}(\mathcal{L})$ is Π-continuous if and only if it is *Scott-continuous* (see [20]) if and only if $f(\bigvee D) = \bigvee f(D)$ for every directed subset D of $\mathcal{A}(\mathcal{K})$.

We have already seen how the Π-topology on $\mathcal{L}$ is related to the Scott-topology on $\mathcal{A}(\mathcal{L})$, and next we show how Π-continuous functions between concepts spaces are related to Scott-continuous functions.

Definition 3.1.6 *Let $\mathcal{K}$ and $\mathcal{L}$ be concept spaces, and let $f: \mathcal{K} \to \mathcal{L}$ be a Π-continuous function. We define $\mathcal{A}(f): \mathcal{A}(\mathcal{K}) \to \mathcal{A}(\mathcal{L})$ so that*

$$\mathcal{A}(f)(X) = \bigcup\{\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\} \mid F \text{ is a finite subset of } X\}$$

for each $X \in \mathcal{A}(\mathcal{K})$. □

To see that $\mathcal{A}(f)$ actually is a function into $\mathcal{A}(\mathcal{L})$ (i.e., $\mathcal{A}(f)(X) \in \mathcal{A}(\mathcal{L})$ for each $X \in \mathcal{A}(\mathcal{K})$), first note that $\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\}$ is in $\mathcal{A}(\mathcal{L})$ because closure systems are closed under arbitrary intersections. Now, if F, G are finite and $F \subseteq G$, then

$$\{f(K) \mid F \subseteq K \in \mathcal{K}\} \supseteq \{f(K) \mid G \subseteq K \in \mathcal{K}\}$$

hence

$$\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\} \subseteq \bigcap\{f(K) \mid G \subseteq K \in \mathcal{K}\}.$$

Therefore,

$$\{\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\} \mid F \text{ is a finite subset of } X\}$$

is a directed family of elements of $\mathcal{A}(\mathcal{L})$, so their union is an element of $\mathcal{A}(\mathcal{L})$ by Proposition 2.1.5.

The following theorem relates Π-continuous functions and Scott-continuous functions. The theorem can be proven as a special case of Proposition II-3.9 in [20], but we give a full proof for completeness.

Theorem 3.1.7 *For every Π-continuous $f: \mathcal{K} \to \mathcal{L}$, $\mathcal{A}(f): \mathcal{A}(\mathcal{K}) \to \mathcal{A}(\mathcal{L})$ is Scott-continuous and $\mathcal{A}(f)(L) = f(L)$ for all $L \in \mathcal{L}$. Furthermore, $\mathcal{A}(f)$ is the supremum (ordered pointwise, i.e., $g \leq h \iff g(X) \subseteq h(X)$ for all X) of all such Scott-continuous extensions of f.*

Proof: We first show that $\mathcal{A}(f)(K) = f(K)$ for all $K \in \mathcal{K}$. From the definition of $\mathcal{A}(f)$, $x \in \mathcal{A}(f)(K)$ if and only if $x \in \bigcap\{f(K') \mid F \subseteq K' \in \mathcal{K}\}$ for some finite subset F of K, and therefore $\mathcal{A}(f)(K) \subseteq f(K)$. To show that $\mathcal{A}(f)(K) \supseteq f(K)$, note that for any $y \in f(K)$, $\uparrow_{\mathcal{L}} \{y\}$ is Π-open in $\mathcal{L}$ hence $f^{-1}(\uparrow_{\mathcal{L}} \{y\})$ is Π-open in $\mathcal{K}$ because f is Π-continuous. Since $K \in f^{-1}(\uparrow_{\mathcal{L}} \{y\})$, there must be some finite $F \subseteq K$ such that $K \in \uparrow_{\mathcal{K}} F \subseteq f^{-1}(\uparrow_{\mathcal{L}} \{y\})$. Then for every $K' \in \uparrow_{\mathcal{K}} F$, $y \in f(K')$, hence $y \in \bigcap\{f(K') \mid F \subseteq K' \in \mathcal{K}\}$. Therefore $y \in \mathcal{A}(f)(K)$.

Next we show that $\mathcal{A}(f)$ is Scott-continuous. Let D be a directed subset of $\mathcal{A}(\mathcal{K})$, then

$$\begin{aligned}
\mathcal{A}(f)(\bigvee_{d \in D} d) &= \bigcup\{\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\} \mid F \text{ is a finite subset of } \bigvee_{d \in D} d\} \\
&= \bigcup\{\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\} \mid F \subseteq d \text{ is finite and } d \in D\} \\
&= \bigcup_{d \in D} \bigcup\{\bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\} \mid F \subseteq d \text{ is finite}\} \\
&= \bigcup_{d \in D} \mathcal{A}(f)(d) = \bigvee_{d \in D} \mathcal{A}(f)(d),
\end{aligned}$$

where the second equation holds because $C_{\mathcal{L}}(F)$ is compact in $\mathcal{A}(\mathcal{K})$, and the last equation holds because $\mathcal{A}(\mathcal{K})$ is an algebraic lattice. Therefore $\mathcal{A}(f)$ is Scott-continuous.

Finally, we show that $\mathcal{A}(f)$ is the supremum of all Scott-continuous extensions of f. Assume $g: \mathcal{A}(\mathcal{K}) \to \mathcal{A}(\mathcal{L})$ is Scott-continuous, and $g(X) \not\subseteq \mathcal{A}(f)(X)$ for some $X \in \mathcal{A}(\mathcal{K})$. Let $D = \{d \in \mathcal{A}(\mathcal{K}) \mid d \text{ is compact and } d \subseteq X\}$. Then clearly D is directed and $X = \bigvee D$, so $g(X) = \bigvee_{d \in D} g(d)$ because g is Scott-continuous, and $\bigvee_{d \in D} g(d) = \bigcup_{d \in D} g(d)$ by Proposition 2.1.5. Applying the same argument to $\mathcal{A}(f)$, it follows that $g(X) = \bigcup_{d \in D} g(d) \not\subseteq \bigcup_{d \in D} \mathcal{A}(f)(d) =$

$\mathcal{A}(f)(X)$, hence there is compact $d \in D$ such that $g(d) \not\subseteq \mathcal{A}(f)(d)$. Therefore, it suffices to show that if g is a Scott-continuous extension of f, then $g(X) \subseteq \mathcal{A}(f)(X)$ for every compact $X \in \mathcal{A}(\mathcal{K})$.

Let X be compact in $\mathcal{A}(\mathcal{K})$ and assume some Scott-continuous function $g: \mathcal{A}(\mathcal{K}) \to \mathcal{A}(\mathcal{L})$ exists such that $g(X) \not\subseteq \mathcal{A}(f)(X)$. Note that $\mathcal{A}(f)(X) = \bigcap\{f(K) \mid F \subseteq K \in \mathcal{K}\}$ for some finite F such that $X = C_{\mathcal{K}}(F)$. If $x \in g(X) \setminus \mathcal{A}(f)(X)$, then there is some $K \in \mathcal{K}$ such that $F \subseteq X \subseteq K$ and $x \notin f(K)$. Since g is Scott-continuous, $g(X) \subseteq g(K)$ hence $x \in g(K)$. But this shows that $g(K) \neq f(K)$. Thus, if g is a Scott-continuous extension of f, then $g(X) \subseteq \mathcal{A}(f)(X)$ for every compact $X \in \mathcal{A}(\mathcal{K})$. □

We can think of $\mathcal{A}(f): \mathcal{A}(\mathcal{K}) \to \mathcal{A}(\mathcal{L})$ as a mapping from partial information about some $K \in \mathcal{K}$ to partial information about $f(K) \in \mathcal{L}$. The fact that it is the supremum of all continuous extensions means that it is the "best" such mapping of partial information.

Finally, we mention that despite our notation, $\mathcal{A}$ is *not* a functor because although it preserves identities it does not preserve composition. Here we give a simple example where $\mathcal{A}(g) \circ \mathcal{A}(f) \neq \mathcal{A}(g \circ f)$. We will denote subsets of $\{0, 1, 2, 3, 4\}$ by writing X with a subscript listing the elements it contains (e.g., $X_{024} = \{0, 2, 4\}$). Let $\mathcal{J} = \{X_{01}, X_{02}\}$, $\mathcal{K} = \{X_{01}, X_{02}, X_{03}\}$, and $\mathcal{L} = \{X_{014}, X_{024}, X_{03}\}$. Define $f: \mathcal{J} \to \mathcal{K}$ and $g: \mathcal{K} \to \mathcal{L}$ so that $f(X_{01}) = X_{01}$, $f(X_{02}) = X_{02}$, $g(X_{01}) = X_{014}$, $g(X_{02}) = X_{024}$, and $g(X_{03}) = X_{03}$. Then $(\mathcal{A}(g) \circ \mathcal{A}(f))(X_0) = X_0$, but $\mathcal{A}(g \circ f)(X_0) = X_{04}$.

3.1.3 Texts and the Π-topology

In this subsection, we explore the relationship between the Π-topology and texts for concepts.

Given a finite sequence σ of elements of $\omega \cup \{\#\}$, we define *content*$(\sigma) = \{n \in \omega \mid \exists i : \sigma(i) = n\}$. Recall that given a text T, $T(n)$ is the $(n+1)$th element of T, and $T[n]$ is the initial segment of T of length n.

Given a text T for some $L \in \mathcal{L}$, we can let $D = \{X_0, X_1, X_2, \ldots\}$ represent the ascending chain of the closed sets produced by applying $C_{\mathcal{L}}$ to the set of natural numbers appearing in each initial finite segment of T. Since $L = \bigvee_{i \in \omega} X_i = \bigvee D$ and D is directed, in fact a chain, $f(L) = \mathcal{A}(f)(\bigvee D) = \bigvee \mathcal{A}(f)(D)$. Thus, we can produce a text for $f(L)$ by enumerating in parallel the elements of each $\mathcal{A}(f)(X_i)$, while we can obtain each X_i by seeing more and more of T. We summarize this observation in the following proposition.

Proposition 3.1.8 *Let $f: \mathcal{K} \to \mathcal{L}$ be a Π-continuous function. Then there exists a continuous function $g: \mathcal{T}(\mathcal{K}) \to \mathcal{T}(\mathcal{L})$ such that $f \circ \tau_{\mathcal{K}} = \tau_{\mathcal{L}} \circ g$.*

$$\begin{tikzcd}
\mathcal{T}(\mathcal{K}) \arrow[r, "g"] \arrow[d, "\tau_{\mathcal{K}}"'] & \mathcal{T}(\mathcal{L}) \arrow[d, "\tau_{\mathcal{L}}"] \\
\mathcal{K} \arrow[r, "f"] & \mathcal{L}
\end{tikzcd}$$

Proof: For $T \in \mathcal{T}(\mathcal{K})$, let $X_n = C_{\mathcal{K}}(\mathit{content}(T[n]))$. Define $g: \mathcal{T}(\mathcal{K}) \to \mathcal{T}(\mathcal{L})$ as follows:

$$g(T)(n) \quad = \quad \begin{cases} \min(\mathcal{A}(f)(X_n) \setminus F_n) & \text{if } \mathcal{A}(f)(X_n) \setminus F_n \neq \emptyset \\ \# & \text{otherwise} \end{cases}$$

where $F_n = \mathit{content}(g(T)[n])$.

Let $K = \tau_{\mathcal{K}}(T)$. Then $K = \bigvee_{n \in \omega} X_n$, so using the fact that $\mathcal{A}(f)$ is a Scott-continuous extension of f, we have

$$f(K) = \mathcal{A}(f)(\bigvee_{n\in\omega} X_n) = \bigvee_{n\in\omega} \mathcal{A}(f)(X_n) = \bigcup_{n\in\omega} \mathcal{A}(f)(X_n),$$

where the last equation follows from Proposition 2.1.5. Therefore, if $g(T)(n) \neq \#$, then $g(T)(n) \in \mathcal{A}(f)(X_n) \subseteq f(K)$.

If $x \in f(K)$, then there is some m such that $x \in \mathcal{A}(f)(X_n)$ for all $n \geq m$. So if $x \notin F_m$, then $x \in \mathcal{A}(f)(X_n) \setminus F_m$. There are at most x elements less than x in $f(K) \setminus F_m$ that $g(T)$ will enumerate, so eventually x will be the minimal element of $\mathcal{A}(f)(X_n) \setminus F_n$ for some $n \geq m$, hence x occurs in $g(T)$.

It follows that $g(T)$ is a text for $f(K)$. Therefore, $f \circ \tau_{\mathcal{K}} = \tau_{\mathcal{L}} \circ g$. It is clear that g is continuous. □

A function $f: X \to Y$ between topological spaces is *open* if and only if $f(U)$ is open in Y for every open $U \subseteq Y$. The next theorem shows how the Π-topology is determined by the space of texts.

Theorem 3.1.9 *The function $\tau_{\mathcal{L}}: \mathcal{T}(\mathcal{L}) \to \mathcal{L}$ is an open continuous surjective function with respect to the Π-topology on $\mathcal{L}$. Thus, the Π-topology on $\mathcal{L}$ is the quotient topology on $\mathcal{L}$ with respect to $\tau_{\mathcal{L}}$.*

Proof: To see that $\tau_{\mathcal{L}}$ is open, note that $\tau_{\mathcal{L}}(\uparrow\sigma) = \uparrow_{\mathcal{L}} \mathit{content}(\sigma)$ for any finite sequence σ. If $U \subseteq \mathcal{T}(\mathcal{L})$ is open and non-empty, then $U = \bigcup_{i\in\omega} \uparrow\sigma_i$ for some family $\{\sigma_i\}_{i\in\omega}$ of finite sequences, hence $\tau_{\mathcal{L}}(U) = \tau_{\mathcal{L}}(\bigcup_{i\in\omega} \uparrow\sigma) = \bigcup_{i\in\omega} \tau_{\mathcal{L}}(\uparrow\sigma)$ is a Π-open subset of $\mathcal{L}$.

To see that $\tau_{\mathcal{L}}$ is contiuous, it suffices to check that $\tau_{\mathcal{L}}^{-1}(\uparrow_{\mathcal{L}} F)$ is open in $\mathcal{T}(\mathcal{L})$ for every finite $F \subseteq \omega$. Since $F \subseteq \tau_{\mathcal{L}}(T)$ if and only if there is finite $\sigma \prec T$ such that $F \subseteq \mathit{content}(\sigma)$, $\tau_{\mathcal{L}}^{-1}(\uparrow_{\mathcal{L}} F) = \bigcup\{\uparrow\sigma \mid F \subseteq \mathit{content}(\sigma)\}$, which is clearly open in $\mathcal{T}(\mathcal{L})$.

Now, if $U \subseteq \mathcal{L}$ is Π-open, then $\tau_{\mathcal{L}}^{-1}(U)$ is open in $\mathcal{T}(\mathcal{L})$ because $\tau_{\mathcal{L}}$ is continuous. If $U \subseteq \mathcal{L}$ and $\tau_{\mathcal{L}}^{-1}(U)$ is open, then $\tau_{\mathcal{L}}(\tau_{\mathcal{L}}^{-1}(U)) = U$ is Π-open in $\mathcal{L}$ because $\tau_{\mathcal{L}}$ is open. Therefore, $U \subseteq \mathcal{L}$ is Π-open if and only if $\tau_{\mathcal{L}}^{-1}(U)$ is open in $\mathcal{T}(\mathcal{L})$, so the Π-topology on $\mathcal{L}$ is the quotient topology with respect to $\tau_{\mathcal{L}}$. □

Corollary 3.1.10 *A function $f: \mathcal{K} \to \mathcal{L}$ between concept spaces is Π-continuous if and only if there is a continuous function $g: \mathcal{T}(\mathcal{K}) \to \mathcal{T}(\mathcal{L})$ such that $f \circ \tau_{\mathcal{K}} = \tau_{\mathcal{L}} \circ g$.*

Proof: We have already shown the "only if" part. So let $f: \mathcal{K} \to \mathcal{L}$ be any function and assume $g: \mathcal{T}(\mathcal{K}) \to \mathcal{T}(\mathcal{L})$ is a continuous function satisfying $f \circ \tau_{\mathcal{K}} = \tau_{\mathcal{L}} \circ g$. Let U be any Π-open subset of $\mathcal{L}$. Then $\tau_{\mathcal{K}}^{-1}(f^{-1}(U)) = g^{-1}(\tau_{\mathcal{L}}^{-1}(U))$ is an open subset of $\mathcal{T}(\mathcal{K})$ because g and $\tau_{\mathcal{L}}$ are continuous. Thus, $f^{-1}(U) =$

$\tau_{\mathcal{K}}(\tau_{\mathcal{K}}^{-1}(f^{-1}(U)))$ is Π-open in $\mathcal{K}$ because $\tau_{\mathcal{K}}$ is an open function. Therefore, f is Π-continuous. $\square$

A topological space is *zero-dimensional* if and only if it has a basis of clopen sets. If a space is zero-dimensional and countably based, then it can easily be shown that the space has a countable basis of clopen sets.

The next theorem shows that the space of texts forms an *admissible representation* of $\mathcal{L}$ with respect to the Π-topology. Admissible representations are important in the field of computable analysis (see Weihrauch [59] and Schröder [48]), and will play a major role in Chapter 4.

Theorem 3.1.11 *Let Z be a zero-dimensional countably based T_0-space, and $f\colon Z \to \mathcal{L}$ a continuous function. Then there exists a continuous function $g\colon Z \to \mathcal{T}(\mathcal{L})$ such that $f = \tau_{\mathcal{L}} \circ g$.*

$$\begin{tikzcd} Z \arrow[r, "g"] \arrow[dr, "f"'] & \mathcal{T}(\mathcal{L}) \arrow[d, "\tau_{\mathcal{L}}"] \\ & \mathcal{L} \end{tikzcd}$$

Proof: Let $\{B_i\}_{i\in\omega}$ be a sequence of clopen sets forming a basis for Z. Given $x \in Z$, define the sequence $\{B_j^x\}_{j\in\omega}$ so that for each $j \in \omega$, $B_j^x = B_j$ if $x \in B_j$, and $B_j^x = Z \setminus B_j$ if $x \notin B_j$. Note that B_j^x is open for all $j \in \omega$ because B_j is clopen. Let $\{F_i\}_{i\in\omega}$ be an enumeration of all finite subsets of natural numbers in such a way that every subset occurs infinitely often in the sequence. Let S be the set of all non-empty finite sequences of natural numbers. Given a finite set F of natural numbers, we denote by $seq(F)$ the lexicographically smallest sequence that contains exactly the elements of F if F is non-empty, and define $seq(F) = \#$ if F is the empty set.

We first define a function $p\colon Z \times \omega \to S \cup \{\#\}$. For $x \in Z$ and $i \in \omega$, define

$$p(x,i) = \begin{cases} seq(F_i) & \text{if } f(\bigcap_{j\le i} B_j^x) \subseteq \uparrow_{\mathcal{L}} F_i; \\ \# & \text{otherwise.} \end{cases}$$

We define $g(x) = p(x,0) \diamond p(x,1) \diamond p(x,2) \diamond \cdots$, where $\diamond$ denotes concatenation of sequences.

Choose $x \in Z$ and let $i \in \omega$ be such that $F_i \subseteq f(x)$. Since f is continuous, $f^{-1}(\uparrow_{\mathcal{L}} F_i)$ is an open subset of Z, and since $\{B_i\}_{i\in\omega}$ forms a basis for Z there is some $m \in \omega$ such that $x \in B_m = B_m^x \subseteq f^{-1}(\uparrow_{\mathcal{L}} F_i)$. Since the set F_i occurs infinitely many times in our enumeration of the finite subsets of natural numbers, there exists $n \ge m$ such that $F_n = F_i$ and it follows that $p(x,n) = seq(F_i)$. Furthermore, for each i, if $p(x,i) = seq(F_i)$, then $F_i \subseteq f(x)$ by our definition of p. Thus, it is easily seen that $g(x)$ is a text for $f(x)$, and therefore $f = \tau_{\mathcal{L}} \circ g$.

Finally, we show that g is continuous. Let σ be a finite sequence of elements of $(\omega \cup \{\#\})$. If $g^{-1}(\uparrow \sigma)$ is empty then it is open by definition. Otherwise, let $x \in g^{-1}(\uparrow \sigma)$ be given. Choose n large enough that σ is an initial prefix of $p(x,0) \diamond p(x,1) \diamond \cdots \diamond p(x,n)$. Then $U = \bigcap_{i=0}^{n} B_i^x$ is an open subset of Z containing x. Furthermore, given $y \in U$, we must have $B_i^y = B_i^x$ for $0 \le i \le n$, which implies $p(y,i) = p(x,i)$ for $0 \le i \le n$. Thus, $y \in g^{-1}(\uparrow \sigma)$, and since y

was arbitrary it follows that $U \subseteq g^{-1}(\uparrow\sigma)$. Therefore, $g^{-1}(\uparrow\sigma)$ is equal to the union of open sets hence g is continuous. □

3.1.4 Products

Our construction of products is related to Dana Scott's construction of products of information systems [51].

Definition 3.1.12 *Let $I \subseteq \omega$ and let $\mathcal{L}_i$ be a concept space for each $i \in I$. Define*

$$\prod_{i\in I} \mathcal{L}_i = \{\langle L_i\rangle_{i\in I} \mid L_i \in \mathcal{L}_i\},$$

where $\langle L_i\rangle_{i\in I} = \{\langle x, i\rangle \mid i \in I \text{ and } x \in L_i\}$. For each $j \in I$, define $\pi_j\colon \prod_{i\in I} \mathcal{L}_i \to \mathcal{L}_j$ so that $\pi_j(\langle L_i\rangle_{i\in I}) = L_j$. □

If I only contains finitely many elements, for example $I = \{1, 2, \ldots, n\}$, then we write $\prod_{i\in I} \mathcal{L}_i$ as $\mathcal{L}_1 \times \mathcal{L}_2 \times \cdots \times \mathcal{L}_n$ and $\langle L_i\rangle_{i\in I}$ as $\langle L_1, L_2, \ldots, L_n\rangle$.

Products can be used to model multiple learning problems in parallel. For example, we can combine the problems of identifying $\mathcal{J}$ and $\mathcal{K}$ into the single problem of identifying $\mathcal{J} \times \mathcal{K}$.

Lemma 3.1.13 *For each $j \in I$, $\pi_j\colon \prod_{i\in I} \mathcal{L}_i \to \mathcal{L}_j$ is a Π-continuous open function.*

Proof: For any finite $F \subseteq \omega$, $\pi_j^{-1}(\uparrow_{\mathcal{L}_j} F) = \{\langle L_i\rangle_{i\in I} \mid F \subseteq L_j\}$, hence

$$\pi_j^{-1}(\uparrow_{\mathcal{L}_j} F) = \bigcap_{x\in F} \{\langle L_i\rangle_{i\in I} \mid \langle x, j\rangle \in \langle L_i\rangle_{i\in I}\}$$

is a Π-open subset of $\prod_{i\in I} \mathcal{L}_i$, so π_j is Π-continuous. Furthermore,

$$\pi_j(\uparrow_{(\prod_{i\in I} \mathcal{L}_i)} F) = \bigcap_{\langle x,j\rangle\in F} \uparrow_{\mathcal{L}_j} \{x\},$$

so π_j is an open function. □

The next theorem shows that our definition of products of concept spaces is the correct definition, from the perspective of category theory [35].

Theorem 3.1.14 *Given any concept space $\mathcal{J}$ and Π-continuous $f_j\colon \mathcal{J} \to \mathcal{L}_j$ for each $j \in I$, there exists a unique Π-continuous function $\langle f_i\rangle_{i\in I}\colon \mathcal{J} \to \prod_{i\in I} \mathcal{L}_i$ such that $\pi_j \circ \langle f_i\rangle_{i\in I} = f_j$ for each $j \in I$.*

Proof: Clearly, $\langle f_i\rangle_{i\in I}$ is uniquely determined to be $\langle f_i\rangle_{i\in I}(J) = \langle f_i(J)\rangle_{i\in I}$ for each $J \in \mathcal{J}$.

For any non-empty finite subset F of ω, we can partition F into finitely many sets $F_{i_0}, F_{i_1}, \ldots, F_{i_n}$ so that $F = \bigcup_{k=0}^{n} \{\langle x, i_k\rangle \mid x \in F_{i_k}\}$. Thus,

$$(\langle f_i\rangle_{i\in I})^{-1}(\uparrow_{(\prod_{i\in I} \mathcal{L}_i)} F) = f_{i_0}^{-1}(F_{i_0}) \cap f_{i_1}^{-1}(F_{i_1}) \cap \cdots \cap f_{i_n}^{-1}(F_{i_n})$$

is Π-open in $\mathcal{J}$ because each f_j is Π-continuous. □

Definition 3.1.15 *Given* Π*-continuous function* $f_i\colon \mathcal{J}_i \to \mathcal{K}_i$ *for* $i \in I$*, define*

$$\prod_{i\in I} f_i\colon \prod_{i\in I} \mathcal{J}_i \to \prod_{i\in I} \mathcal{K}_i$$

by $(\prod_{i\in I} f_i)(\langle J_i\rangle_{i\in I}) = \langle f_i(J_i)\rangle_{i\in I}$. □

If $I = \{1, 2, \ldots, n\}$ is finite, then we write $\prod_{i\in I} f_i$ as $f_1 \times f_2 \times \cdots \times f_n$.

Lemma 3.1.16 *If* $f_i\colon \mathcal{J}_i \to \mathcal{K}_i$ *is* Π*-continuous for each* $i \in I$*, then*

$$\prod_{i\in I} f_i\colon \prod_{i\in I} \mathcal{J}_i \to \prod_{i\in I} \mathcal{K}_i$$

is Π*-continuous.* □

3.1.5 (Some) Exponents

Next we define a concept space of all Π-continuous functions from $\mathcal{J}$ to $\mathcal{K}$, which we denote by $\mathcal{K}^{\mathcal{J}}$. Since the category of concept spaces and Π-continuous maps is not cartesian closed, we can not expect $\mathcal{K}^{\mathcal{J}}$ to always be an exponential object [35]. However, we will give a sufficient condition below for $\mathcal{K}^{\mathcal{J}}$ to be an exponential object. Just like products, the constructions in this subsection are also closely related to Dana Scott's information systems [51].

Let $\mathcal{FIN}$ be the collection of all finite subsets of ω. For any concept space $\mathcal{J}$, define

$$\mathcal{FIN}_{\mathcal{J}} = \{F \in \mathcal{FIN} \mid F \neq \emptyset \;\&\; (\uparrow_{\mathcal{J}} F) \neq \emptyset\}$$

if $\mathcal{J}$ does not contain the empty concept, and

$$\mathcal{FIN}_{\mathcal{J}} = \{F \in \mathcal{FIN} \mid (\uparrow_{\mathcal{J}} F) \neq \emptyset\}$$

otherwise. Let $[\cdot,\cdot]\colon \mathcal{FIN} \times \mathcal{FIN} \to \omega$ be a bijection.

Definition 3.1.17 *For any* Π*-continuous function* $f\colon \mathcal{J} \to \mathcal{K}$*, define*

$$\ulcorner f \urcorner = \{\, [F, G] \mid F \in \mathcal{FIN}_{\mathcal{J}} \;\&\; G \in \mathcal{FIN} \;\&\; G \subseteq \mathcal{A}(f)(C_{\mathcal{J}}(F))\}.$$

We define $\mathcal{K}^{\mathcal{J}} = \{\ulcorner f \urcorner \mid f\colon \mathcal{J} \to \mathcal{K}$ *is* Π*-continuous* $\}$. □

Intuitively, the concept $\ulcorner f \urcorner$ encodes the function f in terms of numbers $[F, G]$ which indicate that if F is a subset of the input then G is a subset of the output.

For simplicity, we will assume that $\mathcal{J}$ does not contain the empty concept in the following proofs. All results can easily be shown to hold in their current form if minor modifications are made to the proofs to handle the special case.

Definition 3.1.18 *Define* $\varepsilon_{\mathcal{K}^{\mathcal{J}}}\colon \mathcal{K}^{\mathcal{J}} \times \mathcal{J} \to \mathcal{K}$ *so that for any* $\langle \ulcorner f \urcorner, J\rangle \in \mathcal{K}^{\mathcal{J}} \times \mathcal{J}$,

$$\varepsilon_{\mathcal{K}^{\mathcal{J}}}(\langle \ulcorner f \urcorner, J\rangle) = \bigcup\{G \mid [F, G] \in \ulcorner f \urcorner \text{ and } F \subseteq J\}.$$

We will often abbreviate $\varepsilon_{\mathcal{K}^{\mathcal{J}}}$ *to* ε *when there will be no confusion.* □

Lemma 3.1.19 *$\varepsilon\colon \mathcal{K}^{\mathcal{J}} \times \mathcal{J} \to \mathcal{K}$ is Π-continuous and $\varepsilon(\langle \ulcorner f \urcorner, J\rangle) = f(J)$ for all Π-continuous $f\colon \mathcal{J} \to \mathcal{K}$ and $J \in \mathcal{K}$.*

Proof: To see that $\varepsilon(\langle \ulcorner f \urcorner, J\rangle) = f(J)$ for all Π-continuous $f\colon \mathcal{J} \to \mathcal{K}$ and $J \in \mathcal{K}$, first note that $\varepsilon(\langle \ulcorner f \urcorner, J\rangle) \subseteq f(J)$ because $\mathcal{A}(f)$ is monotonic and $\mathcal{A}(f)(J) = f(J)$. Let G be any finite subset of $f(J)$, and let $\{F_i\}_{i \in I}$ be the collection of all non-empty finite subsets of J (where I is some index set). Then $f(J) = \mathcal{A}(f)(J) = \mathcal{A}(f)(\bigvee_{i \in I} C_{\mathcal{J}}(F_i)) = \bigvee_{i \in I} \mathcal{A}(f)(C_{\mathcal{J}}(F_i))$. Since $C_{\mathcal{K}}(G) \subseteq \bigvee_{i \in I} \mathcal{A}(f)(C_{\mathcal{J}}(F_i))$ is compact and $\{\mathcal{A}(f)(C_{\mathcal{J}}(F_i))\}_{i \in I}$ is directed, there must be $i \in I$ such that $C_{\mathcal{K}}(G) \subseteq \mathcal{A}(f)(C_{\mathcal{J}}(F_i))$. This implies that $[F_i, G] \in \ulcorner f \urcorner$, and so $G \subseteq \varepsilon(\langle \ulcorner f \urcorner, J\rangle)$. Therefore, $\varepsilon(\langle \ulcorner f \urcorner, J\rangle) = f(J)$.

We next show that ε is Π-continuous. Let $G \subseteq \omega$ be finite, and let $\langle \ulcorner f \urcorner, J\rangle \in \varepsilon^{-1}(\uparrow_{\mathcal{K}} G)$. Then there must be a finite sequence $\{\, [F_i, G_i] \,\}_{1 \leq i \leq n}$ in $\ulcorner f \urcorner$ such that $G \subseteq \bigcup_{i=1}^{n} G_i$ and $\bigcup_{i=1}^{n} F_i \subseteq J$. Let $F = \bigcup_{i=1}^{n} F_i$. Since $\mathcal{A}(f)$ is monotonic, $G_i \subseteq \mathcal{A}(f)(C_{\mathcal{J}}(F))$ for $1 \leq i \leq n$, which implies that $G \subseteq \mathcal{A}(f)(C_{\mathcal{J}}(F))$. Therefore, $[F, G] \in \ulcorner f \urcorner$. Then clearly for every $\langle \ulcorner f' \urcorner, J'\rangle \in \mathcal{K}^{\mathcal{J}} \times \mathcal{J}$ such that $\ulcorner f' \urcorner$ contains $[F, G]$ and J' contains F, $G \subseteq \varepsilon(\langle \ulcorner f' \urcorner, J'\rangle)$. Thus,

$$\langle \ulcorner f \urcorner, J\rangle \in \uparrow_{\mathcal{K}^{\mathcal{J}} \times \mathcal{J}} \langle \{\, [F, G] \,\}, F\rangle \subseteq \varepsilon^{-1}(\uparrow_{\mathcal{K}} G),$$

which implies that ε is Π-continuous. □

Combining the above lemma with Corollary 3.1.10, if we are given a text for some $\ulcorner f \urcorner \in \mathcal{K}^{\mathcal{J}}$ and a text for some $J \in \mathcal{J}$, then we can generate a text for $f(J) \in \mathcal{K}$.

Lemma 3.1.20 *For any Π-continuous function $f\colon \mathcal{L} \times \mathcal{J} \to \mathcal{K}$, the function $f_L\colon \mathcal{J} \to \mathcal{K}$, defined as $f_L(J) = f(\langle L, J\rangle)$ for all $J \in \mathcal{J}$, is Π-continuous.*

Proof: Let U be a Π-open subset of $\mathcal{K}$, and let $J \in f_L^{-1}(U)$. Since f is Π-continuous, $f^{-1}(U)$ is Π-open, and since $\langle L, J\rangle \in f^{-1}(U)$, there is a finite set $\langle F, G\rangle \subseteq \omega$ such that $\langle L, J\rangle \in \uparrow_{\mathcal{L} \times \mathcal{J}} \langle F, G\rangle \subseteq f^{-1}(U)$. Therefore, $J \in \uparrow_{\mathcal{J}} G \subseteq f_L^{-1}(U)$, and it follows that f_L is Π-continuous. □

Definition 3.1.21 *For any Π-continuous $f\colon \mathcal{L} \times \mathcal{J} \to \mathcal{K}$, define $\lambda f\colon \mathcal{L} \to \mathcal{K}^{\mathcal{J}}$ so that $\lambda f(L) = \ulcorner f_L \urcorner$ for each $L \in \mathcal{L}$.* □

The next proposition follows immediately from the definition and Lemma 3.1.20.

Proposition 3.1.22 *For any Π-continuous function $f\colon \mathcal{L} \times \mathcal{J} \to \mathcal{K}$, $\lambda f\colon \mathcal{L} \to \mathcal{K}^{\mathcal{J}}$ is the unique function from $\mathcal{L}$ to $\mathcal{K}^{\mathcal{J}}$ that satisfies $f(\langle L, J\rangle) = \varepsilon(\langle \lambda f(L), J\rangle)$ for all $\langle L, J\rangle \in \mathcal{L} \times \mathcal{J}$.*

$$\begin{tikzcd}
\mathcal{K}^{\mathcal{J}} & \mathcal{K}^{\mathcal{J}} \times \mathcal{J} \arrow[r, "\varepsilon"] & \mathcal{K} \\
\mathcal{L} \arrow[u, dashed, "\lambda f"] & \mathcal{L} \times \mathcal{J} \arrow[u, "\lambda f \times 1_{\mathcal{J}}"] \arrow[ur, "f"'] &
\end{tikzcd}$$

□

Definition 3.1.23 *$\mathcal{K}^{\mathcal{J}}$ is the* exponential object *of $\mathcal{J}$ and $\mathcal{K}$ if and only if for any concept space $\mathcal{L}$ and any Π-continuous function $f\colon \mathcal{L}\times\mathcal{J}\to\mathcal{K}$, $\lambda f\colon \mathcal{L}\to\mathcal{K}^{\mathcal{J}}$ is Π-continuous.* □

In general, λf is *not* Π-continuous, but we give a sufficient condition for it being continuous below. Intuitively, the problem is that $\ulcorner f_L \urcorner$ sometimes contains more information about f_L than can be extracted from a text for $L\in\mathcal{L}$.

Definition 3.1.24 *Let X be a topological space. A subset K of X is* topologically compact *if and only if for every family $\{U_i\}_{i\in I}$ of open subsets of X such that $K\subseteq\bigcup_{i\in I}U_i$, there is some finite $F\subseteq I$ such that $K\subseteq\bigcup_{i\in F}U_i$.* □

Theorem 3.1.25 *Let $\mathcal{J}$ be a concept space such that $\uparrow_{\mathcal{J}} F$ is topologically compact for every non-empty finite $F\subseteq\omega$. Then for any concept space $\mathcal{L}$ and any Π-continuous function $f\colon\mathcal{L}\times\mathcal{J}\to\mathcal{K}$, $\lambda f\colon\mathcal{L}\to\mathcal{K}^{\mathcal{J}}$ is Π-continuous.*

Proof: Let $S=\{\,[F_i,G_i]\,\}_{1\le i\le n}$ be a finite set representing pairs of finite subsets of ω. It suffices to show that $\lambda f^{-1}(\uparrow_{\mathcal{K}^{\mathcal{J}}} S)$ is a Π-open subset of $\mathcal{L}$. We assume that $F_i\neq\emptyset$ and $\uparrow_{\mathcal{J}} F_i$ is non-empty for $1\le i\le n$, since otherwise $\uparrow_{\mathcal{K}^{\mathcal{J}}} S$ would be emtpy and the claim would follow trivially. Then $L\in\lambda f^{-1}(\uparrow_{\mathcal{K}^{\mathcal{J}}} S)$ if and only if $G_i\subseteq\mathcal{A}(f_L)(C_{\mathcal{J}}(F_i))$ if and only if $G_i\subseteq f(\langle L,J\rangle)$ for all $J\in\uparrow_{\mathcal{J}} F_i$ (for all $1\le i\le n$).

Choose any $L\in\lambda f^{-1}(\uparrow_{\mathcal{K}^{\mathcal{J}}} S)$. Since f is Π-continuous, $U_i=f^{-1}(\uparrow_{\mathcal{K}} G_i)$ is Π-open for each $1\le i\le n$. For each $J\in\uparrow_{\mathcal{J}} F_i$, let H_i^J and F_i^J be finite subsets of ω such that $\langle L,J\rangle\in\uparrow_{\mathcal{L}\times J}\langle H_i^J,F_i^J\rangle\subseteq U_i$. Such finite sets clearly exists by our assumption on L. Since $\uparrow_{\mathcal{J}} F_i\subseteq\bigcup_{J\in\mathcal{J}}\uparrow_{\mathcal{J}} F_i^J$, from our assumption on $\mathcal{J}$ there is some finite $X_i\subseteq\mathcal{J}$ such that $\uparrow_{\mathcal{J}} F_i\subseteq\bigcup_{J\in X_i}\uparrow_{\mathcal{J}} F_i^J$. Therefore, $H_i=\bigcup\{H_i^J\mid J\in X_i\}$ and $H=\bigcup_{i=1}^n H_i$ are finite subsets of ω.

Let $L'\in\uparrow_{\mathcal{L}} H$. For any i ($1\le i\le n$) and every $J'\in\uparrow_{\mathcal{J}} F_i$, there is some $J\in X_i$ such that $F_i^J\subseteq J'$. Therefore, $\langle L',J'\rangle\in\uparrow_{\mathcal{L}\times J}\langle H_i^J,F_i^J\rangle\subseteq U_i$, which means that $G_i\subseteq f(\langle L',J'\rangle)$. Therefore, $L\in\uparrow_{\mathcal{L}} H\subseteq\lambda f^{-1}(\uparrow_{\mathcal{K}^{\mathcal{J}}} S)$, and it follows that λf is Π-continuous. □

If $\mathcal{J}$ contains the empty concept, then $\uparrow_{\mathcal{J}}\emptyset$ would trivially be topologically compact, so the above theorem still holds in its current form. Also note that the property in the premise of Theorem 3.1.25 is not topologically invariant (i.e., there may be homeomorphic spaces where one has the property and the other does not).

We will be careful to mention $\mathcal{K}^{\mathcal{J}}$ only when it actually is an exponential object. In the next section, we will show that the property of M-finite thickness is useful in proving that $\mathcal{K}^{\mathcal{J}}$ is an exponential object.

Finally, we mention that our definition of $\mathcal{K}^{\mathcal{J}}$ is certainly not the best one possible, in the sense that there are concept spaces that have exponential objects but must be constructed in a different way. Our choice of the definition of $\mathcal{K}^{\mathcal{J}}$ is that it is relatively intuitive and yet still useful in many situations.

3.1.6 The informant topology

Although we will be primarily concerned with the Π-topology on concept spaces, in this subsection we will briefly introduce a topology that is more suitable for identification from positive and negative data.

Definition 3.1.26 *The* informant topology *or* positive and negative information topology *on $\mathcal{L}$ is generated by sets of the form*

$$\{\uparrow_{\mathcal{L}} F \mid F \subseteq \omega \textit{ is finite }\} \cup \{\mathcal{L} \setminus \uparrow_{\mathcal{L}} F \mid F \subseteq \omega \textit{ is finite }\},$$

where $F \subseteq \omega$ is finite. □

It is clear from the above definition that the informant topology on a concept space is countably based, zero-dimensional, and satisfies the T_0 separation axiom. The converse is true up to homeomorphism.

Proposition 3.1.27 *Every zero-dimensional countably based T_0-space is homeomorphic to some concept space with the informant topology.*

Proof: Let X be a zero-dimensional countably based T_0-space and let $\{B_i\}_{i\in\omega}$ be a countable basis for X consisting of clopen sets. For $x \in X$, define $\zeta(x) = \{\langle i,1\rangle \mid x \in B_i\} \cup \{\langle i,0\rangle \mid x \notin B_i\}$, where $\langle\cdot,\cdot\rangle: \omega \times \{0,1\} \to \omega$ is a bijection. Define $\mathcal{L}_X = \{\zeta(x) \mid x \in X\}$. Then $\mathcal{L}_X$ is a concept space, and $\zeta: X \to \mathcal{L}_X$ is easily seen to be a homeomorphism with respect to the informant topology on $\mathcal{L}_X$. □

Note that if $F \subset \omega$ is finite, then $\mathcal{L}_X \setminus \uparrow_{\mathcal{L}_X} F = \uparrow_{\mathcal{L}_X} \overline{F}$ if we define $\overline{F} = \{\langle i, 1-j\rangle \mid \langle i,j\rangle \in F\}$, which means that the Π-topology and informant topology agree on $\mathcal{L}_X$. So the above proof shows that every zero-dimensional space is homeomorphic to a concept space in which the Π-topology and informant topology agree.

Theorem 3.1.28 *$\langle \iota_{\mathcal{L}}, \mathcal{I}(\mathcal{L})\rangle$ is an admissible representation of $\mathcal{L}$ with the informant topology.*

Proof: Define $\zeta(L) = \{\langle i,1\rangle \mid i \in L\} \cup \{\langle i,0\rangle \mid i \notin L\}$ for $L \in \mathcal{L}$, and let $\mathcal{Z} = \{\zeta(L) \mid L \in \mathcal{L}\}$. Then $\zeta: \mathcal{L} \to \mathcal{Z}$ is a homeomorphism from $\mathcal{L}$ (with the informant topology) to $\mathcal{Z}$ (with the Π-topology), and informants for $\mathcal{L}$ are the same as texts for $\mathcal{Z}$. The admissibility of $\langle \tau_{\mathcal{Z}}, \mathcal{T}(\mathcal{Z})\rangle$ can then be used to prove the admissibility of $\langle \iota_{\mathcal{L}}, \mathcal{I}(\mathcal{L})\rangle$. □

As will be shown in Chapter 4, the above theorem implies that the informant topology on $\mathcal{L}$ is the quotient topology with respect to $\iota_{\mathcal{L}}$. Furthermore, a function $f: \mathcal{K} \to \mathcal{L}$ is continuous with respect to the informant topologies on $\mathcal{K}$ and $\mathcal{L}$ if and only if there is a continuous function $g: \mathcal{I}(\mathcal{K}) \to \mathcal{I}(\mathcal{L})$ such that $f \circ \iota_{\mathcal{K}} = \iota_{\mathcal{L}} \circ g$.

3.2 Topologically invariant properties

Since every concept space with the informant topology is homeomorphic to a zero-dimensional concept space with the Π-topology, we obtain more general results by investigating the Π-topology. However, we must be careful when we exchange results between the two paradigms.

For example, a necessary and sufficient criterion for a concept space to be identifiable in the limit from positive data is that every concept has a finite tell-tale. However, this criterion is certainly *not* true if we replace "positive data" by "positive and negative data", because it is well known that every countable concept space is identifiable in the limit from positive and negative data. However, we will show in this section that the notion of a finite tell-tale is related to a purely topological property known as the T_D separation axiom. We can then show that there is a uniform criterion for identifiability in the limit: a concept space is identifiable in the limit from positive data (from positive and negative data) if and only if it is a T_D-space with respect to the Π-topology (the informant topology). A more interesting result is the following: a concept space is identifiable in the limit with a mind-change bound from positive data (from positive and negative data) if and only if it is a scattered space with respect to the Π-topology (the informant topology).

It will become clear from the results of Chapter 4 that the "T_D" and "scattered" criteria apply not only to identification from texts or informants, but for any paradigm in which the representation $\langle \mathcal{R}, \rho \rangle$ of a concept space is admissible with respect to some countably based topology on the concept space. Therefore, by formulating criteria for identifiability in topological terms, we obtain results that apply to all paradigms that have "admissible" information presentations.

In this section, we will investigate topological properties of concept spaces that are relevant to the identification in the limit paradigm. By topological property, we mean a property that is *topologically invariant*, i.e., those properties P such that if P holds for a topological space X, then P also holds for all spaces homeomorphic to X. For comparison, we will also give some examples of well known properties that are *not* topologically invariant. Three invariant properties and one non-invariant property that we will investigate are shown in the figure.

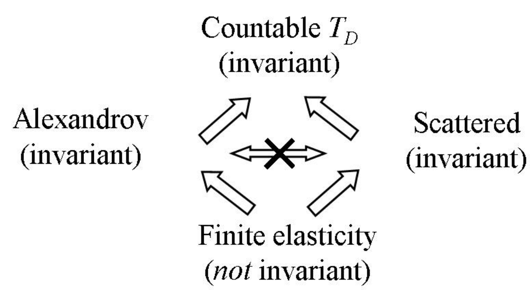


Figure 3.2: The relationship between three topologically invariant properties of concept spaces and one non-topological property.

We will always assume the Π-topology on concept spaces in this section. We will actually be investigating two different forms of topological invariance: purely topological properties and those invariant with respect to concept spaces with the Π-topology. For example, finite tell-tales are not a purely topological property, because the definition does not even make sense for most topological spaces. However, we will see that if a concept space with the Π-topology has finite tell-tales for all concepts, then any other homeomorphic concept space with the Π-topology will also have finite tell-tales for all concepts. Therefore, the notion of a finite tell-tale is invariant with respect to concept spaces with the Π-topology. On the other hand, the T_D separation axiom is a purely topological property that makes sense for any topological space, and in the case of a concept space with the Π-topology, the T_D separation axiom is equivalent to every concept having a finite tell-tale.

3.2.1 Scattered spaces

Scattered concept spaces are *defined* topologically, and therefore the property is easily seen to be a pure topologically invariant property. The notion was introduced to the learning community by Luo and Schulte [34] as a means of characterizing mind-change complexity (see Section 2.2.2).

Definition 3.2.1 *A concept $L \in X \subseteq \mathcal{L}$ is an* isolated point *of X if and only if there is a Π-open subset U of $\mathcal{L}$ such that $X \cap U = \{L\}$. If $L \in X \subseteq \mathcal{L}$ and L is not an isolated point of X, then L is an* accumulation point *of X.* □

Definition 3.2.2 (see [29, 34]) *Let $\mathcal{L}$ be a concept space. For each ordinal α, the α-th derived set of $\mathcal{L}$, denoted $\mathcal{L}^{(\alpha)}$, is defined inductively as follows:*

1. *$\mathcal{L}^{(0)}$ is defined to be $\mathcal{L}$.*
2. *$\mathcal{L}^{(\alpha+1)}$ is defined to be the set of all accumulation points of $\mathcal{L}^{(\alpha)}$.*
3. *If α is a limit ordinal, then $\mathcal{L}^{(\alpha)}$ is defined to be $\bigcap_{\beta<\alpha} \mathcal{L}^{(\beta)}$.*

The accumulation order *of a concept $L \in \mathcal{L}$, denoted* $\mathrm{acc}_{\mathcal{L}}(L)$*, is defined as the maximal ordinal α such that $L \in \mathcal{L}^{(\alpha)}$ (if the maximum exists). The* accumulation order *of a concept space $\mathcal{L}$, denoted* $\mathrm{acc}(\mathcal{L})$*, is the least ordinal α such that $\mathcal{L}^{(\alpha)} = \mathcal{L}^{(\alpha+1)}$. If $acc(\mathcal{L}) = \alpha$ and $\mathcal{L}^{(\alpha)}$ is empty, then $\mathcal{L}$ is said to be* scattered. □

Since all concept spaces are countably-based, it can be shown that $\mathrm{acc}(\mathcal{L})$ is defined for *every* concept space $\mathcal{L}$ and is always strictly less than ω_1, the least uncountable ordinal (see Theorem I.6.9 in [29]). It can also be seen that $\mathcal{L}^{(\alpha)}$ is a Π-closed subset of $\mathcal{L}$ for all α.

Assume $L \in \mathcal{L}$ and $\alpha = acc_{\mathcal{L}}(L)$. By definition, L is an isolated point in $\mathcal{L}^{(\alpha)}$, so there is finite $F \subseteq \omega$ such that $\mathcal{L}^{(\alpha)} \cap (\uparrow_{\mathcal{L}^{(\alpha)}} F) = \{L\}$. Since $\mathcal{L}^{(\alpha)}$ is the subset of $\mathcal{L}$ of concepts with accumulation order greater than or equal to α, it follows that L is the only concept in $\mathcal{L}$ that contains F and has accumulation order α. In particular, any concept containing L must have accumulation order strictly less than L, hence $\langle \mathcal{L}, \supseteq \rangle$ is well-founded if $\mathcal{L}$ is scattered.

If $\mathcal{L}$ is a scattered concept space, then by mapping $L \in \mathcal{L}$ to a finite subset F of L such that no other concept containing F has accumulation order $acc_{\mathcal{L}}(L)$, then we obtain an injection from $\mathcal{L}$ to finite subets of ω. Hence, every scattered concept space contains at most a countable number of concepts.

Luo and Schulte showed a nice connection between scattered spaces and mind-change complexity. We will give a simple proof below.

Proposition 3.2.3 (Luo and Schulte [34]) *A concept space $\mathcal{L}$ is learnable with a mind-change bound if and only if $\mathcal{L}$ is scattered. If $\mathcal{L}$ is scattered, then the mind-change complexity of $\mathcal{L}$ equals $acc(\mathcal{L})$.* □

The following property is important for weak-reductions, which we will investigate later.

Definition 3.2.4 (Jain et al. [24]) *A concept space $\mathcal{L}$ is* quasi-dense *if $\mathcal{L}$ is non-empty and for any finite $F \subseteq \omega$, either there exists no concept in $\mathcal{L}$ containing F or else there exist infinitely many distinct concepts in $\mathcal{L}$ containing F.* □

Although the notion of a quasi-dense subspace is not a purely topological property, the following theorem shows that it is invariant among concept spaces with the Π-topology.

Theorem 3.2.5 *The following are equivalent for any concept space $\mathcal{L}$.*

1. *$\mathcal{L}$ is a scattered concept space.*
2. *$\mathcal{L}$ is identifiable from positive data with a mind-change bound.*
3. *$\mathcal{L}$ does not contain a quasi-dense subspace.*
4. *There exists a concept space $\mathcal{K}$ with finite elasticity and an injective* Π-*continuous function $f\colon \mathcal{L} \to \mathcal{K}$.*

Proof: The equivalence of 1 and 2 is the main result of [34]. To show that 1 implies 3, assume $\mathcal{S}$ is a quasi-dense subspace of $\mathcal{L}$. Since any Π-open subset of $\mathcal{L}$ that intersects $\mathcal{S}$ intersects an infinite subset of $\mathcal{S}$, it can be shown using transfinite induction that $\mathcal{S} \subseteq \mathcal{L}^{(\alpha)}$ for all α, and therefore $\mathcal{L}$ is not scattered. To see that 3 implies 1, note that if $\mathcal{L}$ is not scattered then there is a subset $\mathcal{L}^{(\alpha)} = \mathcal{L}^{(\alpha+1)}$ of $\mathcal{L}$ (for some ordinal α) which is infinite (since non-empty finite T_0-spaces necessarily contain isolated points) and easily seen to be quasi-dense.

That 4 implies 1 is also due to Luo and Schulte [34], since they showed that every concept space with finite elasticity is scattered. The rest follows from Theorem 3.3.4 and Proposition 3.3.7.

To show 2 implies 4, assume that $\mathcal{L}$ is learnable with mind-change bound α. Let ι be an indexing of $\mathcal{L}$ (i.e. $\iota\colon \mathcal{L} \to \omega$ is an injective function). For $L \in \mathcal{L}$, define $K_{\iota(L)} = \{\iota(L') \mid L = L' \text{ or } acc_{\mathcal{L}}(L) < acc_{\mathcal{L}}(L')\}$, and let $\mathcal{K} = \{K_{\iota(L)} \mid L \in \mathcal{L}\}$. $\mathcal{K}$ has finite elasticity, because a proof that $\mathcal{K}$ has infinite elasticity could be used to construct an infinitely decreasing sequence of ordinals. Finally, the function $f\colon \mathcal{L} \to \mathcal{K}$ such that $f(L) = K_{\iota(L)}$ is clearly an injection and can be shown to be Π-continuous in the usual way. □

The above theorem can be used to give a simple proof of the latter half of Proposition 3.2.3. First assume $\mathcal{L}$ is scattered and let $\alpha = acc_{\mathcal{L}}(L)$ and let $\langle \mathcal{H}, h\rangle$ be a hypothesis space for $\mathcal{L}$. First note that the $\mathcal{K}$ in the proof above satisfies $acc(\mathcal{L}) = Ord(\langle \mathcal{K}, \supseteq\rangle)$, because $K_{\iota(L)} \subset K_{\iota(L')}$ iff $acc_{\mathcal{L}}(L) > acc_{\mathcal{L}}(L')$. Also note that $f\colon \mathcal{L} \to \mathcal{K}$ is a bijection. Therefore, $\langle \mathcal{H}, f \circ h\rangle$ is a hypothesis space for $\mathcal{K}$. Since $\mathcal{K}$ has finite elasticity, Corollary 2.2.8 implies that there is continuous $\psi\colon \mathcal{T}(\mathcal{K}) \to \mathcal{H}_\alpha^\omega$ that identifies $\mathcal{K}$ from positive data with α mind-changes. Since $f\colon \mathcal{L} \to \mathcal{K}$ is Π-continuous, there is continuous $g\colon \mathcal{T}(\mathcal{L}) \to \mathcal{T}(\mathcal{K})$ such that $f \circ \tau_{\mathcal{L}} = \tau_{\mathcal{K}} \circ g$. Now, if $T \in \mathcal{T}(\mathcal{L})$ is a text for $L \in \mathcal{L}$, then $g(T)$ is a text for $f(L)$, so $\psi(g(T))$ converges to some hypothesis $H \in \mathcal{H}$ for $f(L)$. This means that H satisfies $f(h(H)) = f(L)$, and since f is injective, $h(H) = L$. Therefore, $\psi \circ g\colon \mathcal{T}(\mathcal{L}) \to \mathcal{H}_\alpha^\omega$ identifies $\mathcal{L}$ from positive data with α mind-changes.

The other half of Proposition 3.2.3 is shown as follows. Assume $\psi\colon \mathcal{T}(\mathcal{L}) \to \mathcal{H}_\alpha^\omega$ identifies $\mathcal{L}$ from positive data with α mind-changes. Then for each $\beta < \alpha$,

$$U_\beta = \{\xi \in \mathcal{H}_\alpha^\omega \mid \exists n : \xi(n) = \langle H, \gamma\rangle \text{ and } \gamma \leq \beta\}$$

is an open subset of $\mathcal{H}_\alpha^\omega$, hence $V_\beta = \tau_{\mathcal{L}}(\psi^{-1}(U_\beta))$ is a Π-open subset of $\mathcal{L}$ because ψ is continuous and $\tau_{\mathcal{L}}$ is an open function. We show by induction that if $L \in V_\beta$, then $acc_{\mathcal{L}}(L) \leq \beta$, which implies that $\mathcal{L}$ is scattered. Assume that the hypothesis holds for all $\gamma < \beta$, and we prove it for β. Assume that there is $L \in V_\beta$, and let T be a text for L such that $\psi(T) \in U_\beta$. If $\psi(T)$ is also in U_γ for some $\gamma < \beta$, then by the induction hypothesis $acc_{\mathcal{L}}(L) \leq \gamma < \beta$ and we are done. So assume that $\psi(T) \notin U_\gamma$ for every $\gamma < \beta$.

Since ψ identifies $\mathcal{L}$ with α mind-changes, it follows that there is $n_0 \in \omega$ such that $\psi(T)(n) = \langle H, \beta \rangle$ for all $n \geq n_0$, where H is a hypothesis for L. Therefore, there is an initial finite segment σ of T such that $\psi(T')(n_0) = \langle H, \beta \rangle$ for all $T' \in \mathcal{T}(\mathcal{L})$ extending σ. Thus, if T' extends σ and $\tau_{\mathcal{L}}(T') \neq L$, then there must be $n > n_0$ such that $\psi(T')(n) = \langle H', \gamma \rangle$ and $H' \neq H$ is a hypothesis for L', hence $\gamma < \beta$. Therefore, $\tau_{\mathcal{L}}(T') \in V_\gamma$ for some $\gamma < \beta$, so by the induction hypothesis $acc_{\mathcal{L}}(\tau_{\mathcal{L}}(T')) = \gamma < \beta$, hence $\tau_{\mathcal{L}}(T') \notin \mathcal{L}^{(\beta)}$ by definition. It follows that $\tau_{\mathcal{L}}(\uparrow \sigma)$ is a Π-open set containing L and no other concepts in $\mathcal{L}^{(\beta)}$. Therefore, $L \notin \mathcal{L}^{(\beta+1)}$, so $acc_{\mathcal{L}}(L) \leq \beta$.

3.2.2 Alexandrov spaces

Alexandrov spaces are topological spaces that are useful because of their close relationship to partial orders (see [28]). The definition is purely topological.

Definition 3.2.6 *An* Alexandrov concept space *is a concept space where, for each concept L, there exists a smallest open set containing L.* □

Note that if U is the smallest open set containing $L \in \mathcal{L}$, then U must be equal to $\uparrow_{\mathcal{L}} L$.

The next definition was given as a sufficient criterion for a concept space to be identifiable from positive data.

Definition 3.2.7 (Sato and Umayahara [47]) *A concept $L \in \mathcal{L}$ has an* infinite cross sequence *if and only if there exists an infinite sequence $L_0, L_1, \ldots$ of concepts in $\mathcal{L}$ such that*

1. *$S_0 (\neq \emptyset), S_1, \ldots$ is strictly monotone-increasing, and*
2. *$\bigcup_{i \in \omega} S_i = L$,*

where $S_i = \bigcap_{j=i}^{\infty} (L_j \cap L)$ for $i \geq 0$. We say that $\mathcal{L}$ has finite cross property *if and only if no concept in $\mathcal{L}$ has an infinite cross sequence.* □

The following theorem shows that the topologically invariant property of being an Alexandrov space is equivalent to each concept having a characteristic set and also equivalent to the finite cross property. It also makes it clear that every Alexandrov concept space is countable, because any mapping that sends a concept to a characteristic set for the concept would necessarily be one-to-one.

Theorem 3.2.8 *The following are equivalent for any concept space $\mathcal{L}$.*

1. *$\mathcal{L}$ is an Alexandrov concept space.*
2. *Every L in $\mathcal{L}$ has a characteristic set.*

3. *$\mathcal{L}$ has finite cross property.*

4. *Every L in $\mathcal{L}$ is compact in $\mathcal{A}(\mathcal{L})$.*

Proof: If $\mathcal{L}$ is Alexandrov and $L \in \mathcal{L}$ then $\uparrow_{\mathcal{L}} L$ is Π-open, hence it must be equal to $\uparrow_{\mathcal{L}} F$ for some $F \subseteq L$ (since Π-open sets are defined in terms of unions of Π-basic open sets), which means that F is a characteristic set for L. Therefore, 1 implies 2. The implication from 2 to 1 is similar. The definition of $C_{\mathcal{L}}(\cdot)$ makes it clear that for any finite $F \subseteq L$, $C_{\mathcal{L}}(F) = L$ iff F is a characteristic set for L, hence the equivalence of 2 and 4.

The equivalence of 2 and 3 is mentioned in [46], but we include the proof for completeness. Assume L has no characteristic set. We construct in stages an infinite sequence of concepts $L_0, L_1, \ldots$ in $\mathcal{L}$, and an infinite sequence of natural numbers $x_0, x_1, \ldots$ such that $L = \bigcup_{i=0}^{\infty}\{x_i\}$ and $\{x_0, x_1, \ldots, x_j\} \subseteq L_j$ and $x_{j+1} \notin L_j$ for all $j \in \omega$.

Stage 0. Let x_0 be any element of L.

Stage n+1. Since $\{x_0, x_1, \ldots, x_n\}$ is not a characteristic set of L, there is some $L_n \in \mathcal{L}$ such that $\{x_0, x_1, \ldots, x_n\} \subseteq L_n$ but $L \not\subseteq L_n$. Let x_{n+1} be any element of $L \setminus L_n$. Go to the next stage.

It is easy to verify that $L_0, L_1, \ldots$ is an infinite cross-sequence for L.

For the converse, assume L has a characteristic set and an infinite cross sequence. Since $\bigcup_{i\in\omega} S_i = L$, there will be some i such that S_i contains a characteristic set of L. Therefore, for all $j \geq i$ it follows that $L_j \cap L = L$, since $S_i \subseteq L_j$ implies $L \subseteq L_j$. But this contradicts $S_1, S_2, \ldots$ being strictly monotone-increasing. $\square$

The topological properties of an Alexandrov concept space are determined by the subset relation among its concepts. Π-continuous mappings to and from Alexandrov concept spaces are particularly easy to work with because of the following two lemmas (we leave the easy proofs to the reader).

Lemma 3.2.9 *For any Alexandrov concept space $\mathcal{L}$ and any concept space $\mathcal{K}$, a map $f\colon \mathcal{L} \to \mathcal{K}$ is Π-continuous if and only if it is monotonic with respect to set inclusion.* $\square$

Note, however, that *all* Π-continuous functions are monotonic. The above lemma shows that monotonicity is sufficient for Alexandrov concept spaces.

Lemma 3.2.10 *Assume that $\mathcal{L}$ is an Alexandrov concept space, and let $f\colon \mathcal{K} \to \mathcal{L}$ be a Π-continuous map from any concept space $\mathcal{K}$ to $\mathcal{L}$. Then for every $K \in \mathcal{K}$ there is a finite $F \subseteq K$ such that $\mathcal{A}(f)(C_{\mathcal{K}}(F)) = f(K)$.* $\square$

Note that there exist non-scattered Alexandrov spaces, such as $\mathcal{FIN}$ (the set of all finite subsets of ω), and also non-Alexandrov scattered spaces, such as $\mathcal{L} = \{L_\omega\} \cup \{L_i \mid i \geq 0\}$, where $L_i = \{\langle i, 1\rangle\} \cup \{\langle j, 0\rangle \mid j \leq i\}$ for $i \geq 0$, and $L_\omega = \{\langle j, 0\rangle \mid j \geq 0\}$.

3.2.3 Scattered Alexandrov spaces

We next investigate concept spaces that are both scattered and Alexandrov. It is clear that every concept space with finite elasticity is both scattered and Alexandrov, but we will see in Section 3.2.5 that there are scattered Alexandrov spaces that do not have finite elasticity.

The *closure* of a subset X of a topological space is the intersection of all closed sets containing X.

Definition 3.2.11 *An* irreducible closed set *is a non-empty closed set that is not equal to the union of any two proper closed subsets. A* sober space *is a topological space in which every irreducible closed set is the closure of a unique point.* □

Sobriety is an important topological property because it guarantees that a topological space can be recovered (up to homeomorphism) just by information about the lattice of its open sets (see [28]).

Note that the closure of a concept $L \in \mathcal{L}$ with respect to the Π-topology is the set $\downarrow_{\mathcal{L}} L$.

Theorem 3.2.12 *Every scattered concept space is sober.*

Proof: Let $X \subseteq \mathcal{L}$ be an irreducible Π-closed set, and let $Y \subseteq X$ be the maximal elements of X, i.e., the subset of X of concepts that are not a subset of any other concept in X. Every concept in X is a subset of some concept in Y, since otherwise X would contain an infinite strictly increasing chain of concepts, contradicting $\mathcal{L}$ being scattered. Let $\alpha = \min\{acc_{\mathcal{L}}(L) \mid L \in Y\}$, and let $L_\alpha \in Y$ be such that $acc_{\mathcal{L}}(L_\alpha) = \alpha$. Since $\mathcal{L}$ is scattered, there is a Π-open subset U of $\mathcal{L}$ that contains L_α and does not contain any other concepts with accumulation order greater than or equal to α. Therefore, $U \cap Y = \{L_\alpha\}$, and it follows that X is the union of the two Π-closed sets $\downarrow_{\mathcal{L}} L_\alpha$ and $(\mathcal{L} \setminus U) \cap X$. The latter Π-closed set does not contain L_α, so the irreducibility of X shows that it must be equal to $\downarrow_{\mathcal{L}} L_\alpha$. □

Lemma 3.2.13 *If $\mathcal{L}$ is a sober concept space and $D \subseteq \mathcal{L}$ is directed with respect to subset inclusion, then $\bigcup_{L \in D} L$ is in $\mathcal{L}$.*

Proof: Let X be the topological closure of D. Assume that X is equal to the union of two closed sets X_1 and X_2. Assume for a contradiction that L_1 and L_2 are concepts in D such that $L_1 \in X_1 \setminus X_2$ and $L_2 \in X_2 \setminus X_1$. Since D is directed, there exists $L \in \mathcal{L}$ that contains both L_1 and L_2. Assume without loss of generality that $L \in X_1$. Then $L_2 \in X_1$ because all closed sets are lower sets, but this contradicts the assumption $L_2 \in X_2 \setminus X_1$. Therefore, either $D \subseteq X_1$ or $D \subseteq X_2$, and it follows that $X = X_1$ or $X = X_2$. Thus, X is an irreducible closed set, so $X = \downarrow_{\mathcal{L}} L^*$ for some $L^* \in \mathcal{L}$. If $L^* \neq \bigcup_{L \in D} L$, then there is some finite subset $F \subseteq L^*$ such that $D \cap \uparrow_{\mathcal{L}} F = \emptyset$. But then $X' = (\mathcal{L} \setminus \uparrow_{\mathcal{L}} F) \cap X$ is a proper closed subset of X containing all of D, which contradicts X being the closure of D. Therefore, $L^* = \bigcup_{L \in D} L$. □

Theorem 3.2.14 *For any concept space $\mathcal{L}$, the following are equivalent:*

1. *$\mathcal{L}$ is scattered and Alexandrov,*

2. *$\mathcal{L}$ is sober and Alexandrov.*

Proof: The implication from 1 to 2 follows from Theorem 3.2.12.

Next we show that 2 implies 1. Assume that $\mathcal{L}$ is a sober Alexandrov concept space, and let α be the accumulation order of $\mathcal{L}$. Assume, for a contradiction, that there is some $L \in \mathcal{L}^{(\alpha)}$. By virtue of being Alexandrov, $\uparrow_{\mathcal{L}} L$ is a Π-open subset of $\mathcal{L}$, but since L is not an isolated point of $\mathcal{L}^{(\alpha)}$ there must be some $L' \in \mathcal{L}^{(\alpha)}$ such that $L' \neq L$ and $L' \in \uparrow_{\mathcal{L}} L$, i.e., L' is a strict superset of L. Since L was arbitrary, this shows that there is an infinite strictly increasing chain D of concepts in $\mathcal{L}^{(\alpha)} \subseteq \mathcal{L}$. By Lemma 3.2.13, $L^* = \bigcup_{L \in D} L$ is in $\mathcal{L}$. However, since D is a chain, any finite subset of L^* must be a finite subset of some $L \in D$, so L^* does not have a characteristic set, which contradicts the assumption that $\mathcal{L}$ is Alexandrov. Therefore, $\mathcal{L}^{(\alpha)}$ must be empty, which means $\mathcal{L}$ is scattered. $\square$

It is not true that every sober concept space that is identifiable in the limit from positive data is scattered. For example, let $\mathcal{Q}$ be a concept space that is homeomorphic to the rationals as a subspace of the reals with the Euclidean topology. $\mathcal{Q}$ is sober because it is Hausdorff, and clearly $\mathcal{Q}$ is not scattered, but we will see in the following subsection that every concept in $\mathcal{Q}$ must have a finite tell-tale.

Finally, we point out that the accumulation order of a scattered Alexandrov concept space is completely determined by the ordering of the concepts by the superset relation.

Theorem 3.2.15 *If $\mathcal{L}$ is scattered and Alexandrov, then $acc(\mathcal{L}) = Ord(\langle \mathcal{L}, \supseteq \rangle)$.*

Proof: First note that $\langle \mathcal{L}, \supseteq \rangle$ is well-founded for all scattered concept spaces, and then compare Theorem 2.2.7 and Proposition 3.2.3. $\square$

3.2.4 Countable T_D-spaces

Next we investigate topological properties that characterize identifiability in the limit from positive data.

Definition 3.2.16 *A subset S of topological space X is* locally closed *in X if and only if there exists an open set U and a closed set A such that $S = U \cap A$.* $\square$

The following is a separation axiom proposed by Aull and Thron [6] that is strictly between the T_0 and T_1 axioms (see also [28]).

Definition 3.2.17 (Aull and Thron [6]) *A T_D-space is a topological space X such that $\{x\}$ is locally closed in X for every $x \in X$.* $\square$

Definition 3.2.18 *Let $\mathcal{L}$ be a concept space containing a countably infinite number of concepts. A* distinct ω-indexing *of $\mathcal{L}$ is an indexed family of concepts $\{L_i\}_{i \in \omega}$ such that $\mathcal{L} = \{L_i \,|\, i \in \omega\}$ and for all $i, j \in \omega$, $L_i = L_j$ if and only if $i = j$.* $\square$

The proof of the following lemma is based on the proof by Jain et al. [24] that every learnable concept space can be strongly reduced to $\mathcal{RINIT}_{0,1}$, which is defined as all subsets of the rationals $\mathbb{Q}$ of the form $\{q \in \mathbb{Q} \,|\, 0 \leq q \leq r\}$ for each rational r between 0 and 1 (inclusive).

Lemma 3.2.19 *Let $\mathcal{L}$ be a countably infinite concept space such that every concept in $\mathcal{L}$ has a finite tell-tale. Let $\{L_i\}_{i\in\omega}$ be any distinct ω-indexing of $\mathcal{L}$. Then there exists a family of finite sets $\{F_i\}_{i\in\omega}$ and a partial ordering $\sqsubseteq$ on $\{F_i\}_{i\in\omega}$ extending $\subseteq$ such that*

1. $(\forall\, i \in \omega)[F_i \subseteq L_i]$,
2. $(\forall\, i, j \in \omega)[F_i \subseteq L_j \Rightarrow F_i \sqsubseteq F_j]$,
3. $(\forall\, i, j \in \omega)[F_i = F_j \Rightarrow L_i = L_j]$,

Proof: First, we define a partial order $\sqsubseteq$ extending $\subseteq$ on the set $\mathcal{FIN}$ of all finite subsets of ω. For any $X \subseteq \omega$, let $\min(X)$ denote the least element of X if X is non-empty and ∞ otherwise. For $F, G \in \mathcal{FIN}$, define

$$F \sqsubseteq G \iff \min(G \setminus F) \leq \min(F \setminus G),$$

where we assume that $\infty \leq \infty$, $n \leq \infty$, $\infty \not\leq n$ for all $n \neq \infty$. Then if $F \subseteq G$, $\min(F \setminus G) = \infty$, hence $F \sqsubseteq G$, so $\sqsubseteq$ extends $\subseteq$ and is reflexive. It is easily seen that $\sqsubseteq$ is anti-symmetric, so it only remains to show that it is transitive. Assume $F, G, H \in \mathcal{FIN}$ and $F \sqsubseteq G$ and $G \sqsubseteq H$. Let us abbreviate

$$\begin{aligned} f &= \min\{\min(F \setminus G), \min(F \setminus H)\}, \\ g &= \min\{\min(G \setminus F), \min(G \setminus H)\}, \\ h &= \min\{\min(H \setminus F), \min(H \setminus G)\}. \end{aligned}$$

We first show that $h \leq g \leq f$. If $g = \infty$, then trivially $h \leq g$. Otherwise, if $g \in H$ then g must be in $G \setminus F$ thus $g \in H \setminus F$, hence $h \leq g$. The only remaining possibility is that $g \in G \setminus H$, so since $F \sqsubseteq H$ implies $\min(H \setminus G) \leq \min(H \setminus G)$, we have $h \leq g$. Thus, we have proven in all cases that $h \leq g$. A similar argument shows that $g \leq f$.

Now, if $h = \infty$, then $f = \infty$, so $\min(F \setminus H) = \infty$, thus $F \sqsubseteq H$ and we are finished. Otherwise, it must be the case that $h \in H \setminus F$. This is because if $h \in F$, then $h = \min(H \setminus G)$, thus $h \in F \setminus G$ which implies $f \leq h$. But this would imply that $h = g = f$, but since $g = h \neq \infty$, $g \in G$, contradicting $h = \min(H \setminus G)$. Therefore, $h = \min(H \setminus F) \leq f \leq \min(F \setminus H)$, so $F \sqsubseteq H$. This completes the proof that $\sqsubseteq$ is a partial order.

For any $x \in \omega$, let $\downarrow x = \{y \in \omega \,|\, y \leq x\}$. Let x_0 be the smallest natural number such that $L_0 \cap \downarrow x_0$ is a finite tell-tale of L_0, and define $F_0 = L_0 \cap \downarrow x_0$. For $i > 0$, let

$$\begin{aligned} F_i' &= \{\min(L_i \setminus L_j) \,|\, j < i \text{ and } L_i \not\subset L_j\}, \\ F_i'' &= \bigcup\{F_j \,|\, j < i \text{ and } F_j \subseteq L_i\}. \end{aligned}$$

Finally, let x_i be the smallest natural number such that $F_i' \cup F_i'' \subseteq L_i \cap \downarrow x_i$ and $L_i \cap \downarrow x_i$ is a finite tell-tale of L_i. Define $F_i = L_i \cap \downarrow x_i$.

It follows immediately that $F_i \subseteq L_i$ for all $i \in \omega$. We show that the finite sets are all distinct. Assume without loss of generality that $i > j$ and $F_i = F_j$. Then since $F_j \subseteq L_i$ and F_j is a finite tell-tale of L_j, it follows that $L_i \not\subset L_j$. But then $\min(L_i \setminus L_j) \in F_i = F_j$, which contradicts F_j being a subset of L_j.

Assume $F_i \subseteq L_j$. If $j < i$ then it must be the case that $L_i \subset L_j$, so it follows that $F_j \not\subseteq L_i$ and $L_i \cap \downarrow x_j \subseteq F_j$, proving $\min(F_j - F_i) < \min(F_i - F_j)$. If $j > i$ then $F_i \subset F_j$ so $\min(F_j - F_i) < \min(F_i - F_j)$. If follows that $F_i \sqsubseteq F_j$. □

We next give a topological characterization of identifiability from positive data.

Theorem 3.2.20 *The following are equivalent for any concept space $\mathcal{L}$.*

1. *$\mathcal{L}$ is countable and every L in $\mathcal{L}$ has a finite tell-tale.*
2. *$\mathcal{L}$ is identifiable in the limit from positive data.*
3. *$\mathcal{L}$ is a countable T_D-space.*
4. *There exists an Alexandrov concept space $\mathcal{K}$ and an injective Π-continuous function $f\colon \mathcal{L} \to \mathcal{K}$.*

Proof: The equivalence of 1 and 2 is due to Angluin [3].

That 1 implies 3 can easily be seen by noting that $\downarrow_{\mathcal{L}} L$ is Π-closed for every $L \in \mathcal{L}$, and a finite tell-tale F of L would imply $(\uparrow_{\mathcal{L}} F) \cap (\downarrow_{\mathcal{L}} L) = \{L\}$.

To show that 3 implies 1, assume there are Π-open U and Π-closed V such that $U \cap V = \{L\}$. Then there is a finite $F \subseteq L$ such that $L \in \uparrow_{\mathcal{L}} F \subseteq U$, and since $\downarrow_{\mathcal{L}} L \subseteq V$, it is clear that $(\uparrow_{\mathcal{L}} F) \cap (\downarrow_{\mathcal{L}} L) = \{L\}$. Therefore, F is a finite tell-tale of L.

Next we show that 1 implies 4. If $\mathcal{L}$ is finite then it is Alexandrov so the result is trivial. So let $\{L_i\}_{i\in\omega}$ be a distinct ω-indexing of $\mathcal{L}$ and let $\{F_i\}_{i\in\omega}$ be a family of finite sets partially ordered by $\sqsubseteq$ as in Lemma 3.2.19. Define $K_i = \{j \in \omega \mid F_j \sqsubseteq F_i\}$ for each $i \in \omega$. Let $\mathcal{K} = \{K_i \mid i \in \omega\}$ and define $f(L_i) = K_i$. Since $\sqsubseteq$ is a partial order, it is immediate that $\{i\}$ is a characteristic set for K_i, so $\mathcal{K}$ is Alexandrov. The third criterion of Lemma 3.2.19 guarantees that f is injective. To see that f is Π-continuous, note that the second criterion of Lemma 3.2.19 guarantees that if $F_i \subseteq L_j$, then $i \in f(L_j)$, hence $f(L_i) \subseteq f(L_j)$. Thus, $f^{-1}(\uparrow_{\mathcal{K}} F) = \bigcup\{\uparrow_{\mathcal{L}} F_i \mid F \subseteq f(L_i)\}$ is Π-open in $\mathcal{L}$ for each finite $F \subseteq \omega$.

To see that 4 implies 1, let $L \in \mathcal{L}$ be given. By Lemma 3.2.10, there is a finite $F \subseteq L$ such that $\mathcal{A}(f)(C_{\mathcal{L}}(F)) = f(L)$. For any $L' \in \mathcal{L}$ such that $L' \subseteq L$ and $F \subseteq L'$, the monotonicity of $\mathcal{A}(f)$ implies that $\mathcal{A}(f)(C_{\mathcal{L}}(F)) \subseteq \mathcal{A}(f)(L') \subseteq \mathcal{A}(f)(L)$. Thus $f(L') = f(L)$. Since f is injective, $L = L'$. □

Given a hypothesis space $\langle \mathcal{H}, h\rangle$ for $\mathcal{L}$ and a Π-continuous injection $f\colon \mathcal{L} \to \mathcal{K}$ to an Alexandrov concept space $\mathcal{K}$, we can define a learner ψ_f for $\mathcal{L}$ in the following way. Let $\psi_f(T)(0) = 0$. Assume $\psi_f(T)(n)$ is defined and let

$$\psi_f(T)(n+1) = \begin{cases} H_L & \text{if } \mathcal{A}(f)(C_{\mathcal{L}}(content(T[n+1]))) = f(L), \\ \psi_f(T)(n) & \text{otherwise;} \end{cases}$$

where $H_L \in \mathcal{H}$ is some pre-defined hypothesis such that $h(H_L) = L$. From the above theorem, it is easy to see that ψ_f learns $\mathcal{L}$ in the limit from positive data. Although not all learners for $\mathcal{L}$ can be defined in this way, many "intuitive" learning strategies are of this form. For example, consider $\mathcal{L} = \{L_\omega\} \cup \{L_i \mid i \geq 0\}$, where $L_i = \{\langle i, 1\rangle\} \cup \{\langle j, 0\rangle \mid j \leq i\}$ for $i \geq 0$, and $L_\omega = \{\langle j, 0\rangle \mid j \geq 0\}$. Note that this is the example we gave earlier of a scattered concept space that is not Alexandrov. We can define a concept space $\mathcal{K} = \{K_\perp\} \cup \{K_i \mid i \geq 0\}$, where $K_\perp = \{0\}$ and $K_i = \{0, i+1\}$. Define the Π-continuous function $f\colon \mathcal{L} \to \mathcal{K}$ so that $f(L_\omega) = K_\perp$ and $f(L_i) = K_i$ for $i \geq 0$. Intuitively, ψ_f chooses a hypothesis for L_ω until it sees evidence that confirms otherwise.

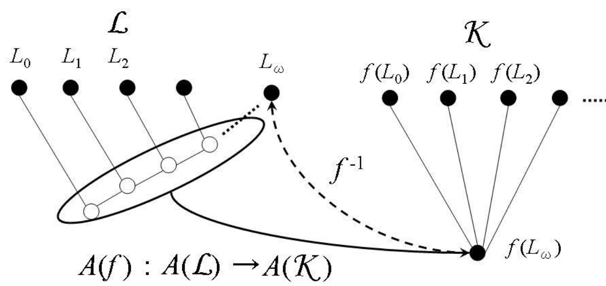


Figure 3.3: Example of a learning strategy for $\mathcal{L}$ based on a Π-continuous injection f from $\mathcal{L}$ to an Alexandrov space $\mathcal{K}$. Solid circles represent concepts in the respective concept spaces, and the open circles on the left represent closed sets in $\mathcal{A}(\mathcal{L})$ that are not in $\mathcal{L}$. $\mathcal{A}(f)$ maps the open circles to $K_\perp = f(L_\omega)$, so ψ_f will output L_ω as a hypothesis until it sees evidence to confirm otherwise.

Finally, we mention that the class of countable T_D concept spaces properly includes both the class of Alexandrov concept spaces and the class of scattered concept spaces. An example of an identifiable concept space that is neither Alexandrov nor scattered is $\mathcal{COSINGLE} = \{\omega \setminus \{n\} \mid n \in \omega\}$. The "intuitive" learning strategy for this space is given by a Π-continuous injection into the ordinal ω (viewed as a concept space, i.e., for each $n \in \omega$ there is a concept $\bar{n} \in \omega$ containing all natural numbers less than n) which maps $\omega \setminus \{n\}$ to $\bar{n}$.

3.2.5 Examples of non-topological properties

In this subsection, we give examples of some well known structural properties of concept spaces that are not topologically invariant.

Finite thickness and finite elasticity

Although Theorem 2.1.16 showed that finite elasticity can be characterized using algebraic closure operators, it is not a topologically invariant property.

Proposition 3.2.21 *Finite thickness and finite elasticity are not topologically invariant properties.*

Proof: Let $\mathcal{SINGLE} = \{\{n\} \mid n \in \omega\}$ and let $\mathcal{L} = \{L_i \mid i \geq 0\} \cup \{J_i \mid i \geq 0\}$ where:

$$
\begin{aligned}
L_0 &= \{\langle n, 0\rangle \mid n \in \omega\} \cup \{\langle 0, 1\rangle\} \\
L_i &= \{\langle n, 0\rangle \mid 0 \leq n \leq 2i \;\&\; n \neq i\} \cup \{\langle i, 1\rangle\} \quad \text{(for } i > 0\text{)}
\end{aligned}
$$

$$\begin{aligned} J_0 &= \{\langle n,1\rangle \mid n \in \omega\} \cup \{\langle 0,0\rangle\} \\ J_i &= \{\langle n,1\rangle \mid 0 \leq n \leq 2i \;\&\; n \neq i\} \cup \{\langle i,0\rangle\} \quad \text{(for } i > 0) \end{aligned}$$

Clearly, $\mathcal{SINGLE}$ has finite thickness. However, since $L_1 \cup J_1 \subset L_2 \cup J_2 \subset \cdots$ and $L_0 \cup J_0 = \bigcup_{i \geq 1} L_i \cup J_i$, it can be proven that $\mathcal{L} \tilde{\cup} \mathcal{L}$ is not identifiable in the limit from positive data. Therefore, $\mathcal{L}$ does not have finite elasticity. But it can easily be shown that $\mathcal{L}$ is homeomorphic to $\mathcal{SINGLE}$, because L_0 is the only concept that contains the subset $\{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,0\rangle\}$, J_0 is the only one that contains $\{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,1\rangle\}$ and all other concepts are finite and not contained in any other concept. Hence, both $\mathcal{L}$ and $\mathcal{SINGLE}$ are countably infinite sets with the discrete topology, thus homeomorphic. □

Note that the concept space $\mathcal{L}$ used in the counter example above is scattered Alexandrov but does not have finite elasticity.

M-finite thickness

Given a concept space $\mathcal{L}$ and $S \subseteq \omega$, $L \in \mathcal{L}$ is a *minimal concept* of S within $\mathcal{L}$ if and only if $S \subseteq L$ and there is no $L' \in \mathcal{L}$ such that $S \subseteq L' \subset L$.

Definition 3.2.22 (Sato and Moriyama [45]) *Let $\mathcal{L}$ be a concept space.*

1. *$\mathcal{L}$ satisfies the* MEF-condition *if and only if for every non-empty finite $F \subseteq \omega$, and any $L \subseteq \mathcal{L}$ such that $F \subseteq L$, then there is $L' \in \mathcal{L}$ such that L' is a minimal concept of F within $\mathcal{L}$ and $L' \subseteq L$.*

2. *$\mathcal{L}$ satisfies the* MFF-condition *if and only if for every non-empty finite $F \subseteq \omega$, the cardinality of*

$$\{L \in \mathcal{L} \mid L \textit{ is a minimal concept of } F \textit{ within } \mathcal{L}\}$$

   is finite.

3. *$\mathcal{L}$ has* M-finite thickness *if and only if $\mathcal{L}$ satisfies both the MEF-condition and the MFF-condition.* □

M-finite thickness was used by Mukouchi [40] in characterizing some variations of approximately identifing concepts in the limit. It was shown by Ambainis et al. [2] that if an indexed family of recursive sets has finite thickness and M-finite thickness, then it is identifiable with a mind-change bound by a computable learner.

For any concept space $\mathcal{L}$ and finite $F \subseteq \omega$, $C_{\mathcal{L}}(F)$ is the unique minimal concept of F within $\mathcal{A}(\mathcal{L})$, hence $\mathcal{A}(\mathcal{L})$ has M-finite thickness. This implies that M-finite thickness is not sufficient for a concept space to be identifiable in the limit from positive data (this fact is also mentioned in [40]). It is clear that every concept space with finite thickness has M-finite thickness, but the concept space $\mathcal{L}'$ in the following proof shows that there are concept spaces with finite elasticity that do not have M-finite thickness.

Proposition 3.2.23 *The MEF-condition, MFF-condition, and M-finite thickness are not topologically invariant properties.*

Proof: For $i \geq 0$ define $K_i = \{x \in \omega \mid x \geq i\}$. Let

$$\mathcal{K} = \{K_i \mid i \geq 0\} \text{ and } \mathcal{K}' = \{K_{i+1} \cup \{0\} \mid i \geq 0\}.$$

Clearly, a function mapping K_i to $K_{i+1} \cup \{0\}$ is a homeomorphism from $\mathcal{K}$ to $\mathcal{K}'$. For any non-empty finite $F \subseteq \omega$, $K_{\min(F)}$ is the unique minimal concept of F within $\mathcal{K}$, hence $\mathcal{K}$ has M-finite thickness. However, there is no minimal concept of $\{0\}$ within $\mathcal{K}'$, so $\mathcal{K}'$ does not satisfy the MEF-condition.

Next, define

$$\mathcal{L} = \{\{x\} \mid x > 0\} \text{ and } \mathcal{L}' = \{\{x, 0\} \mid x > 0\}.$$

Then $\mathcal{L}$ and $\mathcal{L}'$ are homeomorphic and $\mathcal{L}$ has M-finite thickness. However, $\mathcal{L}'$ does not satisfy the MFF-condition because every concept in $\mathcal{L}'$ is a minimal concept of $\{0\}$. $\square$

Although M-finite thickness is not topologically invariant, it does imply some nice topological properties.

Lemma 3.2.24 *If $\mathcal{L}$ has M-finite thickness, then $\uparrow_{\mathcal{L}} F$ is topologically compact for every non-empty finite $F \subseteq \omega$.*

Proof: Let F be a non-empty finite subset of ω. Assume $\uparrow_{\mathcal{L}} F \subseteq \bigcup_{i \in I} U_i$, where $\{U_i\}_{i \in I}$ is a family of Π-open subsets of $\mathcal{L}$. Let

$$Y = \{L \in \mathcal{L} \mid L \text{ is a minimal concept of } F \text{ within } \mathcal{L}\}.$$

Then by the M-finite thickness of $\mathcal{L}$, Y is finite, and for any $L' \in \uparrow_{\mathcal{L}} F$ there is $L \in Y$ such that $L \subseteq L'$. Therefore, there is finite $G \subseteq I$ such that for each $L \in Y$ there is some $i \in G$ such that $L \in U_i$. It follows that $\uparrow_{\mathcal{L}} F \subseteq \bigcup_{i \in G} U_i$, and so $\uparrow_{\mathcal{L}} F$ is topologically compact. $\square$

The converse does not hold. For example, every Π-open subset of

$$\mathcal{COSINGLE} = \{\omega \setminus \{n\} \mid n \in \omega\}$$

is cofinite (i.e., its complement is finite), hence topologically compact. However, it is clear that $\mathcal{COSINGLE}$ does not satisfy the MFF-condition.

It follows from Theorem 3.1.25 and Lemma 3.2.24 that if $\mathcal{J}$ has M-finite thickness, then for any concept space $\mathcal{K}$, $\mathcal{K}^{\mathcal{J}}$ is an exponential object for $\mathcal{J}$ and $\mathcal{K}$. In particular, if $\mathcal{J}$ has finite thickness, or if $\mathcal{J} = \mathcal{A}(\mathcal{L})$ for any concept space $\mathcal{L}$, then the exponential object for $\mathcal{J}$ with any other concept space can be defined.

Pairs of finite tell-tales

The following is another structural property of concept spaces that is sufficient to guarantee identifiability in the limit from positive data.

Definition 3.2.25 (Sato and Umayahara [47]) *Let $\mathcal{L}$ be a concept space, and let $L \in \mathcal{L}$ be a concept. A* pair of finite tell-tales *for L consists of a pair $\langle U, V \rangle$ of finite (possibly empty) sets $U \subseteq L$ and $V \subseteq \omega \setminus L$ such that*

1. *U is a finite tell-tale of L, and*

2. $(\forall L' \in \mathcal{L})[U \subseteq L'$ *and* $V \subseteq \omega \setminus L'$ *implies* $L \subseteq L']$.

A concept space has the PFT-property *if and only if every concept has a pair of finite tell-tales.* □

The first condition is implied by the second condition. Assume $\langle U, V \rangle$ is a pair of finite tell-tales for $L \in \mathcal{L}$, and let $L' \in \mathcal{L}$ be a strict subset of L. Since $V \subseteq \omega \setminus L$ and $L' \subset L$, it is clear that $V \subseteq \omega \setminus L'$. Clearly, $L \not\subseteq L'$, and so it follows that $U \not\subseteq L'$. Thus U is a finite tell-tale of L.

If $\langle U, V \rangle$ is a pair of finite tell-tales for both L and L' in $\mathcal{L}$, then $L = L'$. This implies that any concept space with the PFT-property contains at most a countably infinite number of concepts.

Every Alexandrov concept space has the PFT-property, and every concept space with the PFT-property is a countable T_D-space. The counter example in the following proof shows that the PFT-property is incomparable with the property of being scattered.

Proposition 3.2.26 *The PFT-property is not topologically invariant.*

Proof: We define the following concept spaces:

1. $\mathcal{L} = \{L_\omega\} \cup \{L_i \,|\, i \geq 0\}$,
2. $\mathcal{L}' = \{L_\omega\} \cup \{L_i \cup \{\langle 0, 2 \rangle\} \,|\, i \geq 0\}$,

where $L_i = \{\langle i, 1 \rangle\} \cup \{\langle j, 0 \rangle \,|\, j \leq i\}$ for $i \geq 0$, and $L_\omega = \{\langle j, 0 \rangle \,|\, j \geq 0\}$.

Clearly $\mathcal{L}$ and $\mathcal{L}'$ are homeomorphic. $\mathcal{L}'$ has the PFT-property because the only infinite concept is L_ω, which has $\langle \emptyset, \{\langle 0, 2 \rangle\} \rangle$ as a pair of finite tell-tales. On the other hand, $\mathcal{L}$ does not have the PFT-property, because L_ω has no pair of finite tell-tales. For assume that $\langle U, V \rangle$ is a pair of finite tell-tales for L_ω. Let i be large enough that $j < i$ for all $\langle j, 0 \rangle \in U$ and $k < i$ for all $\langle k, 1 \rangle \in V$. Then clearly $U \subseteq L_i$ and $V \subseteq \omega \setminus L_i$, but $L_\omega \not\subseteq L_i$. □

3.3 Reductions between concept spaces

In this section, we analyze reductions between concept spaces. Reductions are a way of reducing one learning problem into another learning problem. Reductions for identification in the limit of concept spaces was introduced by Jain and Sharma [27], and further investigated by Jain et al. [24]. Here we give topological characterizations of when concept spaces are reducible to one another.

The basic idea is that if we have a reduction from a concept space $\mathcal{J}$ to a concept space $\mathcal{K}$, then we can convert any text T for a concept in $\mathcal{J}$ to a text T' for some concept in $\mathcal{K}$, in a way that if we can identify $\tau_{\mathcal{K}}(T')$ then we can identify $\tau_{\mathcal{J}}(T)$. This gives us a means of reducing the problem of learning $\mathcal{J}$ to the problem of learning $\mathcal{K}$.

3.3.1 Weak and strong reductions

The following is a variation of the definition given in [27] for reductions between concept spaces, where enumeration operators are replaced by continuous functions. It is easy to see that the definitions give equivalent notions of reducibility when we allow non-computable enumeration operators.

Given a hypothesis space $\langle \mathcal{H}, h \rangle$ for a concept space $\mathcal{L}$, we let

$$\lim_{\mathcal{H}} : \subseteq \mathcal{H}^\omega \to \mathcal{H}$$

send a converging sequence of hypotheses to the hypothesis to which it converges, and let it be undefined on non-converging sequences. It follows that a continuous function $\psi : \mathcal{T}(\mathcal{L}) \to \mathcal{H}^\omega$ identifies $\mathcal{L}$ in the limit from positive data if and only if $\tau_{\mathcal{L}} = h \circ \lim_{\mathcal{H}} \circ \psi$.

Definition 3.3.1 *Let $\mathcal{J}$ and $\mathcal{K}$ be concept spaces, and $\langle \mathcal{H}_\mathcal{J}, h_\mathcal{J} \rangle$ and $\langle \mathcal{H}_\mathcal{K}, h_\mathcal{K} \rangle$ be hypothesis spaces for $\mathcal{J}$ and $\mathcal{K}$, respectively. We let $\lim_{\mathcal{H}_\mathcal{J}} : \mathcal{H}_\mathcal{J}^\omega \to \mathcal{H}_\mathcal{J}$ and $\lim_{\mathcal{H}_\mathcal{K}} : \mathcal{H}_\mathcal{K}^\omega \to \mathcal{H}_\mathcal{K}$ denote the respective limit functions. A* weak reduction *from $\mathcal{J}$ to $\mathcal{K}$ (with respect to $\langle \mathcal{H}_\mathcal{J}, h_\mathcal{J} \rangle$ and $\langle \mathcal{H}_\mathcal{K}, h_\mathcal{K} \rangle$) is a pair $\langle \Theta, \Psi \rangle$ such that*

1. *$\Theta : \mathcal{T}(\mathcal{J}) \to \mathcal{T}(\mathcal{K})$ and $\Psi : \subseteq \mathcal{H}_\mathcal{K}^\omega \to \mathcal{H}_\mathcal{J}^\omega$ are Π-continuous functions, and*
2. *For any $f : \mathcal{T}(\mathcal{K}) \to \mathcal{H}_\mathcal{K}^\omega$,*

$$\tau_\mathcal{K} = h_\mathcal{K} \circ \lim\nolimits_{\mathcal{H}_\mathcal{K}} \circ f \Rightarrow \tau_\mathcal{J} = h_\mathcal{J} \circ \lim\nolimits_{\mathcal{H}_\mathcal{J}} \circ \Psi \circ f \circ \Theta.$$

$$\begin{tikzcd} \mathcal{T}(\mathcal{K}) \arrow[rr, "f"] \arrow[dr, "\tau_\mathcal{K}"'] & & \mathcal{H}_\mathcal{K}^\omega \arrow[dl, "h_\mathcal{K} \circ \lim_{\mathcal{H}_\mathcal{K}}"] \\ & \mathcal{K} & \end{tikzcd} \quad \Rightarrow \quad \begin{tikzcd} \mathcal{T}(\mathcal{J}) \arrow[r, "\Theta"] \arrow[drr, "\tau_\mathcal{J}"'] & \mathcal{T}(\mathcal{K}) \arrow[r, "f"] & \mathcal{H}_\mathcal{K}^\omega \arrow[r, "\Psi"] & \mathcal{H}_\mathcal{J}^\omega \arrow[dl, "h_\mathcal{J} \circ \lim_{\mathcal{H}_\mathcal{J}}"] \\ & & \mathcal{J} & \end{tikzcd}$$

If, in addition, $\tau_\mathcal{J}(T) = \tau_\mathcal{J}(T') \Rightarrow \tau_\mathcal{K} \circ \Theta(T) = \tau_\mathcal{K} \circ \Theta(T')$ for all T and T' in $\mathcal{T}(\mathcal{J})$, then we say that $\langle \Theta, \Psi \rangle$ is a strong reduction *from $\mathcal{J}$ to $\mathcal{K}$.* □

Note that $f : \mathcal{T}(\mathcal{K}) \to \mathcal{H}_\mathcal{K}^\omega$ in the above definition varies over all functions, and not just continuous ones. The definition above says that if $\langle \Theta, \Psi \rangle$ is a weak reduction from $\mathcal{J}$ to $\mathcal{K}$ and f learns $\mathcal{K}$, then $\Psi \circ f \circ \Theta$ learns $\mathcal{J}$. Thus, we can reduce the problem of learning $\mathcal{J}$ to the problem of learning $\mathcal{K}$. Below, we will write $\mathcal{J} \leq_W \mathcal{K}$ if there exists a weak reduction from $\mathcal{J}$ to $\mathcal{K}$, and write $\mathcal{J} \leq_S \mathcal{K}$ if there exists a strong reduction.

Proposition 3.3.2 *Let $\mathcal{J}$ and $\mathcal{K}$ be concept spaces. $\mathcal{J} \leq_W \mathcal{K}$ (with respect to* any *choice of hypothesis spaces) if and only if there exists a Π-continuous function $\Theta : \mathcal{T}(\mathcal{J}) \to \mathcal{T}(\mathcal{K})$ and an equivalence relation $\equiv_\mathcal{K}$ on $\mathcal{K}$ such that*

$$\tau_\mathcal{J}(T) = \tau_\mathcal{J}(T') \iff \tau_\mathcal{K}(\Theta(T)) \equiv_\mathcal{K} \tau_\mathcal{K}(\Theta(T'))$$

for every $T, T' \in \mathcal{T}(\mathcal{J})$. Furthermore, $\mathcal{J} \leq_S \mathcal{K}$ if and only if there exists a Θ and $\equiv_\mathcal{K}$ as above in which $\equiv_\mathcal{K}$ is the usual equality on $\mathcal{K}$.

Proof: Assume that $\langle \Theta, \Psi \rangle$ is a weak reduction from $\mathcal{J}$ to $\mathcal{K}$. Define a relation $\equiv_\mathcal{K}$ on $\mathcal{K}$ such that $K \equiv_\mathcal{K} K'$ if and only if $K = K'$ or else there are $T_K, T_{K'} \in \mathcal{T}(\mathcal{J})$ such that $\tau_\mathcal{J}(T_K) = \tau_\mathcal{J}(T_{K'})$ and $\tau_\mathcal{K}(\Theta(T_K)) = K$ and $\tau_\mathcal{K}(\Theta(T_{K'})) = K'$. It is easy to check that $\equiv_\mathcal{K}$ is an equivalence relation, and that $\tau_\mathcal{J}(T) = \tau_\mathcal{J}(T')$ implies $\tau_\mathcal{K}(\Theta(T)) \equiv_\mathcal{K} \tau_\mathcal{K}(\Theta(T'))$ for all $T, T' \in \mathcal{T}(\mathcal{J})$.

It remains to show that $\tau_\mathcal{K}(\Theta(T)) \equiv_\mathcal{K} \tau_\mathcal{K}(\Theta(T'))$ always implies $\tau_\mathcal{J}(T) = \tau_\mathcal{J}(T')$. First note that $\tau_\mathcal{K}(\Theta(T)) = \tau_\mathcal{K}(\Theta(T'))$ implies that $\tau_\mathcal{J}(T) = \tau_\mathcal{J}(T')$ for all texts T and T' in $\mathcal{T}(\mathcal{J})$. This is because we can define a (possibly non-continuous) $f : \mathcal{T}(\mathcal{K}) \to \mathcal{H}_\mathcal{K}^\omega$ that, for each $K \in \mathcal{K}$, sends every text in $\tau_\mathcal{K}^{-1}(K)$ to a *unique* element of $(h_{\mathcal{K}} \circ \lim_{\mathcal{H}_{\mathcal{K}}})^{-1}(K)$. Clearly f satisfies $\tau_{\mathcal{K}} = h_{\mathcal{K}} \circ \lim_{\mathcal{H}_{\mathcal{K}}} \circ f$, so it follows that $\tau_{\mathcal{J}} = h_{\mathcal{J}} \circ \lim_{\mathcal{H}_{\mathcal{J}}} \circ \Psi \circ f \circ \Theta$. Therefore, if $\tau_{\mathcal{K}}(\Theta(T)) = \tau_{\mathcal{K}}(\Theta(T'))$, then since $f(\Theta(T)) = f(\Theta(T'))$, it follows that $\tau_{\mathcal{J}}(T) = \tau_{\mathcal{J}}(T')$.

Now, assume that $K = \tau_{\mathcal{K}}(\Theta(T)) \equiv_{\mathcal{K}} \tau_{\mathcal{K}}(\Theta(T')) = K'$. If $K = K'$ then $\tau_{\mathcal{J}}(T) = \tau_{\mathcal{J}}(T')$ follows from the argument in the previous paragraph. Otherwise, by definition there are T_K and $T_{K'}$ in $\mathcal{T}(\mathcal{J})$ such that $\tau_{\mathcal{J}}(T_K) = \tau_{\mathcal{J}}(T_{K'})$ and $\tau_{\mathcal{K}}(\Theta(T_K)) = K$ and $\tau_{\mathcal{K}}(\Theta(T_{K'})) = K'$. Therefore, $\tau_{\mathcal{J}}(T) = \tau_{\mathcal{J}}(T_K) = \tau_{\mathcal{J}}(T_{K'}) = \tau_{\mathcal{J}}(T')$.

Thus, Θ and $\equiv_{\mathcal{K}}$ fulfill the claim of the proposition, and $\equiv_{\mathcal{K}}$ is the usual equality on $\mathcal{K}$ if and only if $\langle \Theta, \Psi \rangle$ is a strong reduction.

For the converse, assume that Θ and $\equiv_{\mathcal{K}}$ satisfy the conditions of the proposition. For each $J \in \mathcal{J}$, let $H_J \in \mathcal{H}_{\mathcal{J}}$ be such that $h_{\mathcal{J}}(H_J) = J$. We next define a function $\phi: \mathcal{H}_{\mathcal{K}} \to \mathcal{H}_{\mathcal{J}}$. For $H \in \mathcal{H}_{\mathcal{K}}$, if there is $T \in \mathcal{T}(\mathcal{J})$ such that $\tau_{\mathcal{J}}(T) = J$ and $\tau_{\mathcal{K}}(\Theta(T)) = h_{\mathcal{K}}(H)$, then define $\phi(H) = H_J$. If there is no such $T \in \mathcal{T}(\mathcal{J})$, then define $\phi(H)$ to be any element of $\mathcal{H}_{\mathcal{J}}$. Define $\Psi: \mathcal{H}_{\mathcal{K}}^{\omega} \to \mathcal{H}_{\mathcal{J}}^{\omega}$ so that $\Psi(S)(n) = \phi(S(n))$ for each $S \in \mathcal{H}_{\mathcal{K}}^{\omega}$ and $n \in \omega$. Ψ is clearly Π-continuous, and for every $J \in \mathcal{J}$ and $T \in \tau_{\mathcal{J}}^{-1}(J)$, if $S \in \mathcal{H}_{\mathcal{K}}^{\omega}$ is such that $h_{\mathcal{K}}(\lim_{\mathcal{H}_{\mathcal{K}}}(S)) = \tau_{\mathcal{K}}(\Theta(T))$, then $h_{\mathcal{J}}(\lim_{\mathcal{H}_{\mathcal{J}}}(\Psi(S))) = J$. Therefore, $\langle \Theta, \Psi \rangle$ is a weak reduction from $\mathcal{J}$ to $\mathcal{K}$, and it is a strong reduction if $\equiv_{\mathcal{K}}$ is the equality on $\mathcal{K}$. □

3.3.2 Multivalued functions

In order to characterize weak reducibility, the following notion of multivalued functions is needed. Theorem 3.3.3 below is interesting independent of reductions.

A *multivalued function* from X to Y is a function $f: X \to \mathcal{P}(Y)$, where $\mathcal{P}(Y)$ denotes the powerset of Y. We write $f: X \rightrightarrows Y$ to denote a multivalued function f from X to Y. For $A \subseteq X$ and $B \subseteq Y$, define $f(A) = \bigcup_{x \in A} f(x)$ and $f^{-1}(B) = \{x \in X \mid f(x) \cap B \neq \emptyset\}$. A multivalued function $f: X \rightrightarrows Y$ is *lower semicontinuous* if and only if $f^{-1}(U)$ is open in X for every open $U \subseteq Y$. We say that $f: X \rightrightarrows Y$ is *absolutely injective* if and only if $x \neq y$ implies $f(x) \cap f(y) = \emptyset$ for all $x, y \in X$.

Theorem 3.3.3 *Let $\mathcal{J}$ be a countable T_D space and $f: \mathcal{J} \rightrightarrows \mathcal{K}$ a lower semicontinuous multivalued function. Then there is a continuous function $g: \mathcal{T}(\mathcal{J}) \to \mathcal{T}(\mathcal{K})$ such that $\tau_{\mathcal{K}}(g(T)) \in f(\tau_{\mathcal{J}}(T))$ for every $T \in \mathcal{T}(\mathcal{J})$.*

Proof: Since $\mathcal{J}$ and $\mathcal{K}$ are countable, we can assume there are well-orderings $\leq_{\mathcal{J}}$ of $\mathcal{J}$ and $\leq_{\mathcal{K}}$ of $\mathcal{K}$ with order type less than or equal to ω. Then it makes sense to say "the $\leq_{\mathcal{J}}$*-least* $J \in \mathcal{J}$ such that...", and for any $J \in \mathcal{J}$, the set $\{J' \in \mathcal{J} \mid J' \leq_{\mathcal{J}} J\}$ is finite (and similarly for $\mathcal{K}$).

Using the fact that $\mathcal{J}$ is a T_D-space and $\tau_{\mathcal{J}}$ is continuous, for each $J \in \mathcal{J}$ we can choose open sets $U_J, N_J \subseteq \mathcal{T}(\mathcal{J})$ such that $\tau_{\mathcal{J}}^{-1}(\{J\}) = U_J \setminus N_J$.

We now define $g(T)$ in stages for any given $T \in \mathcal{T}(\mathcal{J})$. For stage $n \geq 0$, consider the following two cases depending on whether or not there is $J \in \mathcal{J}$ such that

$$\uparrow T[n] \subseteq U_J \text{ and } \uparrow T[n] \not\subseteq N_J.$$

Case 1: No such $J \in \mathcal{J}$ exists. Then define $g(T)(n) = \#$.

Case 2: There is such a $J \in \mathcal{J}$. Then let $J_n \in \mathcal{J}$ be the $\leq_{\mathcal{J}}$-least such J. Let

$$F_n = \{x \in \omega \mid \exists m < n\colon g(T)(m) = x\},$$

and let K_n be the $\leq_{\mathcal{K}}$-least element of $f(J_n) \cap \uparrow_{\mathcal{K}} F_n$ (we will see that such a K_n always exists). Now consider the following three subcases.

Subcase 2.A: $K_n \setminus F_n = \emptyset$. Then define $g(T)(n) = \#$.

If $K_n \setminus F_n \neq \emptyset$, then let $m = \min(K_n \setminus F_n)$ in the following subcases.

Subcase 2.B: $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}}(F_n \cup \{m\}))) \not\supseteq \uparrow T[n]$. Define $g(T)(n) = \#$.

Subcase 2.C: $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}}(F_n \cup \{m\}))) \supseteq \uparrow T[n]$. Define $g(T)(n) = m$.

This completes the definition of $g(T)$.

We first confirm that the K_n in Case 2 always exists. Note that it suffices to show that $f(J_n) \cap \uparrow_{\mathcal{K}} F_n$ is non-empty, since every non-empty subset of $\mathcal{K}$ contains a $\leq_{\mathcal{K}}$-least element by the definition of a well-ordering.

Claim: For all $n \geq 0$ in which Case 2 applies,

1. $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} F_n)) \supseteq \uparrow T[n]$,
2. $f(J_n) \cap \uparrow_{\mathcal{K}} F_n$ is non-empty.

Proof: Let n_0 be the least n such that Case 2 holds. Then clearly $F_{n_0} = \emptyset$, so $\uparrow_{\mathcal{K}} F_{n_0} = \mathcal{K}$. It follows that $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} F_{n_0})) = \mathcal{T}(\mathcal{J}) \supseteq \uparrow T[n_0]$ and $f(J_{n_0}) \cap \uparrow_{\mathcal{K}} F_n$ is non-empty.

Next, assume Case 2 applies to stage n_i and the claim holds. Let $n_{i+1} > n_i$ be the next stage in which Case 2 holds. Clearly, $\uparrow T[n_{i+1}] \subseteq \uparrow T[n_i]$ holds. If $F_{n_{i+1}} = F_{n_i}$, then $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} F_{n_{i+1}})) = \tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} F_{n_i})) \supseteq \uparrow T[n_i] \supseteq \uparrow T[n_{i+1}]$. If $F_{n_{i+1}} \neq F_{n_i}$, then Subcase 2.C must apply to stage n_i, so $F_{n_{i+1}} = F_{n_i} \cup \{m\}$ where m satisfies $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}}(F_{n_i} \cup \{m\}))) \supseteq \uparrow T[n_i] \supseteq \uparrow T[n_{i+1}]$.

For any $J \in \mathcal{J}$, if $\uparrow T[n_{i+1}] \subseteq U_J$ and $\uparrow T[n_{i+1}] \not\subseteq N_J$, there must be $T' \in \uparrow T[n_{i+1}]$ such that $T' \in U_J \setminus N_J$. Note that $\tau_{\mathcal{J}}(T') = J$. Since $\tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} F_{n_{i+1}})) \supseteq \uparrow T[n_{i+1}]$, $T' \in \tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} F_{n_{i+1}}))$, hence $J = \tau_{\mathcal{J}}(T') \in f^{-1}(\uparrow_{\mathcal{K}} F_{n_{i+1}})$, thus $f(J) \cap \uparrow_{\mathcal{K}} F_{n_{i+1}}$ is non-empty. In particular, $f(J_{n_{i+1}}) \cap \uparrow_{\mathcal{K}} F_{n_{i+1}}$ is non-empty. *(End of proof of Claim)*

Now fix $T \in \mathcal{T}(\mathcal{J})$ and assume that $\tau_{\mathcal{J}}(T) = J$. Then

$$U = U_J \cap \bigcup_{J' <_{\mathcal{J}} J} N_{J'}$$

is open in $\mathcal{T}(\mathcal{J})$ because $\leq_{\mathcal{J}}$ has order type less than or equal to ω. Clearly, $T \in U$ and $T \notin N_J$, so since U is open there is $n_0 \geq 0$ such that $\uparrow T[n] \subseteq U \subseteq U_J$ and $\uparrow T[n] \not\subseteq N_J$ for all $n \geq n_0$. Thus, Case 2 holds for all $n \geq n_0$, and clearly J is the $\leq_{\mathcal{J}}$-least element of $\mathcal{J}$ satisfying the condition for Case 2, hence $J_n = J$ for all $n \geq n_0$. Let K be the $\leq_{\mathcal{K}}$-least element of $\mathcal{K}$ of $f(J) \cap \uparrow_{\mathcal{K}} F_{n_0}$. Then clearly $F_n \subseteq K$ for all $n \geq n_0$, hence $K_n = K$ for all $n \geq n_0$.

We now show that $\tau_{\mathcal{K}}(g(T)) = K$. Clearly any natural number occuring in $g(T)$ is an element of K, because $F_n \subseteq K$ for all $n \geq n_0$. Choose any $n \geq n_0$ and assume $K \setminus F_n$ is non-empty. Let $m = \min(K \setminus F_n)$. Then $V = \tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}}$

$(F_n \cup \{m\})))$ is an open subset of $\mathcal{T}(\mathcal{J})$. Clearly, $T \in V$, so let n' be the least number greater than or equal to n such that $\uparrow T[n'] \subseteq V$. Then obviously $F_{n'} = F_n$, thus $V = \tau_{\mathcal{J}}^{-1}(f^{-1}(\uparrow_{\mathcal{K}} (F_{n'} \cup \{m\})))$, so $g(T)(n') = m$. It follows that, $g(T)$ is a text for K.

Therefore, $g\colon \mathcal{T}(\mathcal{J}) \to \mathcal{T}(\mathcal{K})$ satisfies $\tau_{\mathcal{K}}(g(T)) \in f(\tau_{\mathcal{J}}(T))$ for every $T \in \mathcal{T}(\mathcal{J})$. Clearly, g is continuous because each initial finite segment of $g(T)$ only depends on an initial finite segment of T. □

3.3.3 Topological characterization of reductions

In the subsection we characterize reducibility in terms of continuous functions and multivalued functions. Luo and Schulte [34] were the first to notice that a strong reduction between concept spaces induces an injective continuous function between them. We next show that the converse also holds.

Theorem 3.3.4 *For any concept spaces $\mathcal{J}$ and $\mathcal{K}$, $\mathcal{J} \leq_S \mathcal{K}$ if and only if there exists a Π-continuous injection from $\mathcal{J}$ to $\mathcal{K}$.*

Proof: If $\langle \Theta, \Psi \rangle$ is a strong reduction from $\mathcal{J}$ to $\mathcal{K}$, then by Proposition 3.3.2 $\tau_{\mathcal{J}}(T) = \tau_{\mathcal{J}}(T') \iff \tau_{\mathcal{K}}(\Theta(T)) = \tau_{\mathcal{K}}(\Theta(T'))$ for every $T, T' \in \mathcal{T}(\mathcal{J})$. This implies that $f\colon \mathcal{J} \to \mathcal{K}$, uniquely determined as satisfying $f(\tau_{\mathcal{J}}(T)) = f(\tau_{\mathcal{K}}(\Theta(T)))$ for all $T \in \mathcal{T}(\mathcal{J})$, is a well defined function and injective. By Corollary 3.1.10, f is Π-continuous.

Conversely, if $g\colon \mathcal{J} \to \mathcal{K}$ is a Π-continuous injection, then from Corollary 3.1.10 there exists a Π-continuous $\Theta\colon \mathcal{T}(\mathcal{J}) \to \mathcal{T}(\mathcal{K})$ such that $g \circ \tau_{\mathcal{J}} = \tau_{\mathcal{K}} \circ \Theta$. Clearly Θ satisfies Proposition 3.3.2 with $\equiv_{\mathcal{K}}$ as the usual equality on $\mathcal{K}$. □

Next we give a characterization of when $\mathcal{J}$ is weakly reducible to $\mathcal{K}$, provided that $\mathcal{J}$ is identifiable in the limit from positive data. The following characterization is essentially Theorem 3.3.4 with single valued functions replaced by multivalued functions.

Theorem 3.3.5 *Let $\mathcal{J}$ be a countable T_D concept space and let $\mathcal{K}$ be an arbitrary concept space. Then $\mathcal{J} \leq_W \mathcal{K}$ if and only if there exists an absolutely injective lower semicontinuous multivalued function $f\colon \mathcal{J} \rightrightarrows \mathcal{K}$.*

Proof: ($\Rightarrow$). Assume that $\langle \Theta, \Psi \rangle$ is a weak reduction from $\mathcal{J}$ to $\mathcal{K}$. Define $f\colon \mathcal{J} \rightrightarrows \mathcal{K}$ by $K \in f(J)$ if and only if there is $T \in \mathcal{T}(\mathcal{J})$ such that $\tau_{\mathcal{J}}(T) = J$ and $\tau_{\mathcal{K}}(\Theta(T)) = K$. As shown in the proof of Proposition 3.3.2, $\tau_{\mathcal{K}}(\Theta(T)) \equiv_{\mathcal{K}} \tau_{\mathcal{K}}(\Theta(T'))$ implies $\tau_{\mathcal{J}}(T) = \tau_{\mathcal{J}}(T')$, which implies that f is absolutely injective.

Let U be a Π-open subset of $\mathcal{K}$. Then $W = \tau_{\mathcal{J}}(\Theta^{-1}(\tau_{\mathcal{K}}^{-1}(U)))$ is Π-open in $\mathcal{J}$ because Θ and $\tau_{\mathcal{K}}$ are Π-continuous and $\tau_{\mathcal{J}}$ is an open map. Furthermore, it is clear from the definition of f that $W = f^{-1}(U)$. It follows that f is an absolutely injective lower semicontinuous multivalued function from $\mathcal{J}$ to $\mathcal{K}$.

($\Leftarrow$). Assume $f\colon \mathcal{J} \rightrightarrows \mathcal{K}$ is an absolutely injective lower semicontinuous multivalued function. By Theorem 3.3.3, there is a continuous $\Theta\colon \mathcal{T}(\mathcal{J}) \to \mathcal{T}(\mathcal{K})$ such that $\tau_{\mathcal{K}}(\Theta(T)) \in f(\tau_{\mathcal{J}}(T))$ for every $T \in \mathcal{T}(\mathcal{J})$. Define an equivalence relation $\equiv_{\mathcal{K}}$ on $\mathcal{K}$ by $K \equiv_{\mathcal{K}} K'$ if and only if $K = K'$ or else there is $J \in \mathcal{J}$ such that both K and K' are in $f(K)$. Since f is absolutely injective, it is clear that

$$\tau_{\mathcal{J}}(T) = \tau_{\mathcal{J}}(T') \iff \tau_{\mathcal{K}}(\Theta(T)) \equiv_{\mathcal{K}} \tau_{\mathcal{K}}(\Theta(T'))$$

for every $T, T' \in \mathcal{T}(\mathcal{J})$. By Proposition 3.3.2, it follows that $\mathcal{J} \leq_W \mathcal{K}$. □

By reviewing the proof of Theorem 3.3.5, it is clear that $\mathcal{J} \leq_W \mathcal{K}$ implies there exists an absolutely injective lower semicontinuous multivalued function $\mathcal{J}$ to $\mathcal{K}$ even for *arbitrary* $\mathcal{J}$. In general, however, this criterion is not sufficient to guarantee weak reducibility when $\mathcal{J}$ is not a countable T_D space, which we will now show.

Proposition 3.3.6 *Assume $\mathcal{J}$ is a countable concept space and $\mathcal{K}$ is quasi-dense. Then there exists an absolutely injective lower semicontinuous multivalued function $f\colon \mathcal{J} \rightrightarrows \mathcal{K}$.*

Proof: Let $\{F_i\}_{i\in\omega}$ be an enumeration of all finite subsets of ω such that $\uparrow_{\mathcal{K}} F_i$ is non-empty for all $i \in \omega$. Let $q\colon \omega\times\omega \to \mathcal{K}$ be an injection such that $F_i \subseteq q(\langle i, j\rangle)$ for each $\langle i, j\rangle \in \omega \times \omega$. Such an injection can easily be constructed because by assumption $\uparrow_{\mathcal{K}} F_i$ is an infinite subset of $\mathcal{K}$ for each $i \in \omega$. Let $g\colon \mathcal{J} \to \omega$ be injective, and define $f\colon \mathcal{J} \rightrightarrows \mathcal{K}$ so that $f(J) = \{q(\langle i, g(J)\rangle) \mid i \in \omega\}$. Clearly f is absolutely injective because q and g are injective. If $F \subseteq \omega$ is finite and $\uparrow_{\mathcal{K}} F$ is non-empty, then $F = F_i$ for some $i \in \omega$. Since $F_i \subseteq q(\langle i, j\rangle)$ for each $j \in \omega$, clearly $f^{-1}(\uparrow_{\mathcal{K}} F) = \mathcal{J}$. Therefore, f is lower semicontinuous. □

Thus, if $\mathcal{K}$ contains a quasi-dense subspace and $\mathcal{J}$ is countable, then there exists an absolutely injective lower semicontinuous multivalued function from $\mathcal{J}$ to $\mathcal{K}$. Since there are countable concept spaces that are not T_D spaces, and concept spaces with quasi-dense subspaces that are T_D spaces, this shows that Theorem 3.3.5 cannot be generalized in its current form to allow $\mathcal{J}$ that are not countable T_D spaces.

A *weak-complete* [27] concept space is one that is identifiable in the limit, and for which every other identifiable concept space is weakly reducible to it. Jain et al. [24] characterized weak-complete concept spaces as precisely those that are identifiable in the limit and contain a quasi-dense subspace. This characterization is easily seen to follow from Proposition 3.3.6 above. It follows from Theorem 3.2.5 that a concept space is weak-complete if and only if it is a non-scattered countable T_D space.

Finally, we mention a result by Luo and Schulte that shows that weak reductions preserve mind-change complexity.

Proposition 3.3.7 (Luo and Schulte [34]) *If $\mathcal{K}$ is a scattered concept space and $\mathcal{J} \leq_W \mathcal{K}$ then $\mathcal{J}$ is scattered and $acc(\mathcal{J}) \leq acc(\mathcal{K})$.* □

Chapter 4

Representations of Concept Spaces

In the identification in the limit paradigm, a learner essentially converts one representation (a stream of information) of some concept into another representation (a sequence of hypotheses) of the concept. It is important to understand how these representations are different, and why one representation may be preferred over another one.

When a concept space is represented by texts, the natural topology induced on the concept space is the Π-topology as we saw in the previous chapter. On the other hand, when a concept space is represented by a hypothesis space, the induced topology is the discrete topology (i.e., every subset of the concept space is open). The representation by a hypothesis space is in a sense more informative than texts because it allows every subset of the concept space to be observable. From this perspective, the learner's goal is to convert a representation of the concept space into a representation that allows more properties to be observable. From a more semantical perspective, the learner's goal is to refine the topology of the concept space.

However, in general this is an impossible task because the operation of converting a topological space into a space with strictly more open sets is not continuous. Therefore, the learner can not directly convert a representation in terms of texts into a representation by a hypothesis space. Instead, the identification in the limit paradigm asks if the learner can make this conversion in the limit. Thus, instead of the representation $\langle \mathcal{H}, h \rangle$ of the concept space in terms of individual hypotheses, the learner produces a *limiting* representation $\langle \mathcal{H}^\omega, \lim_{\mathcal{H}} \rangle$ in terms of a converging sequence of hypotheses. Although identification in the limit produces a concept space with a finer topology, the cost is that the representation of the concepts becomes more abstract.

In this chapter we investigate a hierarchy of representations of topological spaces which allow us to quantify the level of "abstraction" of representations. For example, texts form a $\mathbf{\Sigma}^0_1$-admissible representation of the concept space with respect to the Π-topology, whereas the limiting representation $\langle \mathcal{H}^\omega, \lim_{\mathcal{H}} \rangle$ is a $\mathbf{\Sigma}^0_2$-admissible representation with respect to the discrete topology. We also characterize which functions between topological spaces can be "realized" or computed with respect to different levels of representations. We will apply these results in the last section of this chapter to show how several variations of the identification in the limit paradigm can be modeled in terms of realizing functions between topological spaces.

We introduce the notion of a $\mathbf{\Sigma}^0_\alpha$-admissible representation and investigate their properties in the following section. In Section 4.2 we construct some specific examples of $\mathbf{\Sigma}^0_\alpha$-admissible representations for various spaces. In Section 4.3 we apply the results in this chapter to characterizing some variations of the identification in the limit model.

4.1 $\mathbf{\Sigma}^0_\alpha$-admissible representations

In this section we introduce the notion of a $\mathbf{\Sigma}^0_\alpha$-admissible representation. The basic idea is a generalization of an "admissible representation" that is often used in the field of computable analysis (see Weihrauch [59] and Schröder [48]). We have seen earlier that texts and informants are admissible representations of concept spaces with respect to specific topologies. In the hierarchy that we present next, these representations are called $\mathbf{\Sigma}^0_1$-admissible. In later sections, we will see that the representation $\lim_{\mathcal{H}} :\subseteq \mathcal{H} \to \mathcal{L}$, which maps a converging sequence of hypotheses to the concept represented by the limit of the sequence, is an example of $\mathbf{\Sigma}^0_2$-admissible representation with respect to the discrete topology on $\mathcal{L}$. By characterizing the representations of concept spaces and the function between concept spaces that can be "realized", we can reduce some problems in learning theory to the problem of "computing" functions between represented spaces. The results in this section were presented in [16].

4.1.1 The Borel hierarchy

In this subsection we define the Borel hierarchy on arbitrary topological spaces and introduce some basic properties. We will use a definition of the Borel hierarchy that differs from the classical definition (e.g., the definition in [29]) on non-metrizable spaces, but is more suitable for general topological spaces.

We let ω_1 denote the least uncountable ordinal and ω the set of natural numbers (or the first infinite ordinal), and for sets A and B we let $A \setminus B$ denote the subset of A of elements not in B.

Definition 4.1.1 *Let X be a topological space. For each ordinal α ($1 \leq \alpha < \omega_1$) we define $\mathbf{\Sigma}^0_\alpha(X)$ inductively as follows.*

1. *$\mathbf{\Sigma}^0_1(X)$ is the set of all open subsets of X.*

2. *For $\alpha > 1$, $\mathbf{\Sigma}^0_\alpha(X)$ is the set of all subsets A of X which can be expressed in the form*
$$A = \bigcup_{i \in \omega} B_i \setminus B'_i,$$
where for each i, B_i and B'_i are in $\mathbf{\Sigma}^0_{\beta_i}(X)$ for some $\beta_i < \alpha$.

We define $\mathbf{\Pi}^0_\alpha(X) = \{X \setminus A \mid A \in \mathbf{\Sigma}^0_\alpha(X)\}$, $\mathbf{\Delta}^0_\alpha(X) = \mathbf{\Sigma}^0_\alpha(X) \cap \mathbf{\Pi}^0_\alpha(X)$, and $\mathbf{B}(X) = \bigcup_{1 \leq \alpha < \omega_1} \mathbf{\Sigma}^0_\alpha(X)$. □

The above definition of the Borel hierarchy is equivalent to the definition that was used by Tang [58] in studying descriptive set theory on $\mathcal{P}(\omega)$ (the power set of the natural numbers with the Scott-topology), and more systematically investigated by Selivanov (see [52] for a survey of results and an extensive list of references).

The classical definition of the Borel hierarchy (which requires $B_i = X$ for all i in the second clause of Definition 4.1.1) is not suitable for non-metrizable spaces. For example, consider the Sierpinski space $\mathcal{S} = \{\perp, \top\}$ (where $\{\top\}$ is open, but $\{\perp\}$ is not). If we used the classical definition then $\mathbf{\Sigma}^0_{2n+1}(\mathcal{S})$ is the set of open subsets of $\mathcal{S}$ and $\mathbf{\Sigma}^0_{2n+2}(\mathcal{S})$ is the closed subsets, so $\mathbf{\Sigma}^0_{2n+1}(\mathcal{S}) \not\subseteq \mathbf{\Sigma}^0_{2n+2}(\mathcal{S})$ (for $0 \leq n < \omega$). The Borel hierarchy defined in Definition 4.1.1 is equivalent to the classical definition for all metrizable spaces, and behaves as we expect it should even for non-metrizable spaces.

In the following, X and Y will denote arbitrary topological spaces, unless stated otherwise. The following results are easily proven, and can also be found in [52].

Proposition 4.1.2 *For each α ($1 \leq \alpha < \omega_1$),*

1. *$\mathbf{\Sigma}^0_\alpha(X)$ is closed under countable unions and finite intersections,*
2. *$\mathbf{\Pi}^0_\alpha(X)$ is closed under countable intersections and finite unions,*
3. *$\mathbf{\Delta}^0_\alpha(X)$ is closed under finite unions, finite intersections, and complementation.*

Proposition 4.1.3 *If $\beta < \alpha$ then $\mathbf{\Sigma}^0_\beta(X) \cup \mathbf{\Pi}^0_\beta(X) \subseteq \mathbf{\Delta}^0_\alpha(X)$.*

Proposition 4.1.4 *For $\alpha > 2$, each $A \in \mathbf{\Sigma}^0_\alpha(X)$ can be expressed in the form*

$$A = \bigcup_{i \in \omega} B_i,$$

where for each i, B_i is in $\mathbf{\Pi}^0_{\beta_i}(X)$ for some $\beta_i < \alpha$.

Proposition 4.1.5 *If X is a metrizable space, then every $A \in \mathbf{\Sigma}^0_2(X)$ is equal to a countable union of closed sets.*

Proposition 4.1.6 *If X is a subspace of Y, then $\mathbf{\Sigma}^0_\alpha(X) = \{A \cap X \mid A \in \mathbf{\Sigma}^0_\alpha(Y)\}$ and $\mathbf{\Pi}^0_\alpha(X) = \{A \cap X \mid A \in \mathbf{\Pi}^0_\alpha(Y)\}$.*

Next we show how the complexity of some subsets of a topological space relate to the separation axioms the space satisfies.

Proposition 4.1.7 *For any countably based topological space X,*

1. *Every singleton set $\{x\} \subseteq X$ is in $\mathbf{\Pi}^0_2(X) \iff X$ is a T_0-space,*
2. *Every singleton set $\{x\} \subseteq X$ is in $\mathbf{\Delta}^0_2(X) \iff X$ is a T_D-space,*
3. *Every singleton set $\{x\} \subseteq X$ is in $\mathbf{\Pi}^0_1(X) \iff X$ is a T_1-space,*
4. *Every singleton set $\{x\} \subseteq X$ is in $\mathbf{\Delta}^0_1(X) \iff X$ is a discrete space.*

Therefore, if X is a countable space and $\mathcal{P}(X)$ is the power set of X, then

5. $\mathcal{P}(X) = \mathbf{\Delta}^0_3(X) \iff X$ *is a T_0-space,*
6. $\mathcal{P}(X) = \mathbf{\Delta}^0_2(X) \iff X$ *is a T_D-space,*
7. $\mathcal{P}(X) = \mathbf{\Delta}^0_1(X) \iff X$ *is a discrete space.*

Proof: We only prove (1), since the other claims are easy. First assume X is T_0 and $x \in X$. Let $\{U_i\}_{i\in I}$ be a countable neighborhood basis for x, and let $\overline{\{x\}}$ denote the closure of $\{x\}$. Then using the fact that X is T_0 it is easily seen that $\{x\} = \overline{\{x\}} \cap \bigcap_{i\in I} U_i$, which is clearly in $\mathbf{\Pi}^0_2(X)$. For the converse, if X is not T_0, then there are x and y that are contained by exactly the same open sets, hence the singleton $\{x\}$ is not even Borel. $\square$

Given an arbitrary countably-based topological space X, let

$$\Delta_X = \{\langle x, y\rangle \in X \times X \mid x = y\}.$$

The following proposition assumes the product topology on $X \times X$, which is the coarsest topology such that each projection function is continuous (see [30]). Note that this definition of product is equivalent (up to homeomorphism) with the construction of products for concept spaces given in Definition 3.1.12.

Proposition 4.1.8 *For an arbitrary countably-based topological space X,*

1. $\Delta_X \in \mathbf{\Pi}^0_2(X \times X) \iff X$ *is a T_0-space,*
2. $\Delta_X \in \mathbf{\Delta}^0_2(X \times X) \Longrightarrow X$ *is a T_D-space,*
3. $\Delta_X \in \mathbf{\Pi}^0_1(X \times X) \iff X$ *is a T_2-space,*
4. $\Delta_X \in \mathbf{\Delta}^0_1(X \times X) \iff X$ *is a discrete space.*

Furthermore, if X is countable, then

5. $\Delta_X \in \mathbf{\Delta}^0_2(X \times X) \iff X$ *is a T_D-space.*

Proof: (1). If X is not T_0, then it is clear that the diagonal of X is not a Borel set. For the converse, just note that

$$X \times X \setminus \Delta_X = \big(\bigcup_{i\in\omega} B_i \times (X \setminus B_i)\big) \cup \big(\bigcup_{i\in\omega} (X \setminus B_i) \times B_i\big),$$

where $\{B_i\}_{i\in\omega}$ is a countable basis for X.

(2). Assume that $\Delta_X = \bigcup_{i\in\omega} U_i \setminus V_i$ for U_i, V_i open in $X \times X$. Let x be any element of X. Then there is some $i \in \omega$ such that $(x, x) \in U_i$ and $(x, x) \notin V_i$. Therefore, there exists an open set $U \subseteq X$ such that $x \in U$ and $(x, x) \subseteq U \times U \subseteq U_i$. Let $y \neq x$ be an element of U. Then $(x, y) \in U_i$, and therefore (x, y) must be in V_i. Let V and V' be open subsets of X such that $(x, y) \in V \times V' \subseteq V_i$. Since $(x, x) \notin V_i$, $x \notin V'$. This implies that y is not in the closure of $\{x\}$, and since $y \in U$ was arbitrary, $\{x\} = U \cap cl(\{x\})$ is locally closed in X.

(3). This result is well known.

(4). Clearly $\Delta_X \in \mathbf{\Sigma}^0_1(X \times X)$ if X is discrete, since discrete countably-based spaces have only countably many points. For the converse, if $\Delta_X \in \mathbf{\Delta}^0_1(X \times X)$, then for each $x \in X$ there is an open set $U \subseteq X$ such that $\langle x, x\rangle \in U \times U \subseteq \Delta_X$. Assume that there is $y \in U$ distinct from x. Then $\langle x, y\rangle \in U \times U$, but $\langle x, y\rangle \notin \Delta_X$, a contradiction. Therefore, $U = \{x\}$. Since x was arbitrary, X is discrete.

(5). Assume X is a countable T_D-space. For $x \in X$, let U_x, V_x be open such that $\{x\} = U_x \setminus V_x$. Then $\Delta_X = \bigcup_{x \in X} U_x \setminus V_x \times U_x \setminus V_x$ is in $\mathbf{\Sigma}^0_2(X \times X)$ and therefore Δ_X is in $\mathbf{\Delta}^0_2(X \times X)$. □

4.1.2 $\mathbf{\Sigma}^0_\alpha$-measurable functions

In this subsection we will investigate some basic properties of $\mathbf{\Sigma}^0_\alpha$-measurable functions. Below, we will write $f \colon\subseteq X \to Y$ to indicate that f is a partial function from X to Y. The domain of definition of f will be denoted $dom(f)$. We say that $f \colon\subseteq X \to Y$ is *continuous* if and only if for every open $U \subseteq Y$, there is open $V \subseteq X$ such that $f^{-1}(U) = V \cap dom(f)$. In other words, $f \colon\subseteq X \to Y$ is continuous if and only if the total function $f \colon dom(f) \to Y$ is continuous with respect to the subspace topology on $dom(f)$.

Definition 4.1.9 *A function $f \colon X \to Y$ is* $\mathbf{\Sigma}^0_\alpha$-measurable *if and only if for every open $U \subseteq Y$, $f^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(X)$. A partial function $f \colon\subseteq X \to Y$ is said to be* $\mathbf{\Sigma}^0_\alpha$-measurable *if and only if for every open $U \subseteq Y$, there is $A \in \mathbf{\Sigma}^0_\alpha(X)$ such that $f^{-1}(U) = A \cap dom(f)$.* □

Equivalently, a partial function $f \colon\subseteq X \to Y$ is $\mathbf{\Sigma}^0_\alpha$-measurable if and only if for every open $U \subseteq Y$, $f^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(dom(f))$, where $dom(f)$ is given the relative topology.

For any fixed $\alpha > 1$, the $\mathbf{\Sigma}^0_\alpha$-measurable functions are not closed under composition. To characterize how composition behaves, we will need ordinal addition. Addition on ordinals is defined recursively as follows:

1. $\alpha + 0 = \alpha$
2. $\alpha + (\beta + 1) = (\alpha + \beta) + 1 =$ the successor of $\alpha + \beta$.
3. $\alpha + \lambda = \lim_{\beta<\lambda}(\alpha + \beta)$ for limit ordinal λ.

Note that ordinal addition is non-commutative. For example, $1+\omega = \omega \neq \omega+1$. Also note that if $\alpha < \beta$, then there is a unique ordinal γ such that $\alpha + \gamma = \beta$.

Composing with continuous functions does not change the level of a function. For that reason it would have been more convenient for our purposes to define the Borel hierarchy so that open sets and continuous functions were of level 0 (the additive identity for ordinals). To simplify the statement of some of the following theorems and proofs, we will often make use of the following "hat" notation, so that we can treat the Borel hierarchy as if we defined the open sets to be at level 0.

Definition 4.1.10 *For $0 \leq \alpha < \omega_1$, define*

$$\widehat{\alpha} = \begin{cases} \alpha + 1 & \text{if } \alpha < \omega \\ \alpha & \text{if } \alpha \geq \omega \end{cases}$$

□

Note that $\alpha < \beta \iff \widehat{\alpha} < \widehat{\beta}$ and $\widehat{\alpha+\beta} = \widehat{\alpha} + \beta$ hold for any countable ordinals α and β.

Lemma 4.1.11 *Let X and Y be countably based T_0-spaces. If $f\colon\subseteq X \to Y$ is $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable ($0 \le \alpha < \omega_1$) and $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(Y)$ ($0 \le \beta < \omega_1$), then $f^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(f))$.*

Proof: If $\beta = 0$ and $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(Y)$, then A is open so $f^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(f)) = \mathbf{\Sigma}^0_{\widehat{\alpha}}(dom(f))$ by definition of a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable function.

For $\beta > 1$, if $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(Y)$ then

$$A = \bigcup_{i\in\omega} B_i \setminus B_i',$$

where for each i, B_i and B_i' are in $\mathbf{\Sigma}^0_{\widehat{\beta_i}}(Y)$ for some $\beta_i < \beta$. Hence,

$$\begin{aligned} f^{-1}(A) &= f^{-1}(\bigcup_{i\in\omega} B_i \setminus B_i') \\ &= \bigcup_{i\in\omega} f^{-1}(B_i) \setminus f^{-1}(B_i'). \end{aligned}$$

By induction hypothesis $f^{-1}(B_i), f^{-1}(B_i') \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta_i}}(dom(f))$ for each $i \in \omega$. Since $\beta_i < \beta$, $\widehat{\alpha+\beta_i} = \widehat{\alpha} + \beta_i < \widehat{\alpha} + \beta = \widehat{\alpha+\beta}$ for all $i \in \omega$. Therefore, $f^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(f))$. $\square$

Theorem 4.1.12 *Let X, Y, and Z be countably based T_0-spaces, $f\colon\subseteq X \to Y$ a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable function ($0 \le \alpha < \omega_1$), and $g\colon\subseteq Y \to Z$ a $\mathbf{\Sigma}^0_{\widehat{\beta}}$-measurable function ($0 \le \beta < \omega_1$). Then $g \circ f\colon\subseteq X \to Z$ is $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-measurable.*

Proof: Let $U \subseteq Z$ be open. Then $g^{-1}(U) \in \mathbf{\Sigma}^0_{\widehat{\beta}}(dom(g))$ and, by restricting the domain of f to $dom(g \circ f)$ if necessary, from Lemma 4.1.11 it follows that $(g\circ f)^{-1}(U) = f^{-1}(g^{-1}(U)) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(g \circ f))$. $\square$

In particular, if f is $\mathbf{\Sigma}^0_2$-measurable and g is $\mathbf{\Sigma}^0_\omega$-measurable, then due to the non-commutativity of ordinal addition, $g \circ f$ is $\mathbf{\Sigma}^0_\omega$-measurable but $f \circ g$ is $\mathbf{\Sigma}^0_{\omega+1}$-measurable (assuming the compositions make sense).

The following is due to Wadge (this is Theorem 22.10 in [29]). We let ω^ω denote the Baire space.

Proposition 4.1.13 (Wadge) *If $B \subseteq \omega^\omega$ is in $\mathbf{B}(\omega^\omega)\setminus\mathbf{\Pi}^0_{\widehat{\alpha}}(\omega^\omega)$ ($0 \le \alpha < \omega_1$), then for any $A \in \mathbf{\Sigma}^0_{\widehat{\alpha}}(\omega^\omega)$ there is continuous total $f\colon \omega^\omega \to \omega^\omega$ such that $A = f^{-1}(B)$.* $\square$

We will need the following generalization of Wadge's results that characterize reductions using measurable functions.

Theorem 4.1.14 *For $0 \le \alpha < \omega_1$ and $0 \le \beta < \omega_1$, if $B \in \mathbf{B}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\beta}}(\omega^\omega)$, then for any $A \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$ there exists a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable total function $f\colon \omega^\omega \to \omega^\omega$ such that $A = f^{-1}(B)$.*

Proof: Fix α and choose $\mathcal{U}_0 \in \mathbf{\Sigma}^0_{\widehat{\alpha}}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\alpha}}(\omega^\omega)$, and let $\{U_i\}_{i\in\omega}$ be $\mathbf{\Delta}^0_{\widehat{\alpha}}$ sets such that $\mathcal{U}_0 = \bigcup_{i\in\omega} U_i$. Define $V_i = \{y \in \omega^\omega \,|\, y(i) = 1\}$ and $\mathcal{V}_0 = \bigcup_{i\in\omega} V_i$. For $1 \le \beta < \omega_1$, let $\{\eta^\beta_n\}_{n\in\omega}$ be a non-decreasing countable sequence of ordinals such that $\beta = \sup_{n\in\omega}(\eta^\beta_n + 1)$. Define $\mathcal{U}_\beta \subseteq \omega^\omega$ and $\mathcal{V}_\beta \subseteq \omega^\omega$ as

$$\begin{aligned} y \in \mathcal{U}_\beta &\iff (\exists n \in \omega)[(y)_n \notin \mathcal{U}_{\eta^\beta_n}],\\ y \in \mathcal{V}_\beta &\iff (\exists n \in \omega)[(y)_n \notin \mathcal{V}_{\eta^\beta_n}], \end{aligned}$$

where $(y)_n(m) = y(\langle n, m\rangle)$ for some bijection $\langle\cdot,\cdot\rangle\colon \omega\times\omega \to \omega$.

We first prove the claim that for $0 \le \beta < \omega_1$, $\mathcal{U}_\beta \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$ and $\mathcal{V}_\beta \in \mathbf{\Sigma}^0_{\widehat{\beta}}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\beta}}(\omega^\omega)$. The proof for $\mathcal{U}_\beta$ and $\mathcal{V}_\beta$ are essentially the same, so we will only prove the claim for $\mathcal{U}_\beta$. It is immediate for $\beta = 0$, so assume $\beta \ge 1$ and that the hypothesis holds for η^β_n for all $n \in \omega$. Since $y \mapsto (y)_n$ is a continuous mapping, it is clear that $\mathcal{U}_\beta \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$. Choose any $A \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$, then A can be expressed as $A = \bigcup_{i\in\omega} W_i$, where $W_i \in \mathbf{\Pi}^0_{\widehat{\gamma_i}}(\omega^\omega)$ ($\gamma_i < \alpha + \beta$) for each $i \in \omega$. Define $p(0) = \min\{n \in \omega \,|\, \gamma_0 \le \alpha + \eta^\beta_n\}$ and for $i \ge 0$ define $p(i+1) = \min\{n \in \omega \,|\, n > p(i) \text{ and } \gamma_{i+1} \le \alpha + \eta^\beta_n\}$. Clearly $p(i)$ is defined for all $i \in \omega$ because $\beta = \sup_{n\in\omega}(\eta^\beta_n + 1)$ and $\{\eta^\beta_n\}_{n\in\omega}$ is non-decreasing. It is also obvious that p is injective. By induction hypothesis $\mathcal{U}_{\eta^\beta_n}$ is not $\mathbf{\Pi}^0_1$ hence non-empty for all $n \in \omega$, so let u_n be an arbitrary element of $\mathcal{U}_{\eta^\beta_n}$. We define continuous functions $q_n\colon \omega^\omega \to \omega^\omega$ for $n \in \omega$ as follows. If $n \notin range(p)$ then define $q_n(y) = u_n$ for all $y \in \omega^\omega$. If $n \in range(p)$, then $n = p(i)$ for a unique $i \in \omega$, so let q_n be continuous such that $W_i = q_n^{-1}(\omega^\omega \setminus \mathcal{U}_{\eta^\beta_n})$ (such a q_n exists by induction hypothesis). Now define $q\colon \omega^\omega \to \omega^\omega$ so that $(q(y))_n = q_n(y)$. Then q is continuous and

$$\begin{aligned} q(y) \in \mathcal{U}_\beta &\iff (\exists n \in \omega)[(q(y))_n \notin \mathcal{U}_{\eta^\beta_n}]\\ &\iff (\exists n \in \omega)[q_n(y) \notin \mathcal{U}_{\eta^\beta_n}]\\ &\iff (\exists i \in \omega)[q_{p(i)}(y) \notin \mathcal{U}_{\eta^\beta_{p(i)}}]\\ &\iff (\exists i \in \omega)[y \in W_i] \end{aligned}$$

Thus, $A = q^{-1}(\mathcal{U}_\beta)$, which implies $\mathcal{U}_\beta \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$, and the proof of our claim is complete.

Next, define $g\colon \omega^\omega \to \omega^\omega$ so that

$$g(y)(n) = \begin{cases} 1 & \text{if } y \in U_n \\ 0 & \text{otherwise.} \end{cases}$$

Then $g(y) \in V_n \iff y \in U_n$ which implies $g(y) \in \mathcal{V}_0 \iff y \in \mathcal{U}_0$, and so by induction $g(y) \in \mathcal{V}_\beta \iff y \in \mathcal{U}_\beta$ for $0 \le \beta < \omega_1$.

To see that g is $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable, just note that the range of g is a subset of 2^ω, that $\{V_n \cap 2^\omega \,|\, n \in \omega\} \cup \{(\omega^\omega \setminus V_n) \cap 2^\omega \,|\, n \in \omega\}$ is a subbasis for the topology on 2^ω, and that the preimage under g of any of these subbasic sets is in $\mathbf{\Delta}^0_{\widehat{\alpha}}(\omega^\omega)$.

Finally, for $A \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$ and $B \in \mathbf{B}(\omega^\omega) \setminus \mathbf{\Pi}^0_{\widehat{\beta}}(\omega^\omega)$, by Proposition 4.1.13 there are continuous functions $f', f''\colon \omega^\omega \to \omega^\omega$ such that $A = (f')^{-1}(\mathcal{U}_\beta)$ and $\mathcal{V}_\beta = (f'')^{-1}(B)$. So $f = f'' \circ g \circ f'$ is $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable and $A = f^{-1}(B)$. $\square$

4.1.3 Existence of $\mathbf{\Sigma}^0_\alpha$-admissible representations

The goal of this subsection is to show that every countably based T_0-space has a $\mathbf{\Sigma}^0_\alpha$-admissible representation for $1 \leq \alpha < \omega_1$ (Theorem 4.1.22 below). We also show the complexity of converting between representations of different levels (Theorem 4.1.23), and consider representations of representations of a space (Corollary 4.1.24), which is a generalization of Ziegler's "jump" of a representation [62].

Definition 4.1.15 *A $\mathbf{\Sigma}^0_\alpha$-*admissible representation *of a topological space X is a $\mathbf{\Sigma}^0_\alpha$-measurable partial function $\rho\colon\subseteq \omega^\omega \to X$ such that for every $\mathbf{\Sigma}^0_\alpha$-measurable partial function $f\colon\subseteq \omega^\omega \to X$, there exists continuous $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f = \rho \circ g$.*

$$\begin{tikzcd} \omega^\omega \arrow[r, "g"] \arrow[dr, "f"'] & \omega^\omega \arrow[d, "\rho"] \\ & X \end{tikzcd}$$

□

Note that the above definition implies that $\mathbf{\Sigma}^0_\alpha$-admissible representations are always surjective. Clearly, a $\mathbf{\Sigma}^0_1$-admissible representation is equivalent to what is usually called an "admissible representation" in the computable analysis literature (see, e.g., [59] and [48]). The above definition applies to arbitrary topological spaces, but most of our results will focus on countably based spaces.

We let $\mathcal{S}$ denote the Sierpinski space, which has only two points $\top$ and $\bot$, and where $\{\top\}$ is open but $\{\bot\}$ is not open.

Proposition 4.1.16 *Let $A \in \mathbf{\Sigma}^0_\alpha(\omega^\omega) \setminus \mathbf{\Pi}^0_\alpha(\omega^\omega)$ and define $\rho\colon \omega^\omega \to \mathcal{S}$ so that $\rho(y) = \top$ if $y \in A$ and $\rho(y) = \bot$ if $y \notin A$. Then ρ is a $\mathbf{\Sigma}^0_\alpha$-admissible representation for $\mathcal{S}$.*

Proof: It is clear that ρ is $\mathbf{\Sigma}^0_\alpha$-measurable. Let $f\colon\subseteq \omega^\omega \to \mathcal{S}$ be a $\mathbf{\Sigma}^0_\alpha$-measurable partial function. Then $f^{-1}(\{\top\}) \in \mathbf{\Sigma}^0_\alpha(dom(f))$, so there is $B \in \mathbf{\Sigma}^0_\alpha(\omega^\omega)$ such that $f^{-1}(\{\top\}) = B \cap dom(f)$. From Proposition 4.1.13 there is continuous $g\colon \omega^\omega \to \omega^\omega$ such that $g^{-1}(A) = B$. Then for all $y \in dom(f)$, $f(y) = \top \iff g(y) \in A \iff \rho(g(y)) = \top$. Hence, by restricting the domain of g if necessary, $f = \rho \circ g$. □

Corollary 4.1.17 *For $0 \leq \alpha < \omega_1$ and $0 \leq \beta < \omega_1$, if $\rho_{\alpha+\beta}\colon\subseteq \omega^\omega \to \mathcal{S}$ is a $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible representation of $\mathcal{S}$ and $\rho_\beta\colon\subseteq \omega^\omega \to \mathcal{S}$ is a $\mathbf{\Sigma}^0_{\widehat{\beta}}$-admissible representation of $\mathcal{S}$, then there exists a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable function $f\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\rho_{\alpha+\beta} = \rho_\beta \circ f$.*

Proof: Immediate from Theorem 4.1.14 and Proposition 4.1.16. □

Proposition 4.1.18 *If X is a subspace of Y and $\rho\colon\subseteq \omega^\omega \to Y$ is a $\mathbf{\Sigma}^0_\alpha$-admissible representation of Y, then $\rho_X\colon\subseteq \omega^\omega \to X$ defined as the restriction of ρ to $dom(\rho_X) = \rho^{-1}(X)$, is a $\mathbf{\Sigma}^0_\alpha$-admissible representation of X.*

Proof: Let $f\colon\subseteq \omega^\omega \to X$ be a $\mathbf{\Sigma}^0_\alpha$-measurable partial function. Since $X \subseteq Y$ and ρ is $\mathbf{\Sigma}^0_\alpha$-admissible for Y, there is continuous $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f = \rho \circ g$. Since $range(f) \subseteq X$ we can assume that $range(g) \subseteq \rho^{-1}(X) = dom(\rho_X)$, hence $f = \rho_X \circ g$. □

Proposition 4.1.19 *If $\{X_i\}_{i\in\omega}$ and $\{Y_i\}_{i\in\omega}$ are all countably based T_0-spaces, and for each i $f_i\colon\subseteq X_i \to Y_i$ is $\mathbf{\Sigma}^0_\alpha$-measurable ($1 \leq \alpha < \omega_1$), then $f^\omega\colon\subseteq \prod X_i \to \prod Y_i$ is $\mathbf{\Sigma}^0_\alpha$-measurable, where $\prod X_i$ and $\prod Y_i$ are given the product topologies and f^ω is defined so that $f^\omega(\xi)(i) = f_i(\xi(i))$.*

Proof: First note that $dom(f^\omega) = \prod dom(f_i)$. $\prod Y_i$ has as a countable subbasis sets of the form $U = \prod U_i$ where U_n is a basic open subset from some set countable basis for Y_n for some $n \in \omega$ and $U_i = Y_i$ for all $i \neq n$. Then $(f^\omega)^{-1}(U) = \prod f_i^{-1}(U_i) = \pi_n^{-1}(f_n^{-1}(U_n)) \cap dom(f^\omega)$, where $\pi_n\colon \prod X_i \to X_n$ is the n-th projection. Hence, $(f^\omega)^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(dom(f^\omega))$. Since $\mathbf{\Sigma}^0_\alpha$ is closed under finite intersections and countable unions, it follows that f^ω is $\mathbf{\Sigma}^0_\alpha$-measurable. □

Proposition 4.1.20 *If X and $\{Y_i\}_{i\in\omega}$ are all countably based T_0-spaces, and for each i $f_i\colon\subseteq X \to Y_i$ is $\mathbf{\Sigma}^0_\alpha$-measurable ($1 \leq \alpha < \omega_1$), then $\langle f_i\rangle_{i\in\omega}\colon\subseteq X \to \prod Y_i$ is $\mathbf{\Sigma}^0_\alpha$-measurable, where $\langle f_i\rangle_{i\in\omega}$ is defined so that $\langle f_i\rangle_{i\in\omega}(x)(i) = f_i(x)$ for all $x \in X$.*

Proof: The proof is similar to the proof of Proposition 4.1.19. □

If $X_i = X$ for all $i \in \omega$, then we will often abbreviate $\prod X_i$ to X^ω. For the following proposition, let $\phi\colon \omega^\omega \to (\omega^\omega)^\omega$ be a homeomorphism.

Proposition 4.1.21 *Let X_i be a countably based T_0-space and $\rho_i\colon\subseteq \omega^\omega \to X_i$ a $\mathbf{\Sigma}^0_\alpha$-admissible representation for X_i ($i \in \omega$). Then $\rho^\omega \circ \phi$ is a $\mathbf{\Sigma}^0_\alpha$-admissible representation for $\prod X_i$.*

Proof: The proof that $\rho^\omega \circ \phi$ is $\mathbf{\Sigma}^0_\alpha$-measurable follows from Proposition 4.1.19.

Let $f\colon\subseteq \omega^\omega \to \prod X_i$ be a $\mathbf{\Sigma}^0_\alpha$-measurable partial function. By the $\mathbf{\Sigma}^0_\alpha$-admissibility of $\rho_i\colon\subseteq \omega^\omega \to X_i$, for $i \in \omega$ there is continuous $g_i\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\pi_i \circ f = \rho_i \circ g_i$, where $\pi_i\colon \prod X_i \to X$ is the i-th projection. Since π_i is a total function, we must have that $dom(f) = dom(\pi_i \circ f) \subseteq dom(g_i)$ for all $i \in \omega$. Define $g\colon\subseteq \omega^\omega \to (\omega^\omega)^\omega$ so that $g(\xi)(i) = g_i(\xi)$. Then $dom(f) \subseteq dom(g)$ and

$$\begin{aligned} \rho^\omega(g(\xi))(i) &= \rho_i(g(\xi)(i)) \\ &= \rho_i(g_i(\xi)) \\ &= \pi_i(f(\xi)) \\ &= f(\xi)(i), \end{aligned}$$

so $f = \rho^\omega \circ g$. Define $h\colon\subseteq \omega^\omega \to \omega^\omega$ as $h = \phi^{-1} \circ g$. Clearly, h is continuous and $f = \rho^\omega \circ g = \rho^\omega \circ \phi \circ h$. □

Theorem 4.1.22 *For every countably based T_0-space X and every α ($1 \leq \alpha < \omega_1$), there exists a $\mathbf{\Sigma}^0_\alpha$-admissible representation of X.*

Proof: From Propositions 4.1.16 and 4.1.21 we can see that there exists a $\mathbf{\Sigma}^0_\alpha$-admissible representation for $\mathcal{S}^\omega$. It then follows from Proposition 4.1.18 that every subspace of $\mathcal{S}^\omega$ has a $\mathbf{\Sigma}^0_\alpha$-admissible representation.

Let $\{U_i\}_{i\in\omega}$ be a countable basis for X. Define $f\colon X \to \mathcal{S}^\omega$ so that $f(x)(i) = \top$ if $x \in U_i$ and $f(x)(i) = \bot$, otherwise. Then f is easily seen to be a topological embedding of X into $\mathcal{S}^\omega$. Therefore, since X is homeomorphic to a subspace of S^ω, X has a $\mathbf{\Sigma}^0_\alpha$-admissible representation. □

Theorem 4.1.23 (Reductions between representations) *Assume X is a countably based T_0-space. For $0 \le \alpha < \omega_1$ and $0 \le \beta < \omega_1$, if $\rho_{\alpha+\beta}\colon\subseteq \omega^\omega \to X$ is a $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible representation of X and $\rho_\beta\colon\subseteq \omega^\omega \to X$ is a $\mathbf{\Sigma}^0_{\widehat{\beta}}$-admissible representation of X, then there exists a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable function $f\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\rho_{\alpha+\beta} = \rho_\beta \circ f$.*

Proof: We prove the case for $X = \mathcal{S}^\omega$, from which the general case follows. Let $\delta_{\alpha+\beta}\colon\subseteq \omega^\omega \to \mathcal{S}$ be a $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible representation of $\mathcal{S}$ and $\delta_\beta\colon\subseteq \omega^\omega \to \mathcal{S}$ is a $\mathbf{\Sigma}^0_{\widehat{\beta}}$-admissible representation of $\mathcal{S}$, and let $f'\colon\subseteq \omega^\omega \to \omega^\omega$ be a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable function such that $\delta_{\alpha+\beta} = \delta_\beta \circ f$.

Define $f^\omega\colon\subseteq (\omega^\omega)^\omega \to (\omega^\omega)^\omega$ so that $f^\omega(\xi)(i) = f'(\xi(i))$ and define $\delta^\omega_{\alpha+\beta}\colon\subseteq \omega^\omega \to \mathcal{S}^\omega$ and $\delta^\omega_\beta\colon\subseteq \omega^\omega \to \mathcal{S}^\omega$ similarly. Then

$$\begin{aligned}
\delta^\omega_\beta(f^\omega(\xi))(i) &= \delta_\beta(f^\omega(\xi)(i)) \\
&= \delta_\beta(f'(\xi(i))) \\
&= \delta_{\alpha+\beta}(\xi(i)) \\
&= \delta^\omega_{\alpha+\beta}(\xi)(i),
\end{aligned}$$

which implies $\delta^\omega_{\alpha+\beta} = \delta^\omega_\beta \circ f^\omega$.

Let $\phi\colon \omega^\omega \to (\omega^\omega)^\omega$ be a homeomorphism. We have that $\delta^\omega_{\alpha+\beta} \circ \phi$ is a $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible representation of $\mathcal{S}^\omega$ by Proposition 4.1.21, and also that $\delta^\omega_\beta \circ \phi$ is $\mathbf{\Sigma}^0_{\widehat{\beta}}$-measurable. Therefore, there are continuous $g, h \subseteq \omega^\omega \to \omega^\omega$ such that $\rho_{\alpha+\beta} = \delta^\omega_{\alpha+\beta} \circ \phi \circ g$ and $\delta^\omega_\beta \circ \phi = \rho_\beta \circ h$. Define $f\colon\subseteq \omega^\omega \to \omega^\omega$ as $f = h \circ \phi^{-1} \circ f^\omega \circ \phi \circ g$. Then

$$\begin{aligned}
\rho_{\alpha+\beta} &= \delta^\omega_{\alpha+\beta} \circ \phi \circ g \\
&= \delta^\omega_\beta \circ f^\omega \circ \phi \circ g \\
&= \delta^\omega_\beta \circ \phi \circ \phi^{-1} \circ f^\omega \circ \phi \circ g \\
&= \rho_\beta \circ h \circ \phi^{-1} \circ f^\omega \circ \phi \circ g \\
&= \rho_\beta \circ f.
\end{aligned}$$

The case that X is a subspace of $\mathcal{S}^\omega$ is handled by restricting the domain of f^ω, $\delta^\omega_{\alpha+\beta}$ and δ^ω_β in the above argument. □

Corollary 4.1.24 (Representations of representations) *Let X be a countably based T_0-space, $\rho_\beta\colon\subseteq \omega^\omega \to X$ a $\mathbf{\Sigma}^0_{\widehat{\beta}}$-admissible representation of X, and $\rho_\alpha\colon\subseteq \omega^\omega \to dom(\rho_\beta)$ a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation of $dom(\rho_\beta)$, ($0 \le \alpha < \omega_1$, $0 \le \beta < \omega_1$). Then $\rho_\beta \circ \rho_\alpha\colon\subseteq \omega^\omega \to X$ is a $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible representation of X.*

Proof: First note that $\rho_\beta \circ \rho_\alpha$ is $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-measurable by Theorem 4.1.12. Let $\rho\colon\subseteq \omega^\omega \to X$ be a $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible representation of X. By Theorem 4.1.23, there is a $\mathbf{\Sigma}^0_\alpha$-measurable $f\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\rho = \rho_\beta \circ f$. We can assume without loss of generality that $range(f) \subseteq dom(\rho_\beta)$, and so by the $\mathbf{\Sigma}^0_\alpha$-admissibility of ρ_α there is a continuous $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f = \rho_\alpha \circ g$. It follows that g is a continuous reduction of ρ to $\rho_\beta \circ \rho_\alpha$, thus $\rho_\beta \circ \rho_\alpha$ is $\mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}$-admissible. $\square$

Let $\iota'\colon\subseteq \omega^\omega \to \omega^\omega$ be a $\mathbf{\Sigma}^0_2$-admissible representation of ω^ω. By the above theorem, if $\rho\colon\subseteq \omega^\omega \to X$ is a $\mathbf{\Sigma}^0_\beta$-admissible representation ($1 \leq \beta < \omega$) of a countably based T_0-space X, then $\rho \circ \iota'$ is a $\mathbf{\Sigma}^0_{\beta+1}$-admissible representation of X. This corresponds to Ziegler's "jump" of a representation [62]. However, it should be noted that if ρ is $\mathbf{\Sigma}^0_\beta$-admissible for $\beta \geq \omega$, then $\rho \circ \iota'$ is still $\mathbf{\Sigma}^0_\beta$-measurable and thus *not* $\mathbf{\Sigma}^0_{\beta+1}$-admissible.

4.1.4 Properties of $\mathbf{\Sigma}^0_\alpha$-admissible representations

The main purpose of this subsection is to relate the Borel complexity of a subset of a space with the complexity of the preimage of the subset under a $\mathbf{\Sigma}^0_\alpha$-admissible representation. These results will be useful in the following subsection where we characterize the functions that are realizable with respect to these representations.

The first step is Corollary 4.1.31, which shows that the complexity of a subset of a countably based T_0-space is exactly the complexity of the preimage of the subset under a $\mathbf{\Sigma}^0_1$-admissible representation. This result follows almost immediately from a recent result by J. Saint Raymond (Lemma 17 in [43]), but as the original statement was for metrizable spaces, we will reproduce the proof here (Proposition 4.1.30 below) so that the reader can easily verify that Saint Raymond's argument applies to more general spaces when we define the Borel hierarchy according to Definition 4.1.1.

The following definitions and propositions concerning meager sets can be found in Kechris [29] Chapter I Section 8. Recall that a topological space is a *polish space* if and only if it is countably based and completely metrizable.

Definition 4.1.25 *A subset A of a topological space X is* nowhere dense *if its closure has empty interior. A set $A \subseteq X$ is* meager *if it is equal to the union of a countable collection of nowhere dense sets. A set $A \subseteq X$ is* comeager *if its complement is meager.* $\square$

Note that a subset of a meager set is meager.

Proposition 4.1.26 *Every non-empty open subset of a Polish space is non-meager.* $\square$

Definition 4.1.27 *A subset A of a topological space X has the* Baire Property *if $(A \setminus U) \cup (U \setminus A)$ is meager for some open $U \subseteq X$.* $\square$

Proposition 4.1.28 *Every Borel subset of a Polish space has the Baire Property.* $\square$

Proposition 4.1.29 *If $A \subseteq X$ is non-meager and has the Baire Property, then there is a non-empty open $U \subseteq X$ in which $U \setminus A$ is meager in X.* $\square$

Proposition 4.1.30 (Saint-Raymond [43]) *Let $\phi\colon X \to Y$ be an open continuous surjective total function with Polish fibers (i.e. $\phi^{-1}(y)$ is Polish for each $y \in Y$), where X is a separable metric space and Y is a countably based T_0 topological space. Then for every $A \subseteq Y$ and $1 \le \alpha < \omega_1$, $A \in \mathbf{\Sigma}^0_\alpha(Y)$ if and only if $\phi^{-1}(A) \in \mathbf{\Sigma}^0_\alpha(X)$.*

Proof: The "only if" part holds because ϕ is continuous, so we only need to prove the "if" part.

For all Borel subsets B of X, define

$$\begin{aligned} N_0(B) &= \{y \in Y \mid B \cap \phi^{-1}(y) \text{ is non-meager in } \phi^{-1}(y)\} \\ N_1(B) &= \{y \in Y \mid B \cap \phi^{-1}(y) \text{ is comeager in } \phi^{-1}(y)\} \end{aligned}$$

Let $\{U_k\}_{k\in\omega}$ be a countable basis for X. We show that for $1 \le \alpha < \omega_1$, if $B \in \mathbf{\Sigma}^0_\alpha(X)$ then $N_0(B) \in \mathbf{\Sigma}^0_\alpha(Y)$ and if $B \in \mathbf{\Pi}^0_\alpha(X)$ then $N_1(B) \in \mathbf{\Pi}^0_\alpha(Y)$. Note that

$$\begin{aligned} N_1(X \setminus B) &= \{y \in Y \mid (X \setminus B) \cap \phi^{-1}(y) \text{ is comeager in } \phi^{-1}(y)\} \\ &= \{y \in Y \mid B \cap \phi^{-1}(y) \text{ is meager in } \phi^{-1}(y)\} \\ &= Y \setminus \{y \in Y \mid B \cap \phi^{-1}(y) \text{ is non-meager in } \phi^{-1}(y)\} \\ &= Y \setminus N_0(B). \end{aligned}$$

Therefore, the statements we wish to prove for N_0 and N_1 are equivalent. For the case $\alpha = 1$, B is open and since $\phi^{-1}(y)$ is Polish for every $y \in Y$, by Proposition 4.1.26 $B \cap \phi^{-1}(y)$ is non-meager in $\phi^{-1}(y)$ if and only if $B \cap \phi^{-1}(y)$ is non-empty. Therefore, $N_0(B) = \phi(B)$ is open in Y by our assumption that ϕ is an open map.

For $\alpha > 1$, $B = \bigcup_{i\in\omega} B_i$, where each $B_i \in \mathbf{\Pi}^0_{\beta_i}(X)$ for some $\beta_i < \alpha$. For any $y \in Y$, $B \cap \phi^{-1}(y)$ is non-meager in $\phi^{-1}(y)$ if and only if $B_i \cap \phi^{-1}(y)$ is non-meager for some $i \in \omega$, because the countable union of meager sets are meager. If $B_i \cap \phi^{-1}(y)$ is non-meager, then by Propositions 4.1.28 and 4.1.29 there is an open $U \subseteq X$ such that $U \cap \phi^{-1}(y)$ is non-empty and $(U \setminus B_i) \cap \phi^{-1}(y)$ is meager in $\phi^{-1}(y)$. Let $U_k \subseteq U$ be a basic open set such that $U_k \cap \phi^{-1}(y)$ is non-empty, then $(U_k \setminus B_i) \cap \phi^{-1}(y)$ is meager in $\phi^{-1}(y)$, since it is a subset of $(U \setminus B_i) \cap \phi^{-1}(y)$. Since $(B_i \cup (X \setminus U_k)) = (X \setminus (U_k \setminus B_i))$, $(B_i \cup (X \setminus U_k)) \cap \phi^{-1}(y)$ is comeager in $\phi^{-1}(y)$.

On the other hand, if $y \in \phi(U_k)$ and $(B_i \cup (X \setminus U_k)) \cap \phi^{-1}(y)$ is comeager in $\phi^{-1}(y)$, then $(U_k \setminus B_i) \cap \phi^{-1}(y)$ is meager in $\phi^{-1}(y)$. $U_k \cap \phi^{-1}(y)$ is non-empty, and therefore non-meager in $\phi^{-1}(y)$ by Proposition 4.1.26. Since $U_k \cap \phi^{-1}(y) = [(U_k \setminus B_i) \cap \phi^{-1}(y)] \cup [(U_k \cap B_i) \cap \phi^{-1}(y)]$, $(U_k \cap B_i) \cap \phi^{-1}(y)$ is non-meager in $\phi^{-1}(y)$ (otherwise $U_k \cap \phi^{-1}(y)$ would be equal to the union of two meager sets and be meager). It follows that $B \cap \phi^{-1}(y)$ is non-meager in $\phi^{-1}(y)$.

Therefore,

$$N_0(B) = \bigcup_{i\in\omega} \bigcup_{k\in\omega} N_1(B_i \cup (X \setminus U_k)) \cap \phi(U_k).$$

Since $B_i \in \mathbf{\Pi}^0_{\beta_i}(X)$ and $(X \setminus U_k)$ is closed, $B_i \cup (X \setminus U_k)$ is in $\mathbf{\Pi}^0_{\beta_i}(X)$. By the induction hypothesis, $N_1(B_i \cup (X \setminus U_k)) \in \mathbf{\Pi}^0_{\beta_i}(Y)$. Hence, $N_1(B_i \cup (X \setminus U_k)) \cap \phi(U_k) \in \mathbf{\Sigma}^0_\alpha(Y)$ for all i and k, and it follows that $N_0(B) \in \mathbf{\Sigma}^0_\alpha(Y)$, which completes the induction argument.

Let $A \subseteq Y$ and assume that $\phi^{-1}(A) \in \mathbf{\Sigma}^0_\alpha(X)$. Then $A = N_0(\phi^{-1}(A)) \in \mathbf{\Sigma}^0_\alpha(Y)$. □

We can now prove that the Borel hierarchy is preserved under $\mathbf{\Sigma}^0_1$-admissible representations of countably based T_0-spaces.

Corollary 4.1.31 *Let X be a countably based T_0-space and $\rho\colon\subseteq \omega^\omega \to X$ a $\mathbf{\Sigma}^0_1$-admissible representation of X. Then for $1 \leq \alpha < \omega_1$, $A \in \mathbf{\Sigma}^0_\alpha(X)$ if and only if $\rho^{-1}(A) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$.*

Proof: Define $\delta\colon \omega^\omega \to \mathcal{P}(\omega)$ such that $\delta(\xi) = \{n-1 \mid \exists j(\xi(j) = n \neq 0)\}$, where $\mathcal{P}(\omega)$ is the power set of ω. Note that δ is a continuous (total) surjection with respect to the Scott-topology on $\mathcal{P}(\omega)$, and furthermore δ is an open map.

For every $x \in \mathcal{P}(\omega)$, the singleton $\{x\}$ is in $\mathbf{\Pi}^0_2(\mathcal{P}(\omega))$ by Proposition 4.1.7. Since δ is continuous and total, $\delta^{-1}(x)$ is in $\mathbf{\Pi}^0_2(\omega^\omega)$, and hence a G_δ set. Every G_δ subspace of a Polish space is Polish (Theorem 3.11 in [29]). Therefore, $\delta^{-1}(x)$ is a Polish subspace of ω^ω for every $x \in \mathcal{P}(\omega)$.

Since X is a countably based T_0-space, we can assume without loss of generality that X is a subspace of $\mathcal{P}(\omega)$. Let $\phi\colon \delta^{-1}(X) \to X$ be the restriction of δ to $\delta^{-1}(X)$. It follows that ϕ is an open continuous surjective total function with Polish fibers, and from Proposition 4.1.30 that for $1 \leq \alpha < \omega_1$ and $A \subseteq X$, $A \in \mathbf{\Sigma}^0_\alpha(X)$ if and only if $\phi^{-1}(A) \in \mathbf{\Sigma}^0_\alpha(\phi^{-1}(X))$.

Let $f\colon\subseteq \omega^\omega \to \omega^\omega$ be continuous such that $\phi = \rho \circ f$, which exists because ρ is $\mathbf{\Sigma}^0_1$-admissible. Let $A \subseteq X$ be such that $\rho^{-1}(A) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$. Then $f^{-1}(\rho^{-1}(A)) \in \mathbf{\Sigma}^0_\alpha(dom(f))$ since f is continuous. Therefore, $\phi^{-1}(A) = f^{-1}(\rho^{-1}(A))$ is in $\mathbf{\Sigma}^0_\alpha(\phi^{-1}(X))$ because $\phi^{-1}(X) \subseteq dom(f)$. It follows that $A \in \mathbf{\Sigma}^0_\alpha(X)$. □

Our next goal is to generalize Corollary 4.1.31 to some $\mathbf{\Sigma}^0_\alpha$-admissible representations.

Let ω^* have as a base set $\omega \cup \{\infty\}$ and the topology so that U is open if and only if either $\infty \notin U$ or else U is cofinite (i.e., for some $m < \omega$, $n \in U$ for all $n \geq m$). Note that ω^* is the one-point compactification of ω with the discrete topology, hence the notation (which should not be confused with the set of finite strings of natural numbers).

Lemma 4.1.32 *Let $\rho\colon\subseteq \omega^\omega \to \omega^*$ be $\mathbf{\Sigma}^0_\alpha$-admissible ($1 \leq \alpha < \omega_1$). Then $S \subseteq \omega^*$ is open if and only if $\rho^{-1}(S) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$.*

Proof: Since ρ is $\mathbf{\Sigma}^0_\alpha$-measurable, if S is an open subset of ω^* then $\rho^{-1}(S) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$.

Assume, for a contradiction, that $\rho^{-1}(S) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$ but S is not open in ω^*. Then ∞ must be in S, so $\omega^* \setminus S$ is open and it follows that $\rho^{-1}(S) \in \mathbf{\Delta}^0_\alpha(dom(\rho))$ because ρ is $\mathbf{\Sigma}^0_\alpha$-measurable.

First assume that S is finite, say $S = \{n_0, n_1, \ldots, n_k, \infty\}$. First note that $\rho^{-1}(\infty) \in \mathbf{\Delta}^0_\alpha(dom(\rho))$ because $\rho^{-1}(\infty) = \rho^{-1}(S) \setminus \rho^{-1}(\{n_0, n_1, \ldots, n_k\})$ and $\rho^{-1}(\{n_0, n_1, \ldots, n_k\}) \in \mathbf{\Delta}^0_\alpha(dom(\rho))$. Now let $X \in \mathbf{\Sigma}^0_\alpha(\omega^\omega) \setminus \mathbf{\Pi}^0_\alpha(\omega^\omega)$. Set $X = \bigcup_{i\in\omega} W_i$, where $W_i \in \mathbf{\Delta}^0_\alpha(\omega^\omega)$ for each $i \in \omega$. Define $f\colon \omega^\omega \to \omega^*$ so that for $\xi \in \omega^\omega$,

$$f(\xi) = \begin{cases} \min\{i \in \omega \mid \xi \in W_i\} & \text{if } \xi \in X \\ \infty & \text{if } \xi \notin X \end{cases}$$

For any open $U \subseteq \omega^*$, if $\infty \notin U$, then $f^{-1}(U)$ equals the countable union of some $\mathbf{\Delta}^0_\alpha$ sets (for example, $f^{-1}(0) = W_0$ and $f^{-1}(i) = W_i \setminus (W_0 \cup \cdots \cup W_{i-1})$ for $0 < i < \infty$), and is therefore $\mathbf{\Sigma}^0_\alpha$. If $\infty \in U$, then $\omega^* \setminus U$ is finite and does not contain ∞, so $f^{-1}(\omega^* \setminus U) \in \mathbf{\Delta}^0_\alpha(\omega^\omega)$, and it follows that $f^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(\omega^\omega)$. Therefore, f is $\mathbf{\Sigma}^0_\alpha$-measurable. Since ρ is $\mathbf{\Sigma}^0_\alpha$-admissible, there is continuous $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f = \rho \circ g$. Since g is continuous and $\rho^{-1}(\infty) \in \mathbf{\Delta}^0_\alpha(dom(\rho))$, $\omega^\omega \setminus X = f^{-1}(\infty) = g^{-1}(\rho^{-1}(\infty)) \in \mathbf{\Delta}^0_\alpha(\omega^\omega)$, which is a contradiction.

Next, assume that S is infinite. Let $X, Y \in \mathbf{\Sigma}^0_\alpha(\omega^\omega)$ be disjoint (i.e. $X \cap Y = \emptyset$). Set $X = \bigcup_{i\in\omega} V_i$, where $V_i \in \mathbf{\Delta}^0_\alpha(\omega^\omega)$ for each $i \in \omega$. Define $A_0 = V_0$ and $A_i = V_i \setminus (V_0 \cup \cdots \cup V_{i-1})$ for $i > 0$. Then $X = \bigcup_{i\in\omega} A_i$, where the A_i are all pairwise disjoint $\mathbf{\Delta}^0_\alpha$ sets (i.e. $A_i \in \mathbf{\Delta}^0_\alpha(\omega^\omega)$ for all $i \in \omega$ and $A_i \cap A_j = \emptyset$ whenever $i \neq j$). Similarly, $Y = \bigcup_{i\in\omega} B_i$, where the B_i are pairwise disjoint $\mathbf{\Delta}^0_\alpha$ sets. Define $f\colon \omega^\omega \to \omega^*$ so that for $\xi \in \omega^\omega$,

$$f(\xi) = \begin{cases} \min\{j \in S \mid j \geq i\} & \text{if } \xi \in A_i \\ \min\{j \in \omega^* \setminus S \mid j \geq i\} & \text{if } \xi \in B_i \\ \infty & \text{if } \xi \notin X \cup Y \end{cases}$$

Since S is infinite but not open, it is clear that $f^{-1}(\infty) = \omega^\omega \setminus (X \cup Y)$, and furthermore $f^{-1}(n)$ $(n < \omega)$ is a finite union of $\mathbf{\Delta}^0_\alpha$ sets and is therefore in $\mathbf{\Delta}^0_\alpha(\omega^\omega)$. It then easily follows that f is $\mathbf{\Sigma}^0_\alpha$-measurable. Let $g\colon\subseteq \omega^\omega \to \omega^\omega$ be continuous such that $f = \rho\circ g$. Since g is continuous and $\rho^{-1}(S) \in \mathbf{\Delta}^0_\alpha(dom(\rho))$, $f^{-1}(S) \in \mathbf{\Delta}^0_\alpha(\omega^\omega)$. Furthermore, $X \subseteq f^{-1}(S)$ and $Y \cap f^{-1}(S) = \emptyset$. Since $X, Y \in \mathbf{\Sigma}^0_\alpha(\omega^\omega)$ were arbitrary, this implies that any two $\mathbf{\Sigma}^0_\alpha$ subsets of ω^ω can be separated by a $\mathbf{\Delta}^0_\alpha$ set, which is a contradiction (see Proposition 22.15 (iv) in [29]). □

Definition 4.1.33 *Let X be an arbitrary topological space. A subset $A \subseteq X$ is* sequentially open *if and only if for every sequence $\{x_i\}_{i\in\omega}$ that converges to $x \in A$, there is some m such that $x_n \in A$ for all $n \geq m$. X is a* sequential space *if and only if all sequentially open subsets of X are open.* □

Note that all countably based spaces are sequential spaces (see Theorem 1.6.14 in [17]).

Theorem 4.1.34 *Let X be a sequential T_0-space and $\rho\colon\subseteq \omega^\omega \to X$ be $\mathbf{\Sigma}^0_\alpha$-admissible ($1 \leq \alpha < \omega_1$). Then $U \subseteq X$ is open if and only if $\rho^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$.*

Proof: If U is open then $\rho^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$ holds because ρ is $\mathbf{\Sigma}^0_\alpha$-measurable.

Assume that $\rho^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(dom(\rho))$ and let $\{x_i\}_{i\in\omega}$ be a sequence converging to $x \in U$. Define $f\colon \omega^* \to X$ so that $f(n) = x_n$ and $f(\infty) = x$. Then f is clearly continuous. If δ is a $\mathbf{\Sigma}^0_\alpha$-admissible representation of ω^*, then $f \circ \delta$ is $\mathbf{\Sigma}^0_\alpha$-measurable, so by the $\mathbf{\Sigma}^0_\alpha$-admissibility of ρ there is continuous $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f \circ \delta = \rho \circ g$. Since g is continuous, $\delta^{-1}(f^{-1}(U)) = g^{-1}(\rho^{-1}(U)) \in \mathbf{\Sigma}^0_\alpha(dom(\delta))$. It follows that $f^{-1}(U)$ is open by Lemma 4.1.32. Since $\infty \in f^{-1}(U)$, there is $m < \omega$ such that $n \in f^{-1}(U)$ for all $n \geq m$. Therefore, $x_n \in U$ for all $n \geq m$. Since $\{x_i\}_{i\in\omega}$ and its limit $x \in U$ were arbitrary, U is sequentially open, hence open because X is a sequential space. □

The rest of this subsection extends Theorem 4.1.34 to the entire hierarchy for a special class of topological spaces.

Lemma 4.1.35 *Let $\rho :\subseteq \omega^\omega \to \omega^\omega$ be a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation of ω^ω ($0 \leq \alpha < \omega_1$). For $0 \leq \beta < \omega_1$ and $A \subseteq \omega^\omega$, $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(\omega^\omega)$ if and only if $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$.*

Proof: If $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(\omega^\omega)$ then $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$ follows from Lemma 4.1.11.

For the converse, let $0 \leq \beta < \omega_1$ and $A \subseteq \omega^\omega$ be such that $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$. First note that $A \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega) \subseteq \mathbf{B}(\omega^\omega)$, because by the $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissibility of ρ there is continuous $h: \omega^\omega \to \omega^\omega$ such that $\rho \circ h$ is the identity on ω^ω. Now assume for a contradiction that $A \notin \mathbf{\Sigma}^0_{\widehat{\beta}}(\omega^\omega)$. Choose $B \in \mathbf{\Pi}^0_{\widehat{\alpha+\beta}}(\omega^\omega) \setminus \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$, and let $f: \omega^\omega \to \omega^\omega$ be a total $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable function such that $B = f^{-1}(A)$, which exists by Theorem 4.1.14. Since f is $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable, there exists continuous $g: \omega^\omega \to \omega^\omega$ such that $f = \rho \circ g$ (note that g is total because f is). Since g is continuous, $B = g^{-1}(\rho^{-1}(A)) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$, a contradiction. □

Lemma 4.1.36 *Let X be a zero-dimensional Polish space and $\rho :\subseteq \omega^\omega \to X$ a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation of X ($0 \leq \alpha < \omega_1$). For $0 \leq \beta < \omega_1$, $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(X)$ if and only if $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$.*

Proof: For the non-trivial part of the lemma, we can assume that X is a closed subset of ω^ω (see Theorem 7.8 in [29]) and $\rho :\subseteq \omega^\omega \to X$ is the restriction of a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation $\rho' :\subseteq \omega^\omega \to \omega^\omega$ of ω^ω as in Proposition 4.1.18 (i.e., $dom(\rho) = (\rho')^{-1}(X)$, and $\rho = \rho'|_{dom(\rho)}$). It follows from these assumptions that $dom(\rho) \in \mathbf{\Pi}^0_{\widehat{\alpha}}(\omega^\omega)$ because X is a closed subset of ω^ω and ρ' is $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable.

The case $\beta = 0$ is the statement of Theorem 4.1.34, so assume $\beta \geq 1$ and $A \subseteq X$ is such that $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$. By Proposition 4.1.6 there is $B \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$ such that $\rho^{-1}(A) = B \cap dom(\rho)$. Since $\alpha < \alpha + \beta$ and $dom(\rho) \in \mathbf{\Pi}^0_{\widehat{\alpha}}(\omega^\omega)$, $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(\omega^\omega)$. Since $(\rho')^{-1}(A) = \rho^{-1}(A)$, it follows from Lemma 4.1.35 that $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(\omega^\omega)$ and hence $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(X)$. □

Definition 4.1.37 *We will say that a space X has a* Polish representation *if and only if there is a $\mathbf{\Sigma}^0_1$-admissible representation $\rho :\subseteq \omega^\omega \to X$ of X such that $dom(\rho)$ with the subspace topology is a (zero-dimensional) Polish space.*

In particular, the real numbers with the Euclidean topology and $\mathcal{P}(\omega)$ with the Scott-topology have Polish representations (an admissible representation of the reals with closed domain of definition is given in [60], and the representation δ of $\mathcal{P}(\omega)$ used in the proof of Corollary 4.1.31 can be shown to be admissible).

Theorem 4.1.38 *Let X be a countably based T_0-space with a Polish representation and $\rho :\subseteq \omega^\omega \to X$ a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation of X ($0 \leq \alpha < \omega_1$). For $0 \leq \beta < \omega_1$, $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(X)$ if and only if $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$.*

Proof: For the non-trivial part of the proof, let $\delta\colon\subseteq \omega^\omega \to X$ be $\mathbf{\Sigma}^0_1$-admissible such that $dom(\delta)$ is Polish. Let $\delta'\colon\subseteq \omega^\omega \to dom(\delta)$ be a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation of $dom(\delta)$. Since $\delta \circ \delta'$ is $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-measurable, there is continuous $f\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\delta \circ \delta' = \rho \circ f$.

Assume $A \subseteq X$ is such that $\rho^{-1}(A) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\rho))$. Then

$$(\delta')^{-1}(\delta^{-1}(A)) = f^{-1}(\rho^{-1}(A)) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\beta}}(dom(\delta'))$$

because f is continuous (here we are using the fact that $dom(\delta') \subseteq dom(f)$). It follows from Lemma 4.1.36 that $\delta^{-1}(A)$ is in $\mathbf{\Sigma}^0_{\widehat{\beta}}(dom(\delta))$, hence $A \in \mathbf{\Sigma}^0_{\widehat{\beta}}(X)$ from Corollary 4.1.31. $\square$

4.1.5 Realizability Theorems

In this subsection we will investigate which functions are realizable with respect to $\mathbf{\Sigma}^0_\alpha$-admissible representations. We only consider topological realizability, and do not consider computational issues.

Definition 4.1.39 *Let X and Y be arbitrary topological spaces, and $f\colon X \to Y$ a function. We say that f is $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\beta \rangle$-realizable by a $\mathbf{\Sigma}^0_\gamma$-measurable function if there is a $\mathbf{\Sigma}^0_\alpha$-admissible representation ρ_X of X and a $\mathbf{\Sigma}^0_\beta$-admissible representation ρ_Y of Y and a $\mathbf{\Sigma}^0_\gamma$-measurable partial function $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f \circ \rho_X = \rho_Y \circ g$. If a continuous such g exists, then we say that f is $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\beta \rangle$-continuously realizable.* $\square$

Lemma 4.1.40 *Let X be an arbitrary topological space, and $\rho\colon\subseteq \omega^\omega \to X$ be a $\mathbf{\Sigma}^0_\alpha$-admissible representation of X $(1 \leq \alpha < \omega_1)$. Then X is a T_0-space.*

Proof: Exactly like Schröder's proof for $\mathbf{\Sigma}^0_1$-admissible representations (Theorem 13 in [48]). $\square$

Lemma 4.1.41 *For $1 \leq \beta < \alpha < \omega_1$, a function from the discrete two point space $\mathbf{2}$ to the Sierpinski space $\mathcal{S}$ is $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\beta \rangle$-continuously realizable if and only if it is a constant function.*

Proof: It is clear that constant functions are always continuously realizable.

For the converse, let $\rho_{\mathbf{2}}$ be a $\mathbf{\Sigma}^0_\alpha$-admissible representation of the discrete two point space $\mathbf{2} = \{0, 1\}$, and let $\rho_{\mathcal{S}}$ be a $\mathbf{\Sigma}^0_\beta$-admissible representation of the Sierpinski space $\mathcal{S} = \{\bot, \top\}$ ($\{\top\}$ is open). Let $f\colon \mathbf{2} \to \mathcal{S}$ be such that $f(0) \neq f(1)$. We will assume $f(0) = \bot$ and $f(1) = \top$, as the proof for the other non-constant function is similar. Assume, for a contradiction, that there is continuous $g\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f \circ \rho_{\mathbf{2}} = \rho_{\mathcal{S}} \circ g$. Let $A \in \mathbf{\Pi}^0_\beta(\omega^\omega) \setminus \mathbf{\Sigma}^0_\beta(\omega^\omega)$, and define $\delta\colon \omega^\omega \to \mathbf{2}$ so that $\delta(A) = \{1\}$ and $\delta(\omega^\omega \setminus A) = \{0\}$. Since $\alpha > \beta$, both A and $\omega^\omega \setminus A$ are in $\mathbf{\Sigma}^0_\alpha(\omega^\omega)$, which implies that δ is $\mathbf{\Sigma}^0_\alpha$-measurable. Hence, there exists continuous $h\colon \omega^\omega \to \omega^\omega$ such that $\delta = \rho_{\mathbf{2}} \circ h$. Since g and h are continuous, $A = \delta^{-1}(f^{-1}(\top)) = h^{-1}(g^{-1}(\rho_{\mathcal{S}}^{-1}(\top))) \in \mathbf{\Sigma}^0_\beta(\omega^\omega)$, a contradiction. $\square$

Note that the following theorem does not assume that X and Y are countably based.

Theorem 4.1.42 *Let X and Y be any topological spaces such that X has a $\mathbf{\Sigma}^0_\alpha$-admissible representation and Y has a $\mathbf{\Sigma}^0_\beta$-admissible representation, where $1 \leq \beta < \alpha < \omega_1$. Then a function from X to Y is $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\beta \rangle$-continuously realizable if and only if it is a constant function.*

Proof: For the non-trivial part, let $f: X \to Y$ be such that $f(x_0) \neq f(x_1)$ for some $x_0, x_1 \in X$. Let $\rho_X :\subseteq \omega^\omega \to X$ be a $\mathbf{\Sigma}^0_\alpha$-admissible representation of X, $\rho_Y :\subseteq \omega^\omega \to Y$ a $\mathbf{\Sigma}^0_\beta$-admissible representation of Y, $\rho_{\mathbf{2}}$ a $\mathbf{\Sigma}^0_\alpha$-admissible representation of the discrete two point space $\mathbf{2}$, and $\rho_{\mathcal{S}}$ a $\mathbf{\Sigma}^0_\beta$-admissible representation of the Sierpinski space $\mathcal{S}$.

Assume, for a contradiction, that there is continuous $g :\subseteq \omega^\omega \to \omega^\omega$ such that $f \circ \rho_X = \rho_Y \circ g$. By Lemma 4.1.40, both X and Y are T_0-spaces, so let $U \subseteq Y$ be an open set such that (without loss of generality) $f(x_1) \in U$ and $f(x_0) \notin U$. Let $p: \mathbf{2} \to X$ be defined as $p(0) = x_0$ and $p(1) = x_1$ and let $q: Y \to \mathcal{S}$ be defined so that $q(U) = \top$ and $q(Y \setminus U) = \bot$.

Since p is continuous, $p \circ \rho_{\mathbf{2}}$ is $\mathbf{\Sigma}^0_\alpha$-measurable, so there is continuous $p' :\subseteq \omega^\omega \to \omega^\omega$ such that $p \circ \rho_{\mathbf{2}} = \rho_X \circ p'$. Likewise, $q \circ \rho_Y$ is $\mathbf{\Sigma}^0_\beta$-measurable so there is continuous $q' :\subseteq \omega^\omega \to \omega^\omega$ such that $q \circ \rho_Y = \rho_{\mathcal{S}} \circ q'$. Define $\phi = q \circ f \circ p$ and $\phi' = q' \circ g \circ p'$. Then $\phi: \mathbf{2} \to \mathcal{S}$ is such that $\phi(0) = \bot$ and $\phi(1) = \top$, and ϕ is $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\beta \rangle$-continuously realized by ϕ', a contradiction. □

Statement (3) in the following is a topological generalization of Brattka's extention [8] of the Kreitz-Weihrauch Representation Theorem [59] to all countably based T_0-spaces and all countable ordinals. Statements (1) and (2) are generalizations of some results by Ziegler [62].

Theorem 4.1.43 *Let X and Y be countably based T_0-spaces, $f: X \to Y$ a total function, and $1 \leq \alpha < \omega_1$.*

1. *f is $\langle \mathbf{\Sigma}^0_1, \mathbf{\Sigma}^0_\alpha \rangle$-continuously realizable if and only if f is $\mathbf{\Sigma}^0_\alpha$-measurable,*
2. *f is $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\alpha \rangle$-continuously realizable if and only if f is continuous,*
3. *f is $\langle \mathbf{\Sigma}^0_1, \mathbf{\Sigma}^0_1 \rangle$-realizable by a $\mathbf{\Sigma}^0_\alpha$-measurable function if and only if f is $\mathbf{\Sigma}^0_\alpha$-measurable.*

Proof: The "if" parts of (1) and (2) immediately follow from the definition of admissibility. For (3), assume f is $\mathbf{\Sigma}^0_\alpha$-measurable. From statement (1) it follows that f is $\langle \mathbf{\Sigma}^0_1, \mathbf{\Sigma}^0_\alpha \rangle$-continuously realizable, and by Theorem 4.1.23 there is a $\mathbf{\Sigma}^0_\alpha$-measurable reduction of any $\mathbf{\Sigma}^0_\alpha$ representation of Y to a $\mathbf{\Sigma}^0_1$-admissible representation of Y. Composing the two produces a $\mathbf{\Sigma}^0_\alpha$-measurable function that $\langle \mathbf{\Sigma}^0_1, \mathbf{\Sigma}^0_1 \rangle$-realizes f.

The proof of the "only if" parts are similar for all three statements, so we only prove (1). Let ρ_X be a $\mathbf{\Sigma}^0_1$-admissible representation of X, ρ_Y a $\mathbf{\Sigma}^0_\alpha$-admissible representation of Y, and assume $g :\subseteq \omega^\omega \to \omega^\omega$ is continuous such that $f \circ \rho_X = \rho_Y \circ g$. Let $U \subseteq Y$ be open. Then $\rho_X^{-1}(f^{-1}(U)) = g^{-1}(\rho_Y^{-1}(U)) \in \mathbf{\Sigma}^0_\alpha(dom(\rho_X))$ because ρ_Y is $\mathbf{\Sigma}^0_\alpha$-measurable, g is continuous, and $dom(\rho_X) \subseteq dom(g)$. By Corollary 4.1.31, it follows that $f^{-1}(U) \in \mathbf{\Sigma}^0_\alpha(X)$, hence f is $\mathbf{\Sigma}^0_\alpha$-measurable (for statement (2), use Theorem 4.1.34 instead of Corollary 4.1.31). □

The following shows that, assuming that a representation of a set is admissible at some level with respect to some topology on the set, then the level of the representation and any corresponding sequential topology on the set is uniquely determined. Note, however, that it is easy to construct representations of a set that are not admissible at any level with respect to any topology on the set.

Corollary 4.1.44 *Let X be a set with at least two elements, and let $\rho\colon\subseteq \omega^\omega \to X$ be an arbitrary function. If τ and τ' are two topologies on X such that ρ is $\mathbf{\Sigma}^0_\alpha$-admissible ($1 \leq \alpha < \omega_1$) with respect to τ, and ρ is $\mathbf{\Sigma}^0_\beta$-admissible ($1 \leq \beta < \omega_1$) with respect to τ', then $\alpha = \beta$. If in addition τ and τ' are sequential topologies then $\tau = \tau'$.*

Proof: The identity function $1_X\colon X \to X$ is non-constant because X has at least two elements. The identity on ω^ω both $\langle \mathbf{\Sigma}^0_\alpha, \mathbf{\Sigma}^0_\beta\rangle$ and $\langle \mathbf{\Sigma}^0_\beta, \mathbf{\Sigma}^0_\alpha\rangle$-continuously realizes 1_X with respect to ρ, and therefore $\alpha = \beta$ by Theorem 4.1.42. Finally, if τ and τ' are sequential topologies, then they satisfy the T_0 axiom by Lemma 4.1.40, therefore (as in the proof of statement (2) of Theorem 4.1.43) Theorem 4.1.34 implies that 1_X is a homeomomorphism, hence $\tau = \tau'$. □

Finally, we give a complete characterization for the case that X has a Polish representation (recall that ordinal addition is non-commutative). Note that a generalization of Theorem 4.1.38 to all countably based T_0-spaces would allow us to drop the "Polish representation" restriction on X.

Theorem 4.1.45 *Let X and Y be countably based T_0-spaces, and further assume X has a Polish representation. For any total function $f\colon X \to Y$ and any countable ordinals α, β and γ, there exists a $\mathbf{\Sigma}^0_{\widehat{\gamma}}$-measurable $g\colon\subseteq \omega^\omega \to \omega^\omega$ that $\langle \mathbf{\Sigma}^0_{\widehat{\alpha}}, \mathbf{\Sigma}^0_{\widehat{\beta}}\rangle$-realizes f if and only if:*

1. *$\alpha > \gamma + \beta$ and f is a constant function, or*
2. *$\alpha \leq \gamma + \beta$ and f is a $\mathbf{\Sigma}^0_{\widehat{\eta}}$-measurable function, where η is (the unique ordinal) such that $\alpha + \eta = \gamma + \beta$.*

Proof: Let ρ_X be a $\mathbf{\Sigma}^0_{\widehat{\alpha}}$-admissible representation of X and ρ_Y a $\mathbf{\Sigma}^0_{\widehat{\beta}}$-admissible representation of Y.

We first prove the "if" part. If f is constant, then it is continuously realizable. Otherwise, if f is $\mathbf{\Sigma}^0_{\widehat{\eta}}$-measurable and $\alpha + \eta = \gamma + \beta$, then $f \circ \rho_X$ is $\mathbf{\Sigma}^0_{\widehat{\gamma+\beta}}$-measurable, so there is continuous $g'\colon\subseteq \omega^\omega \to \omega^\omega$ such that $f \circ \rho_X = \rho'_Y \circ g'$, where ρ'_Y is some $\mathbf{\Sigma}^0_{\widehat{\gamma+\beta}}$-admissible representation of Y. By Theorem 4.1.23 there is a $\mathbf{\Sigma}^0_{\widehat{\gamma}}$-measurable $g''\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\rho'_Y = \rho_Y \circ g''$. Then $g = g'' \circ g'$ is $\mathbf{\Sigma}^0_{\widehat{\gamma}}$-measurable and $f \circ \rho_X = \rho'_Y \circ g' = \rho_Y \circ g'' \circ g' = \rho_Y \circ g$.

For the "only if" part, assume g is $\mathbf{\Sigma}^0_{\widehat{\gamma}}$-measurable and $f \circ \rho_X = \rho_Y \circ g$. First assume that $\alpha > \gamma + \beta$. Let ρ'_Y be a $\mathbf{\Sigma}^0_{\widehat{\gamma+\beta}}$-admissible representation of Y. Since $\rho_Y \circ g$ is $\mathbf{\Sigma}^0_{\widehat{\gamma+\beta}}$-measurable there is continuous $g'\colon\subseteq \omega^\omega \to \omega^\omega$ such that $\rho_Y \circ g = \rho'_Y \circ g'$. Then f is $\langle \mathbf{\Sigma}^0_{\widehat{\alpha}}, \mathbf{\Sigma}^0_{\widehat{\gamma+\beta}}\rangle$-continuously realized by g'. Since $\alpha > \gamma + \beta$, it follows from Theorem 4.1.42 that f is a constant function.

Next assume that $\alpha + \eta = \gamma + \beta$ and $U \subseteq Y$ is open. Then $\rho_X^{-1}(f^{-1}(U)) = g^{-1}(\rho_Y^{-1}(U)) \in \mathbf{\Sigma}^0_{\widehat{\alpha+\eta}}(dom(\rho_X))$ since $\rho_Y \circ g$ is $\mathbf{\Sigma}^0_{\widehat{\gamma+\beta}}$-measurable and $dom(\rho_X)$ is a subset of $dom(g)$. From Theorem 4.1.38 it follows that $f^{-1}(U) \in \mathbf{\Sigma}^0_{\widehat{\eta}}(X)$, hence f is $\mathbf{\Sigma}^0_{\widehat{\eta}}$-measurable. □

4.2 Examples of representations

Theorem 3.1.11 showed that $\langle \mathcal{T}(\mathcal{L}), \tau_{\mathcal{L}} \rangle$ is a $\mathbf{\Sigma}^0_1$-admissible representation of $\mathcal{L}$ with respect to the Π-topology. We also saw by Theorem 3.1.28 that $\langle \mathcal{I}(\mathcal{L}), \iota_{\mathcal{L}} \rangle$ is a $\mathbf{\Sigma}^0_1$-admissible representation of $\mathcal{L}$ with respect to the informant topology. In this section, we will give some more examples of representations that we can later apply to learning theory. Our major focus will be on $\mathbf{\Sigma}^0_2$-admissible representations.

4.2.1 Representations of the Sierpinski space

Recall that the Sierpinski space $\mathcal{S} = \{\bot, \top\}$ is defined as having only the open sets $\{\top\}$, $\mathcal{S}$, and $\emptyset$.

Definition 4.2.1 *For $1 \leq n < \omega$, define $\varphi_n \subseteq \omega^\omega \to \mathcal{S}$ as follows:*

$$\begin{aligned} dom(\varphi_n) &= 2^\omega, \\ \varphi_n(\xi) &= \top \iff \exists x_1 \forall x_2 \ldots Q x_n [\xi(\langle x_1, x_2, \ldots, x_n \rangle) = 1]. \end{aligned}$$

Here, $\langle \cdot, \ldots, \cdot \rangle$ is a bijection between ω^n and ω, and Q is $\exists$ if n is odd and $\forall$, otherwise. □

Lemma 4.2.2 *$\varphi_n \subseteq \omega^\omega \to \mathcal{S}$ is $\mathbf{\Sigma}^0_n$-admissible for all $1 \leq n < \omega$.*

Proof: Note that if n is odd, then

$$\varphi_n^{-1}(\top) = \bigcup_{x_1 \in \omega} \bigcap_{x_2 \in \omega} \cdots \bigcup_{x_n \in \omega} U_{\langle x_1, x_2, \ldots, x_n \rangle},$$

and if n is even then

$$\varphi_n^{-1}(\top) = \bigcup_{x_1 \in \omega} \bigcap_{x_2 \in \omega} \cdots \bigcap_{x_n \in \omega} U_{\langle x_1, x_2, \ldots, x_n \rangle},$$

where $U_{\langle x_1, x_2, \ldots, x_n \rangle} = \{\xi \in 2^\omega \mid \xi(\langle x_1, x_2, \ldots, x_n \rangle) = 1\}$ is clearly clopen in 2^ω. Therefore, φ_n is $\mathbf{\Sigma}^0_n$-measurable.

Next, let $X \in \mathbf{\Sigma}^0_n(\omega^\omega)$. Since ω^ω is zero-dimensional, X can be expressed in the form

$$X = \bigcup_{x_1 \in \omega} \bigcap_{x_2 \in \omega} \cdots C_{\langle x_1, x_2, \ldots, x_n \rangle}.$$

where each $C_{\langle x_1, x_2, \ldots, x_n \rangle}$ is clopen. Define $f \colon \omega^\omega \to 2^\omega$ so that for each $\omega \in 2^\omega$,

$$f(\xi)(i) = 1 \iff \xi \in C_i.$$

Then f is continuous and $f^{-1}(\varphi_n^{-1}(\top)) = X$ because $f^{-1}(U_i) = C_i$. Therefore, every set in $\mathbf{\Sigma}^0_n(\omega^\omega)$ can be reduced to $f^{-1}(\varphi_n^{-1}(\top))$, which implies that φ_n is $\mathbf{\Sigma}^0_n$-admissible by Proposition 4.1.16. □

Next we introduce a "limiting" representation of the Sierpinski space, which is implicitly used in the classification in the limit paradigms as studied by, for example, Kelly [31] and Martin et al. [36].

Definition 4.2.3 *Define* $\delta_{\mathcal{S}}: \subseteq \omega^\omega \to \mathcal{S}$ *as follows:*

$$\begin{aligned} dom(\delta_{\mathcal{S}}) &= 2^\omega, \\ \delta_{\mathcal{S}}(\xi) &= \top \iff \exists m \forall n \geq m[\xi(n) = 1]. \end{aligned}$$

□

Intuitively, $\xi \in 2^\omega$ represents $\top$ if and only if the sequence of 0's and 1's in ξ eventually converges to 1. If ξ does not converge, or if it converges to 0, then ξ represents $\bot$.

Theorem 4.2.4 $\delta_{\mathcal{S}}: \subseteq \omega^\omega \to \mathcal{S}$ *is a* $\mathbf{\Sigma}^0_2$*-admissible representation of* $\mathcal{S}$*.*

Proof: Clearly, $\delta_{\mathcal{S}}$ is $\mathbf{\Sigma}^0_2$-measurable because

$$\delta_{\mathcal{S}}^{-1}(\top) = \bigcup_{m \in \omega} \bigcap_{n \geq m} U_n,$$

where $U_n = \{\xi \in 2^\omega \mid \xi(n) = 1\}$ is clopen in 2^ω. So it suffices to show that there is a continuous $f: 2^\omega \to 2^\omega$ such that $\varphi_2 = \delta_{\mathcal{S}} \circ f$. In other words, we want f to satisfy

$$\exists x_1 \forall x_2 [\xi(\langle x_1, x_2 \rangle) = 1] \iff \exists m \forall n \geq m[f(\xi)(n) = 1].$$

Intuitively, f can be constructed as follows. First, f scans $\xi(\langle 0, 0\rangle)$, $\xi(\langle 0, 1\rangle)$, $\xi(\langle 0, 2\rangle), \ldots$, outputting 1 each time until an n is found such that $\xi(\langle 0, n\rangle) = 0$. If such an n is found, then f outputs 0, and then starts searching the next row $\xi(\langle 1, 0\rangle)$, $\xi(\langle 1, 1\rangle)$, $\xi(\langle 1, 2\rangle), \ldots$, outputting 1's until a 0 is found, in which case f outputs a 0 and starts searching the next row, etc. If there exists an x_1 such that $\xi(\langle x_1, x_2\rangle) = 1$ for all x_2, then when f reaches the first such row f will forever after output 1's. If there is no such x_1, then f will output infinitely many 0's. Therefore, $\varphi_2 = \delta_{\mathcal{S}} \circ f$. □

Next we give a representation of $\mathcal{S}^\omega$, which is the product of infinitely many copies of the Sierpinski space. Let $\phi: \omega^\omega \to (\omega^\omega)^\omega$ be a homeomorphism. In the following definition, we will abbreviate $\phi(\xi)(i)$ as $(\xi)_i$ for each $\xi \in \omega^\omega$.

Definition 4.2.5 *Define* $\delta_{\mathcal{S}^\omega} \subseteq \omega^\omega \to \mathcal{S}^\omega$ *as follows:*

$$\begin{aligned} dom(\delta_{\mathcal{S}^\omega}) &= \phi^{-1}((2^\omega)^\omega), \\ \delta_{\mathcal{S}^\omega}(\xi)(i) &= \top \iff \exists m \forall n \geq m[(\xi)_i(n) = 1]. \end{aligned}$$

□

Intuitively, $\delta_{\mathcal{S}^\omega}$ is just a countably infinite number of copies of $\delta_{\mathcal{S}}$. It follows from Theorem 4.2.4 and Proposition 4.1.21 that $\delta_{\mathcal{S}^\omega}$ is a $\mathbf{\Sigma}^0_2$-admissible representation of $\mathcal{S}^\omega$.

4.2.2 Representations of ω and $\omega_\bot$

Next we consider representations of ω viewed as a discrete space (i.e., every subset is open), and $\omega_\bot$ which is ω with an additional "bottom" element.

We will first construct a representation of $\omega_\bot$. Formally, $\omega_\bot = \omega \cup \{\bot\}$ and $U \subseteq \omega_\bot$ is open if and only if $U = \omega_\bot$ or else $\bot \notin U$. We can define an embedding $f\colon \omega_\perp \to \mathcal{S}^\omega$ from $\omega_\perp$ to the product of infinitely many Sierpinski spaces by defining $f(\perp)(m) = \perp$ for each $m \in \omega$, and for $i \neq \perp$ define $f(i)(m) = \top \iff i = m$ for each $m \in \omega$. So $X \in range(f)$ if and only if X contains at most one occurence of $\top$. It is easily seen that $\omega_\perp$ is homeomorphic to $f(\omega) \subseteq \mathcal{S}^\omega$, so a representation for $\omega_\perp$ can be created by restricting $\delta_{\mathcal{S}^\omega}$ to the range of f. In particular, $\delta_{\omega_\perp}\colon \omega^\omega \to \omega_\perp$ defined as:

$$\begin{aligned} dom(\delta_{\omega_\perp}) &= \delta_{\mathcal{S}^\omega}^{-1}(range(f)), \\ \delta_{\omega_\perp}(\xi) &= \perp \iff \forall i \forall m \exists n \geq m[(\xi)_i(n) \neq 1], \\ \delta_{\omega_\perp}(\xi) &= i \iff \exists m \forall n \geq m[(\xi)_i(n) = 1], \end{aligned}$$

is a $\mathbf{\Sigma}^0_2$-admissible representation of $\omega_\perp$. This representation is somewhat non-intuitive, but it is a start. Next consider the representation $\lim_\omega\colon \omega^\omega \to \omega_\perp$ defined as:

$$\lim_\omega(\xi) = \begin{cases} i & \text{if } \exists m \forall n \geq m[\xi(n) = i], \\ \perp & \text{otherwise.} \end{cases}$$

So $\lim_\omega$ sends converging sequences in ω^ω to the limit of the sequence, and sends non-converging sequences to $\perp$.

Theorem 4.2.6 *$\lim_\omega\colon \omega^\omega \to \omega_\perp$ is a $\mathbf{\Sigma}^0_2$-admissible representation of $\omega_\perp$.*

Proof: We see that $\lim_\omega$ is $\mathbf{\Sigma}^0_2$-measurable because for each $i \in \omega$,

$$\lim_\omega^{-1}(\{i\}) = \bigcup_{m\in\omega} \bigcap_{n \geq m} U_n^i,$$

where $U_n^i = \{\xi \in \omega^\omega \mid \xi(n) = i\}$. Since $\mathbf{\Sigma}^0_2$-sets are closed under countable unions, it follows that the preimage under $\lim_\omega$ of any subset of ω is a $\mathbf{\Sigma}^0_2$-subset of ω^ω. Therefore, the preimage of any open subset of $\omega_\perp$ is open in ω^ω.

Now we just have to show that $\delta_{\omega_\perp}$ continuously reduces to $\lim_\omega$. This means we need to construct a continuous function $g\colon\subseteq \omega^\omega \to \omega^\omega$ in a way that for every $\xi \in dom(\delta_{\omega_\perp})$, if there is $i \in \omega$ satisfying $\exists m \forall n \geq m[(\xi)_i(n) = 1]$ then $g(\xi)$ converges to i (which is necessarily unique), and $g(\xi)$ does not converge if no such i exists.

To do this, first we define a function $g'\colon\subseteq \omega^\omega \to \omega \times \omega$. Set $g'(\xi)(0) = \langle 0, 1\rangle$ and for $n \geq 0$ assume $g'(\xi)(n) = \langle j, b\rangle$ and define

$$g'(\xi)(n+1) = \begin{cases} \langle j, b\rangle & \text{if } (\xi)_j(n) = 1, \\ \langle j+1, b\rangle & \text{if } (\xi)_j(n) \neq 1 \text{ and } j < b, \\ \langle 0, b+1\rangle & \text{if } (\xi)_j(n) \neq 1 \text{ and } j \geq b. \end{cases}$$

Intuitively, g' searches up to b for a j satisfying $(\xi)_j(n) = 1$. If $(\xi)_j(n) \neq 1$, then g' checks $j+1$, etc. If g' searches all the way up to b without finding an appropriate j, then g' increments b and starts over the search from zero. This forces g' to check each value of j infinitely many times. It is then easy to see that $g'(\xi)$ converges to $\langle i, b\rangle$ for some $b \in \omega$ if and only if there is some $m \in \omega$ such that $(\xi)_i(n) = 1$ for all $n \geq m$, which holds just in case $\delta_{\omega_\perp}(\xi) = i$. Thus, by defining $g(\xi)(n) = \pi_1(g'(\xi)(n))$, we see that $\delta_{\omega_\perp} = \lim_\omega \circ g$. □

We can let $\lim_\omega\colon\subseteq \omega^\omega \to \omega$ be the restriction of $\lim_\omega\colon \omega^\omega \to \omega_\perp$ to only the sequences that converge. By Proposition 4.1.18 and the above theorem, we immediately obtain the following.

Theorem 4.2.7 $\lim_\omega :\subseteq \omega^\omega \to \omega$ *is a* $\mathbf{\Sigma}^0_2$*-admissible representation of* ω. □

4.2.3 Representations of ω^ω and more

Since ω^ω is homeomorphic to the product of countably many copies of ω, we can use Proposition 4.1.21 and Theorem 4.2.7 to construct a simple $\mathbf{\Sigma}^0_2$-admissible representation of ω^ω. Again, let $\phi: \omega^\omega \to (\omega^\omega)^\omega$ be a homeomorphism and abbreviate $\phi(\xi)(i)$ as $(\xi)_i$ for each $\xi \in \omega^\omega$.

Corollary 4.2.8 *The function* $\lim_{\omega^\omega} :\subseteq \omega^\omega \to \omega^\omega$ *defined as:*

$$\begin{aligned} dom(\lim_{\omega^\omega}) &= \{\xi \in \omega^\omega \mid \forall i\, \exists m \forall n \geq m[(\xi)_i(n) = (\xi)_i(m)]\}, \\ \lim_{\omega^\omega}(\xi) &= \zeta \iff \forall i\, \exists m \forall n \geq m[(\xi)_i(n) = \zeta(i)], \end{aligned}$$

is a $\mathbf{\Sigma}^0_2$*-admissible representation of* ω^ω. □

Given $\xi \in \omega^\omega$, for each $n \in \omega$ let $\xi_n \in \omega^\omega$ be defined as $\xi_n(i) = (\xi)_i(n)$ for each $i \in \omega$. Then from the definition of $\lim_{\omega^\omega}$, we see that $\lim_{\omega^\omega}(\xi) = \zeta$ if and only if for each $i \in \omega$ there exists m such that $\xi_n(i) = \zeta(i)$ for all $n \geq m$. In other words, $\lim_{\omega^\omega}(\xi) = \zeta$ if and only if the sequence $\{\xi_n\}_{i\in\omega}$ converges to ζ with respect to the usual topology on ω^ω. Thus, $\lim_{\omega^\omega}$ as defined above is equivalent to the "jump" introduced by Ziegler [62] and the "derivative" used in Brattka and Makananise [9]. These two papers only investigated metric spaces, but the results of the previous section show that the results in these two papers can be extended to all countably based T_0-spaces.

By Corollary 4.1.24, it follows that $\lim_{\omega^\omega} \circ \lim_{\omega^\omega}$ is a $\mathbf{\Sigma}^0_3$-admissible representation of ω^ω, that $\lim_{\omega^\omega} \circ \lim_{\omega^\omega} \circ \lim_{\omega^\omega}$ is $\mathbf{\Sigma}^0_4$-admissible, and so on. Furthermore, if $\langle \mathcal{R}, \rho \rangle$ is a $\mathbf{\Sigma}^0_1$-admissible representation of a countably based T_0-space X, then $\langle \mathcal{R}, \rho \circ \lim_{\omega^\omega} \rangle$ is $\mathbf{\Sigma}^0_2$-admissible for X, $\langle \mathcal{R}, \rho \circ \lim_{\omega^\omega} \circ \lim_{\omega^\omega} \rangle$ is $\mathbf{\Sigma}^0_3$-admissible, etc. This provides a simple way of constructing $\mathbf{\Sigma}^0_n$-admissible representations for $n < \omega$. Computability aspects of these representations for metric spaces has been investigated in [8], [62], and [9].

4.3 Applications to learning theory

In this section, we will apply the results of the previous sections in the chapter to problems in Gold-style learning models.

Although previous chapters involved specific forms of information presentation, like positive data (the representation $\langle \mathcal{T}(\mathcal{L}), \tau_\mathcal{L} \rangle$) or positive and negative data (the representation $\langle \mathcal{I}(\mathcal{L}), \iota_\mathcal{L} \rangle$), in this section we will be much more general. In the following we will assume that $\langle \mathcal{R}, \rho \rangle$ is a $\mathbf{\Sigma}^0_1$-admissible representation of a concept space $\mathcal{L}$, with respect to some countably based T_0-topology on $\mathcal{L}$. By Corollary 4.1.44, the topology that $\langle \mathcal{R}, \rho \rangle$ induces on $\mathcal{L}$ is uniquely determined, so we will often write $\mathcal{L}_\rho$ to emphasize that we view $\mathcal{L}$ as a topological space with the topology determined by $\langle \mathcal{R}, \rho \rangle$. So, for example, $\mathcal{L}_{\tau_\mathcal{L}}$ is $\mathcal{L}$ with the Π-topology, and $\mathcal{L}_{\iota_\mathcal{L}}$ is $\mathcal{L}$ with the informant topology.

4.3.1 Identification in the limit

It has become clear that the domain of a learner can be interpreted as a representation of $\mathcal{L}$, but the following theorem shows that the learner's codomain is also a representation of $\mathcal{L}$.

Theorem 4.3.1 *Let $\langle \mathcal{H}, h\rangle$ be a hypothesis space for a concept space $\mathcal{L}$, and let $\lim_{\mathcal{H}}\colon\subseteq \mathcal{H}^\omega \to \mathcal{H}$ be the function that maps each converging sequence of hypotheses to its limit. Then,*

1. *$\langle \mathcal{H}, h\rangle$ is $\mathbf{\Sigma}^0_1$-admissible with respect to the discrete topology on $\mathcal{L}$,*
2. *$\langle \mathcal{H}^\omega, \lim_{\mathcal{H}}\rangle$ is $\mathbf{\Sigma}^0_2$-admissible with respect to the discrete topology on $\mathcal{H}$,*
3. *$\langle \mathcal{H}^\omega, h\circ\lim_{\mathcal{H}}\rangle$ is $\mathbf{\Sigma}^0_2$-admissible with respect to the discrete topology on $\mathcal{L}$.*

Proof: The first claim is trivial, although the reader should note that technically we must assume that $\mathcal{H}$ is appropriately encoded as a subspace of ω^ω in order for $\langle \mathcal{H}, h\rangle$ to be a representation as we have defined it. This is done by encoding each $H \in \mathcal{H}$ as the sequence $\xi \in \omega^\omega$ that satisfies $\xi(n) = H$ for all n.

The second claim follows from Theorem 4.2.7, and the third claim follows from the first two and Corollary 4.1.24. □

We let $\mathcal{L}_\rho$ denote $\mathcal{L}$ with the topology induced by $\langle \mathcal{R}, \rho\rangle$, and $\mathcal{L}_h$ denote $\mathcal{L}$ with the topology induced by $\langle \mathcal{H}, h\rangle$. Note that from the above theorem, every subset of $\mathcal{L}_h$ is open. Let $id_{\mathcal{L}}\colon \mathcal{L}_\rho \to \mathcal{L}_h$ be the identity function on $\mathcal{L}$.

By definition, a learner (continuous function) $\psi\colon \mathcal{R} \to \mathcal{H}^\omega$ identifies $\mathcal{L}$ in the limit with respect to $\langle \mathcal{R}, \rho\rangle$ and $\langle \mathcal{H}, h\rangle$ if and only if $\rho = h\circ\lim_{\mathcal{H}}\circ\psi$. Since $id_{\mathcal{L}}$ is the identity on $\mathcal{L}$, this is equivalent to requiring ψ to satisfy $id_{\mathcal{L}}\circ\rho = h\circ\lim_{\mathcal{H}}\circ\psi$. Since $\langle \mathcal{R}, \rho\rangle$ is $\mathbf{\Sigma}^0_1$-admissible for $\mathcal{L}_\rho$, and $\langle \mathcal{H}^\omega, h\circ\lim_{\mathcal{H}}\rangle$ is $\mathbf{\Sigma}^0_2$-admissible for $\mathcal{L}_h$, it follows from Theorem 4.1.43 that such a ψ exists if and only if $id_{\mathcal{L}}\colon \mathcal{L}_\rho \to \mathcal{L}_h$ is $\mathbf{\Sigma}^0_2$-measurable. Since every subset of $\mathcal{L}_h$ is open, $id_{\mathcal{L}}$ is $\mathbf{\Sigma}^0_2$-measurable if and only if every subset of $\mathcal{L}_\rho$ is a $\mathbf{\Sigma}^0_2$-set if and only if every subset of $\mathcal{L}_\rho$ is a $\mathbf{\Delta}^0_2$-set. Applying now Proposition 4.1.7, we have proved the following.

Theorem 4.3.2 *Let $\mathcal{L}$ be a (countable) concept space with representation $\langle \mathcal{R}, \rho\rangle$ and hypothesis space $\langle \mathcal{H}, h\rangle$, such that $\langle \mathcal{R}, \rho\rangle$ is $\mathbf{\Sigma}^0_1$-admissible with respect to a countably based T_0-topology on $\mathcal{L}$. Then the following are equivalent.*

1. *$\mathcal{L}$ is identifiable in the limit with respect to $\langle \mathcal{R}, \rho\rangle$ and $\langle \mathcal{H}, h\rangle$,*
2. *$id_{\mathcal{L}}\colon \mathcal{L}_\rho \to \mathcal{L}_h$ is $\mathbf{\Sigma}^0_2$-measurable,*
3. *$\mathcal{L}_\rho$ is a T_D-space.*

$$
\begin{tikzcd}
\mathcal{R} \arrow[r, "\psi"] \arrow[d, "\rho"'] & \mathcal{H}^\omega \arrow[d, "h\circ\lim_{\mathcal{H}}"] \\
\mathcal{L}_\rho \arrow[r, "id_{\mathcal{L}}"] & \mathcal{L}_h
\end{tikzcd}
$$

□

Note that any learner ψ that identifies $\mathcal{L}$ in the limit (with respect to $\langle\mathcal{R},\rho\rangle$ and $\langle\mathcal{H},h\rangle$) is a $\langle\mathbf{\Sigma}^0_1,\mathbf{\Sigma}^0_2\rangle$-realizer of the function $id_{\mathcal{L}}\colon\mathcal{L}_\rho\to\mathcal{L}_h$. Thus, in a sense, the goal of ψ is to "compute" the function $id_{\mathcal{L}}$.

Now, since $\mathcal{L}_h$ has the discrete topology, every property of $\mathcal{L}_h$ is observable in the sense defined in the introduction. On the other hand, the same cannot always be said about $\mathcal{L}_\rho$, since some subsets may not be open. Although being the identity function may seem trivial, since the topologies on $\mathcal{L}_\rho$ and $\mathcal{L}_h$ are different in general, $id_{\mathcal{L}}$ is in fact modifying the information theoretical structure (topology) of $\mathcal{L}$ so as to make all properties observable.

4.3.2 Variations on convergence

Once the representation $\langle\mathcal{R},\rho\rangle$ and hypothesis space $\langle\mathcal{H},h\rangle$ have been determined, the function $id_{\mathcal{L}}\colon\mathcal{L}_\rho\to\mathcal{L}_h$ and its complexity are uniquely determined. Variations on the convergence requirements of a successful learner can be interpreted as finding out how much we must weaken the representation induced by the hypothesis space until a learner can realize the function $id_{\mathcal{L}}$.

As an example, *finite identification* is a variation of identification in the limit where the learner must read in an information presentation and eventually output a single correct hypothesis and then halt. This paradigm was also introduced by Gold [21] and characterized for the learning from texts and learning from informants paradigms by Mukouchi [38].

This paradigm is modeled in our framework by simply defining a finite learner to be a continuous function $\psi\colon\mathcal{R}\to\mathcal{H}$, and defining ψ to be successful if and only if $\rho = h\circ\psi$. Since $\mathcal{H}$ is assumed to have the discrete topology, the continuity requirement guarantees that the entire output of ψ will only depend on a finite portion of the input, hence ψ finitely identifies $\mathcal{L}$.

Using the fact that $\langle\mathcal{H},h\rangle$ is $\mathbf{\Sigma}^0_1$-admissible for $\mathcal{L}_h$, we can apply Theorem 4.1.43 to obtain the following topological characterization of finite identification.

Theorem 4.3.3 *Let $\mathcal{L}$ be a (countable) concept space with representation $\langle\mathcal{R},\rho\rangle$ and hypothesis space $\langle\mathcal{H},h\rangle$, such that $\langle\mathcal{R},\rho\rangle$ is $\mathbf{\Sigma}^0_1$-admissible with respect to a countably based T_0-topology on $\mathcal{L}$. Then the following are equivalent.*

1. *$\mathcal{L}$ is finitely identifiable with respect to $\langle\mathcal{R},\rho\rangle$ and $\langle\mathcal{H},h\rangle$,*
2. *$id_{\mathcal{L}}\colon\mathcal{L}_\rho\to\mathcal{L}_h$ is continuous,*
3. *$\mathcal{L}_\rho$ is a discrete space.*

$$\begin{tikzcd} \mathcal{R} \arrow[r, "\psi"] \arrow[d, "\rho"'] & \mathcal{H} \arrow[d, "h"] \\ \mathcal{L}_\rho \arrow[r, "id_{\mathcal{L}}"] & \mathcal{L}_h \end{tikzcd}$$

□

Mind-change complexity uses representations like $\mathcal{H}^\omega_\alpha$ that are weaker than $\mathcal{H}$ but stronger than $\mathcal{H}^\omega$, so these representations make up a hierarchy between

$\mathbf{\Sigma}^0_1$ and $\mathbf{\Sigma}^0_2$-admissibility. This can be viewed as a refinement of determining the complexity of $id_{\mathcal{L}}\colon \mathcal{L}_\rho \to \mathcal{L}_h$ between continuity and $\mathbf{\Sigma}^0_2$-measurability.

Although much research on the learning in the limit paradigm has been dedicated on characterizing stronger convergence requirements, little research has been done on relaxing the convergence requirements. By Proposition 4.1.7, we see that $id_{\mathcal{L}}\colon \mathcal{L}_\rho \to \mathcal{L}_h$ is always $\mathbf{\Sigma}^0_3$-measurable because $\mathcal{L}$ is countable. This suggests that a further refinement of the complexity of representations between $\mathbf{\Sigma}^0_2$ and $\mathbf{\Sigma}^0_3$-admissibility (in a fashion similar to mind-change complexity) could be used to better understand the difficulty of identifying concept spaces that are not identifiable in the limit. Since many concept spaces of interest are not identifiable in the limit, a further refinement in this sense is important to let us know what our options are when we are confronted with a "non-learnable" concept space.

4.3.3 Classification

In the classification paradigm, instead of the learner identifying a concept, the learner must determine whether a particular property holds of the concept. In particular, given a property $P \subseteq \mathcal{L}$, and $R \in \mathcal{R}$, the learner must determine whether $\rho(R) \in P$. In this case, a hypothesis space for $\mathcal{L}$ is not needed, so it is meaningful to investigate classification of properties of uncountable concept spaces. The relationship between classification and topology has been investigated by many researches, such as Kelly [31] and Martin et al. [36]. Below we will compare their results with our own.

Here is a summary of some basic variations of classification. The notation we use comes from Kelly [31].

Definition 4.3.4 *Let $\mathcal{L}$ be a (possibly uncountable) concept space, $\langle \mathcal{R}, \rho \rangle$ a representation of $\mathcal{L}$, $P \subseteq \mathcal{L}$, and $\psi\colon \mathcal{R} \to 2^\omega$ be a continuous function.*

1. *ψ decides P with certainty if and only if*

$$\begin{aligned} \rho(R) \in P \quad &\iff \quad \exists n : \psi(R)(n) = 1, \text{ and} \\ \rho(R) \notin P \quad &\iff \quad \exists n : \psi(R)(n) = 0; \end{aligned}$$

2. *ψ verifies P with certainty if and only if*

$$\rho(R) \in P \quad \iff \quad \exists n : \psi(R)(n) = 1;$$

3. *ψ decides P in the limit if and only if*

$$\begin{aligned} \rho(R) \in P \quad &\iff \quad \exists m \forall n \geq m : \psi(R)(n) = 1, \text{ and} \\ \rho(R) \notin P \quad &\iff \quad \exists m \forall n \geq m : \psi(R)(n) = 0; \end{aligned}$$

4. *ψ verifies P in the limit if and only if*

$$\rho(R) \in P \quad \iff \quad \exists m \forall n \geq m : \psi(R)(n) = 1.$$

□

Note that verifying P with certainty is the same as observing P in the sense described in the introduction.

Just as we have viewed the identification paradigms as realizing a function between topological spaces, the classification paradigms above can be viewed as the problem of realizing functions. In particular, let $\mathcal{S} = \{\bot, \top\}$ be the Sierpinski space and let $2 = \{0, 1\}$ be the discrete space with two points. Given $P \subseteq \mathcal{L}$, define a function $\chi_P^{\mathcal{S}}: \mathcal{L} \to \mathcal{S}$ and $\chi_P^2: \mathcal{L} \to 2$ as

$$\begin{aligned} \chi_P^{\mathcal{S}}(L) &= \begin{cases} \top & \text{if } L \in P, \\ \bot & \text{if } L \notin P; \end{cases} \\ \chi_P^{2}(L) &= \begin{cases} 1 & \text{if } L \in P, \\ 0 & \text{if } L \notin P; \end{cases} \end{aligned}$$

We can think of $\chi_P^{\mathcal{S}}$ as the problem of "verifying" P, and χ_P^2 as the problem of "deciding" P. Comparing the criteria for the different classification paradigms with the representations we defined earlier for $\mathcal{S}$ and discrete spaces, we can easily see that:

1. ψ decides P with certainty if and only if ψ $\langle \mathbf{\Sigma}_1^0, \mathbf{\Sigma}_1^0 \rangle$-realizes χ_P^2,

2. ψ verifies P with certainty if and only if ψ $\langle \mathbf{\Sigma}_1^0, \mathbf{\Sigma}_1^0 \rangle$-realizes $\chi_P^{\mathcal{S}}$,

3. ψ decides P in the limit if and only if ψ $\langle \mathbf{\Sigma}_1^0, \mathbf{\Sigma}_2^0 \rangle$-realizes χ_P^2,

4. ψ verifies P in the limit if and only if ψ $\langle \mathbf{\Sigma}_1^0, \mathbf{\Sigma}_2^0 \rangle$-realizes $\chi_P^{\mathcal{S}}$.

We therefore obtain the following simple characterization of classification.

Theorem 4.3.5 *Let $\mathcal{L}$ be a concept space with representation $\langle \mathcal{R}, \rho \rangle$, such that $\langle \mathcal{R}, \rho \rangle$ is $\mathbf{\Sigma}_1^0$-admissible with respect to a countably based T_0-topology on $\mathcal{L}$. The following hold for any $P \subseteq \mathcal{L}$.*

1. *P is decideable with certainty $\iff$ $\chi_P^2: \mathcal{L}_\rho \to 2$ is continuous $\iff$ $P \in \mathbf{\Delta}_1^0(\mathcal{L}_\rho)$,*

2. *P is verifiable with certainty $\iff$ $\chi_P^{\mathcal{S}}: \mathcal{L}_\rho \to \mathcal{S}$ is continuous $\iff$ $P \in \mathbf{\Sigma}_1^0(\mathcal{L}_\rho)$,*

3. *P is decideable in the limit $\iff$ $\chi_P^2: \mathcal{L}_\rho \to 2$ is $\mathbf{\Sigma}_2^0$-measurable $\iff$ $P \in \mathbf{\Delta}_2^0(\mathcal{L}_\rho)$,*

4. *P is verifiable in the limit $\iff$ $\chi_P^{\mathcal{S}}: \mathcal{L}_\rho \to \mathcal{S}$ is $\mathbf{\Sigma}_2^0$-measurable $\iff$ $P \in \mathbf{\Sigma}_2^0(\mathcal{L}_\rho)$.*

$$\begin{tikzcd} \mathcal{R} \arrow[r, "\psi"] \arrow[d, "\rho"'] & 2^\omega \arrow[d, "\delta_{\mathcal{S}}"] \\ \mathcal{L}_\rho \arrow[r, "\chi_P^{\mathcal{S}}"] & \mathcal{S} \end{tikzcd}$$

□

Kelly [31] gave topological characterizations of the classification paradigms in terms of the Borel complexity of the property just like the above theorem. However, the characterizations only applied to the case that $\mathcal{L}$ is a zero-dimensional space, so the results cannot be applied to classification from positive data only. Martin et al. [36] extended the Borel characterizations of classification to a more general logical setting that could also handle positive data only. In their case, $\mathcal{L}$ is essentially a set of structures of a logical language and $\mathcal{R}$ essentially represents each structure by enumerating (a subset of) the formulas that are true in it. Classification in this case is to determine whether or not a structure is a model of some given logical formula. Classification with mind-change complexity was also characterized in [36] using the Hausdorff difference hierarchy (see [29]). Connections between logic and topology are a major contribution of [36], but to the disadvantage that the model is extremely complicated thus difficult to apply.

The advantage of our approach is in its generality, since it applies to all countably based T_0-spaces, of which the topological spaces in [31] and [36] are special cases. We have also given a complete characterization of which properties are classifiable, whereas [36] only considers classification with respect to properties that correspond to the set of models of some given logical formula. Combined with restrictions on the topological spaces considered, this explains why [36] could obtain complete characterizations of classification with respect to a definition of the Borel hierarchy which is in general different from Definition 4.1.1.[1]

Our approach is also relatively simple compared with [36], since the only parameters are the characteristic function χ_P and the representations of the domain and codomain of χ_P. We have also clarified how identification and classification are related, the difference being only in the function, $id_{\mathcal{L}}$ or χ_P, to be realized. Variations on convergence are also made clear, since they simply involve a modification of the representation of the codomain of the function to be realized (the characterization of mind-change complexity using the difference hierarchy as in [36] is extended to our framework in a straightforwad way). In this way, characterizations of various inductive inference paradigms are nicely formulated in our single Theorem 4.1.43.

4.3.4 Borel complexity of concept spaces

This subsection investigates the Borel complexity of concept spaces as subsets of $\mathcal{P}(\omega)$. These results will be applied in the following subsection. We will assume

[1] The Borel hierarchy is defined in [36] for a topological space X generated by B as follows (we put "hats" over the symbols to distinguish them from our definitions): $\widehat{\mathbf{\Pi}^0_0}$-sets are formed by the finite unions and intersections of B; $\widehat{\mathbf{\Sigma}^0_0}$-sets are the complements of $\widehat{\mathbf{\Pi}^0_0}$-sets; $\widehat{\mathbf{\Sigma}^0_\alpha}$-sets are the countable unions of sets from the $\widehat{\mathbf{\Pi}^0_\beta}$-sets ($\beta < \alpha$); $\widehat{\mathbf{\Pi}^0_\alpha}$-sets are the complements of the $\widehat{\mathbf{\Sigma}^0_\alpha}$-sets. It immediately follows that every $\widehat{\mathbf{\Sigma}^0_\alpha}$-set in the sense of [36] is $\mathbf{\Sigma}^0_\alpha$ in our sense (for $\alpha > 0$). However, consider the space $X = \{0,1,2\}$ with open sets $B = \{\emptyset, \{2\}, \{1,2\}, \{0,1,2\}\}$. B is closed under *arbitrary* unions and intersections, so $\widehat{\mathbf{\Sigma}^0_1}(X) = \widehat{\mathbf{\Pi}^0_0}(X) = B$ and $\widehat{\mathbf{\Pi}^0_1}(X) = \widehat{\mathbf{\Sigma}^0_0}(X) = \{\emptyset, \{0\}, \{0,1\}, \{0,1,2\}\}$. $\widehat{\mathbf{\Sigma}^0_2}(X)$ is defined as unions of sets from $\widehat{\mathbf{\Pi}^0_0}(X)$ and $\widehat{\mathbf{\Pi}^0_1}(X)$, hence $\widehat{\mathbf{\Sigma}^0_2}(X) = \{\emptyset, \{0\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\}$. In particular, $\{1\} \notin \widehat{\mathbf{\Sigma}^0_2}(X)$ even though $\mathbf{\Delta}^0_2(X) = \mathcal{P}(X)$ according to our definition. This example shows that there are properties P that are *decideable* in the limit without even being $\widehat{\mathbf{\Sigma}^0_2}$ in the sense of [36].

the Π-topology on all concept spaces in this subsection.

Lemma 4.3.6 *For every concept space $\mathcal{L}$, $\mathcal{A}(\mathcal{L}) \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega))$.*

Proof: It is easy to see that $C_{\mathcal{L}}: \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ is continuous and $C_{\mathcal{L}}(\mathcal{P}(\omega)) = \mathcal{A}(\mathcal{L})$. Let $id: \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ be the identify function, then $\langle id, C_{\mathcal{L}}\rangle: \mathcal{P}(\omega) \to \mathcal{P}(\omega) \times \mathcal{P}(\omega)$, defined as $\langle id, C_{\mathcal{L}}\rangle(X) = \langle X, C_{\mathcal{L}}(X)\rangle$, is continuous by Theorem 3.1.14. By Proposition 4.1.8, $\Delta_{\mathcal{P}(\omega)} = \{\langle X, Y\rangle \in \mathcal{P}(\omega) \times \mathcal{P}(\omega) \mid X = Y\}$ is in $\mathbf{\Pi}^0_2(\mathcal{P}(\omega))$, hence $A = \langle id, C_{\mathcal{L}}\rangle^{-1}(\Delta_{\mathcal{P}(\omega)})$ is in $\mathbf{\Pi}^0_2(\mathcal{P}(\omega))$ because $\langle id, C_{\mathcal{L}}\rangle$ is continuous. Since $X \in A$ if and only if $X = C_{\mathcal{L}}(X)$, $A = \mathcal{A}(\mathcal{L})$, hence $\mathcal{A}(\mathcal{L}) \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega))$. □

The following theorem relates topological properties of concept spaces that are identifiable in the limit from positive data with the Borel complexity of the concept space. This theorem emphasizes the naturalness of the hierarchy of topological properties investigated in Chapter 3.

Theorem 4.3.7 *If $\mathcal{L}$ is a countable T_D-space, then*

1. *$\mathcal{L} \in \mathbf{\Delta}^0_3(\mathcal{A}(\mathcal{L}))$ and $\mathcal{L} \in \mathbf{\Delta}^0_3(\mathcal{P}(\omega))$,*
2. *$\mathcal{L}$ is Alexandrov $\iff$ $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{L}))$,*
3. *$\mathcal{L}$ is scattered $\iff$ $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$ $\iff$ $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega))$,*
4. *$\mathcal{L}$ is scattered Alexandrov $\iff$ $\mathcal{L} \in \mathbf{\Delta}^0_2(\mathcal{A}(\mathcal{L}))$.*

Proof: (4) clearly follows from (2) and (3).

Proof of (2)*:* Assume $\mathcal{L}$ is Alexandrov. For each $L \in \mathcal{L}$ there is finite $F_L \subseteq \omega$ such that $L = C_{\mathcal{L}}(F)$. Thus, $\{L\} = (\downarrow_{\mathcal{A}(\mathcal{L})} L) \cap (\uparrow_{\mathcal{A}(\mathcal{L})} F_L)$, is a $\mathbf{\Sigma}^0_2$-set in $\mathcal{A}(\mathcal{L})$, hence

$$\mathcal{L} = \bigcup_{L \in \mathcal{L}} (\downarrow_{\mathcal{A}(\mathcal{L})} L) \cap (\uparrow_{\mathcal{A}(\mathcal{L})} F_L)$$

is in $\mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{L}))$.

For the converse, assume $\mathcal{L}$ is a countable T_D-space and $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{L}))$. Let $L \in \mathcal{L}$ be given. Then there are open U and closed A subsets of $\mathcal{A}(\mathcal{L})$ such that $L \in U \cap A \subseteq \mathcal{L}$ because $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{L}))$. We can assume that $U = \uparrow_{\mathcal{A}(\mathcal{L})} F$ for some finite $F \subseteq L$ and $A = \downarrow_{\mathcal{A}(\mathcal{L})} L$. Let F' be a finite tell-tale of L that contains F. If $X \in \mathcal{A}(\mathcal{L})$ is a subset of L and contains F', then it contains F so $X \in \mathcal{L}$, hence $X = L$ because F' is a finite tell-tale of L. Thus, $L = C_{\mathcal{L}}(F')$ is compact in $\mathcal{A}(\mathcal{L})$.

Proof of (1)*:* Since every singleton subset of $\mathcal{A}(\mathcal{L})$ is a $\mathbf{\Pi}^0_2$-set and $\mathcal{L}$ is countable, $\mathcal{L} = \bigcup_{L \in \mathcal{L}} \{L\}$ is a $\mathbf{\Sigma}^0_3$-subset of $\mathcal{A}(\mathcal{L})$. Thus, it only remains to show that $\mathcal{A}(\mathcal{L}) \setminus \mathcal{L}$ is in $\mathbf{\Sigma}^0_3(\mathcal{A}(\mathcal{L}))$. Let $\mathcal{K}$ be an Alexandrov concept space and $f: \mathcal{L} \to \mathcal{K}$ a continuous injection. $\mathcal{K} \in \mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{K}))$, so $\mathcal{A}(f)^{-1}(\mathcal{K}) \in \mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{L}))$ because $\mathcal{A}(f)$ is continuous. Thus, $\mathcal{A}(\mathcal{L}) \setminus \mathcal{A}(f)^{-1}(\mathcal{K}) \in \mathbf{\Sigma}^0_3(\mathcal{A}(\mathcal{L}))$. For each $K \in \mathcal{K}$, $\mathcal{A}(f)^{-1}(K) \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$, hence $\mathcal{A}(f)^{-1}(K) \setminus \{L_K\} \in \mathbf{\Sigma}^0_3(\mathcal{A}(\mathcal{L}))$, where L_K is the unique $L_K \in \mathcal{L}$ such that $f(L_K) = K$. It follows that

$$\mathcal{A}(\mathcal{L}) \setminus \mathcal{L} = \left(\bigcup_{K \in \mathcal{K}} \mathcal{A}(f)^{-1}(K) \setminus \{L_K\} \right) \cup \mathcal{A}(f)^{-1}(\mathcal{K})$$

is in $\mathbf{\Sigma}^0_3(\mathcal{A}(\mathcal{L}))$. Therefore, $\mathcal{L} \in \mathbf{\Delta}^0_3(\mathcal{A}(\mathcal{L}))$. To see that $\mathcal{L} \in \mathbf{\Delta}^0_3(\mathcal{P}(\omega))$, note that by Proposition 4.1.6 there is $A \in \mathbf{\Delta}^0_3(\mathcal{P}(\omega))$ such that $\mathcal{L} = A \cap \mathcal{A}(\mathcal{L})$. Since $\mathcal{A}(\mathcal{L}) \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega))$ by Lemma 4.3.6, it follows that $\mathcal{L}$ is the intersection of two $\mathbf{\Delta}^0_3$ subsets of $\mathcal{P}(\omega)$, hence $\mathcal{L} \in \mathbf{\Delta}^0_3(\mathcal{P}(\omega))$.

Proof of (3)*:* $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L})) \iff \mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega))$ follows from Proposition 4.1.6 and Lemma 4.3.6. Assume $\mathcal{L}$ is scattered. Let A be a Noetherian algebraic closure system and $f\colon \mathcal{L} \to A$ a continuous injection. Since $\mathcal{A}(A) = A$, $\mathcal{A}(f)\colon \mathcal{A}(\mathcal{L}) \to A$ is a continuous extension of f to $\mathcal{A}(\mathcal{L})$. Since A is a T_D-space, $\mathcal{A}(f)^{-1}(x) \in \mathbf{\Delta}^0_2(\mathcal{A}(\mathcal{L}))$ for each $x \in A$, and there is at most one $L \in \mathcal{L}$ such that $L \in \mathcal{A}(f)^{-1}(x)$ because f is injective. For each $L \in \mathcal{L}$, define

$$X_L = \{L\} \cup \left(\mathcal{A}(\mathcal{L}) \setminus \mathcal{A}(f)^{-1}(f(L))\right),$$

which is the union of a $\mathbf{\Pi}^0_2$-set and a $\mathbf{\Delta}^0_2$-set, hence $X_L \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$. Define

$$Z = \left(\bigcap_{L\in\mathcal{L}} X_L\right) \cap \mathcal{A}(f)^{-1}(f(\mathcal{L})),$$

which is in $\mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$ (note that $f(\mathcal{L})$, the image of $\mathcal{L}$ under f, is in $\mathbf{\Delta}^0_2(A)$ because every subset of A is $\mathbf{\Delta}^0_2$ by virtue of being a countable T_D-space).

Assume $L \in \mathcal{L}$. Clearly $L \in X_L$, and for any $L' \in \mathcal{L}$, if $L' \neq L$ then $f(L') \neq f(L)$, so $L \notin \mathcal{A}(f)^{-1}(f(L'))$, hence $L \in X_{L'}$. Thus, $\mathcal{L} \subseteq Z$.

Assume $X \in Z$. Since $X \in \mathcal{A}(f)^{-1}(f(\mathcal{L}))$, $\mathcal{A}(f)(X) = f(L)$ for some $L \in \mathcal{L}$. Since $X \in X_L$ and $X \notin \mathcal{A}(\mathcal{L}) \setminus \mathcal{A}(f)^{-1}(f(L))$, $X = L \in \mathcal{L}$. Thus, $Z \subseteq \mathcal{L}$.

Therefore, $\mathcal{L} = Z \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$.

For the converse, assume that $\mathcal{L}$ is a countable T_D-space and $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$. To show that $\mathcal{L}$ is scattered, it is sufficient to show that $\mathcal{L}^{(\alpha)}$ contains an isolated point (i.e., there is $L \in \mathcal{L}^{(\alpha)}$ such that $\{L\}$ is open in $\mathcal{L}^{(\alpha)}$) for each $\alpha \geq 0$. Clearly, if $\mathcal{L}^{(\alpha)}$ is finite then it contains an isolated point, so assume that $\mathcal{L}^{(\alpha)} = \{L_n \mid n \in \omega\}$ where $n \neq m$ implies $L_n \neq L_m$. Since $\mathcal{L}^{(\alpha)}$ is a closed subset of $\mathcal{L}$, it follows that $\mathcal{L}^{(\alpha)} \in \mathbf{\Pi}^0_2(\mathcal{A}(\mathcal{L}))$, so let

$$\mathcal{A}(\mathcal{L}) \setminus \mathcal{L}^{(\alpha)} = \bigcup_{i\in\omega} U_i \setminus V_i$$

for U_i, V_i open subsets of $\mathcal{A}(\mathcal{L})$.

Assume for a contradiction that $\mathcal{L}^{(\alpha)}$ has no isolated point. Let $F_{-1} = \emptyset$. For $n \geq 0$, if there is $L \in V_n \cap \mathcal{L}^{(\alpha)}$ such that $F_{n-1} \subseteq L$, then let $F \supseteq F_{n-1}$ be finite such that $L \in {\uparrow_{\mathcal{A}(\mathcal{L})}} F \subseteq V_n$. If there is no such L, then let $F = F_{n-1}$. Next, if $F \not\subseteq L_n$, then let $F_n = F$. If $F \subseteq L_n$, then let $F' \supseteq F$ be a finite tell-tale for L_n. Since $\{L_n\}$ is not isolated, there is $L' \in \mathcal{L}^{(\alpha)}$ containing F' and distinct from L_n. Clearly, $L' \not\subseteq L_n$ because F' is a finite tell-tale for L_n, so choose $x \in L' \setminus L_n$ and define $F_n = F' \cup \{x\}$. Thus $\{F_n\}_{n\in\omega}$ is an increasing chain of finite subsets of ω such that for each $n \in \omega$:

1. $F_n \not\subseteq L_n$,
2. There is $L \in \mathcal{L}^{(\alpha)}$ containing F_n,
3. If there is $L \in V_n \cap \mathcal{L}^{(\alpha)}$ containing F_n, then ${\uparrow_{\mathcal{A}(\mathcal{L})}} F_n \subseteq V_n$.

Since $\{F_n\}_{n\in\omega}$ is an increasing chain, $X = \bigcup_{n\in\omega} C_{\mathcal{L}}(F_n)$ is in $\mathcal{A}(\mathcal{L})$. Since $\forall n : F_n \not\subseteq L_n$, it is clear that $X \notin \mathcal{L}^{(\alpha)}$, so there is $i \in \omega$ such that $X \in U_i \setminus V_i$. Let $n \geq i$ be such that $X \in \uparrow_{\mathcal{A}(\mathcal{L})} C_{\mathcal{L}}(F_n) \subseteq U_i$ (such an n exists because U_i is Scott-open and $\{C_{\mathcal{L}}(F_n)\}_{n\in\omega}$ is a directed set with supremum in U_i). Since $\uparrow_{\mathcal{A}(\mathcal{L})} F_n = \uparrow_{\mathcal{A}(\mathcal{L})} C_{\mathcal{L}}(F_n)$, there is $L \in \mathcal{L}^{(\alpha)}$ such that $F_n \subseteq L$ hence $L \in U_i$. Thus, $L \in V_i$ and since $i \leq n$, $F_i \subseteq F_n$, so $L \in V_i \cap \mathcal{L}^{(\alpha)}$ and contains F_i. Therefore, $\uparrow_{\mathcal{A}(\mathcal{L})} F_i \subseteq V_i$, which implies $X \in V_i$, a contradiction.

Therefore, $L^{(\alpha)}$ must contain an isolated point. □

Note that Proposition 4.1.6 implies that $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{P}(\omega)) \Rightarrow \mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{A}(\mathcal{L}))$ and $\mathcal{L} \in \mathbf{\Delta}^0_2(\mathcal{P}(\omega)) \Rightarrow \mathcal{L} \in \mathbf{\Delta}^0_2(\mathcal{A}(\mathcal{L}))$, but the converses do not hold. For example, $\mathcal{L} = \{\omega\}$ is clearly scattered Alexandrov, hence $\{\omega\} \in \mathbf{\Delta}^0_2(\mathcal{A}(\{\omega\}))$, but $\{\omega\} \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega)) \setminus \mathbf{\Sigma}^0_2(\mathcal{P}(\omega))$.

4.3.5 Confidence and reliability

In this subsection, we will investigate requirements on how a learner should behave when given an information presentation for a concept that is not in the concept space the learner identifies.

Let $\mathcal{U}$ be a (possibly uncountable) concept space, and let $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$ be a representation of $\mathcal{U}$ that is $\mathbf{\Sigma}^0_1$-admissible with respect to a countably based topology on $\mathcal{U}$. We assume that $\mathcal{L} \subseteq \mathcal{U}$, and $\langle \mathcal{R}, \rho \rangle$ is the restriction of $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$ to $\mathcal{L}$ (i.e., $\mathcal{R} = \rho_{\mathcal{U}}^{-1}(\mathcal{L}) \subseteq \mathcal{R}_{\mathcal{U}}$ and $\rho = \rho_{\mathcal{U}}|_{\mathcal{R}}$, the restriction of $\rho_{\mathcal{U}}$ to $\mathcal{R}$). Let $\langle \mathcal{H}, h \rangle$ be a hypothesis space for $\mathcal{L}$.

We write $\mathcal{L}_\rho$ and $\mathcal{U}_\rho$ to emphasize that $\mathcal{L}$ and $\mathcal{U}$ are viewed as topological spaces induced by $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$, and $\mathcal{L}_h$ is $\mathcal{L}$ with the discrete topology induced by $\langle \mathcal{H}, h \rangle$.

Confident learners

A *confident* learner (see Jain et al. [25]) is a learner that converges no matter what information presentation it receives, even if it does not represent any concept in the concept space the learner is expected to identify. Confident identifiability of concept spaces can be formalized as follows.

Definition 4.3.8 *$\mathcal{L}$ is confidently identifiable within $\mathcal{U}$ (with respect to $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if there is a learner $\psi\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{H}^\omega$ such that $\psi|_{\mathcal{R}}$, the restriction of ψ to $\mathcal{R}$, identifies $\mathcal{L}$ and $\psi(R)$ converges for all $R \in \mathcal{R}_{\mathcal{U}}$.* □

Equivalently, $\mathcal{L}$ is confidently identifiable within $\mathcal{U}$ if and only if there is a learner $\psi\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{H}^\omega$ such that $\psi|_{\mathcal{R}}$ identifies $\mathcal{L}$ and $\lim_{\mathcal{H}} \circ \psi \colon\subseteq \mathcal{R}_{\mathcal{U}} \to \mathcal{H}$ is a *total* function. The following theorem completely characterizes confident identifiability.

Theorem 4.3.9 *$\mathcal{L}$ is confidently identifiable within $\mathcal{U}$ (with respect to $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if $\mathcal{L}_\rho$ is a countable T_D-space and $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{U}_\rho)$.*

Proof: Assume $\mathcal{L}$ is confidently identified by $\psi\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{H}^\omega$. It follows from Theorem 4.3.2 that $\mathcal{L}_\rho$ is a countable T_D-space. Since ψ is confident, $\lim_{\mathcal{H}} \circ \psi \colon \mathcal{R}_{\mathcal{U}} \to$

$\mathcal{H}$ is total, hence $f\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{L}_h$, defined as $f = h \circ \lim_{\mathcal{H}} \circ \psi$, is a total $\mathbf{\Sigma}^0_2$-measurable function. Let $\mathcal{S} = \{\bot, \top\}$ be the Sierpinski space. Every singleton subset of a countably based T_0-space is a $\mathbf{\Pi}^0_2$-set, hence $\chi_{\neg L}\colon \mathcal{U} \to \mathcal{S}$, defined as

$$\chi_{\neg L}(X) \quad = \quad \begin{cases} \bot & \text{if } X = L \ , \\ \top & \text{if } X \neq L, \end{cases}$$

is $\mathbf{\Sigma}^0_2$-measurable for each $L \in \mathcal{L}_h$. Thus, $\chi\colon \mathcal{U} \to \prod_{L\in\mathcal{L}_h} \mathcal{S}$ defined as $\chi(X)(L) = \chi_{\neg L}(X)$ is $\mathbf{\Sigma}^0_2$-measurable by Proposition 4.1.20 (here we are treating $\mathcal{L}_h$ as a subspace of ω to simplify notation). Define $g\colon \mathcal{R}_{\mathcal{U}} \to (\prod_{L\in\mathcal{L}_h} \mathcal{S}) \times \mathcal{L}_h$ by $g = \langle \chi \circ \rho_{\mathcal{U}}, f\rangle$. Clearly, g is $\mathbf{\Sigma}^0_2$-measurable because $\chi \circ \rho_{\mathcal{U}}$ and f are $\mathbf{\Sigma}^0_2$-measurable.

Intuitively, for each $R \in \mathcal{R}_{\mathcal{U}}$, $g(R)$ runs ψ on R and checks in parallel whether R is a representation of L for each $L \in \mathcal{L}$. Let $R \in \mathcal{R}_{\mathcal{U}}$ be given and assume that $g(R) = \langle \xi, L\rangle \in (\prod_{L\in\mathcal{L}_h} \mathcal{S}) \times \mathcal{L}_h$. If $R \in \mathcal{R}$, then since ψ identifies $\mathcal{L}$, $\rho_{\mathcal{U}}(R) = L$, hence $\xi(L) = \chi(\rho_{\mathcal{U}}(R))(L) = \chi_{\neg L}(\rho_{\mathcal{U}}(R)) = \bot$. If $R \notin \mathcal{R}$, then clearly $\rho_{\mathcal{U}}(R) \neq L$, hence $\xi(L) = \top$.

Next, we show that $\varepsilon\colon (\prod_{L\in\mathcal{L}_h} \mathcal{S}) \times \mathcal{L}_h \to \mathcal{S}$, defined as $\varepsilon(\langle \xi, L\rangle) = \xi(L)$, is continuous.[2] If we let $\pi_L\colon \prod_{L\in\mathcal{L}_h} \mathcal{S} \to \mathcal{S}$ be the L-th projection, then $\pi_L^{-1}(\top) = \{\xi \,|\, \xi(L) = \top\}$ is open because π_L is continuous. It follows that $U_L = (\pi_L^{-1}(\top)) \times \{L\}$ is open in $(\prod_{L\in\mathcal{L}_h} \mathcal{S}) \times \mathcal{L}_h$. Since $\varepsilon^{-1}(\top) = \bigcup_{L\in\mathcal{L}} U_L$ is open, ε is continuous.

It follows that $\varepsilon \circ g\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{S}$ is a $\mathbf{\Sigma}^0_2$-measurable function. By our construction, $\varepsilon(g(R)) = \bot \iff R \in \mathcal{R}$ holds, hence $\mathcal{R} \in \mathbf{\Pi}^0_2(\mathcal{R}_{\mathcal{U}})$. It follows from Corollary 4.1.31 that $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{U}_\rho)$.

To prove the converse, assume that $\mathcal{L} \in \mathbf{\Pi}^0_2(\mathcal{U}_\rho)$ and $\mathcal{L}_\rho$ is a countable T_D-space. Let $\mathcal{L}_h^\top = \mathcal{L}_h \cup \{\top\}$ have the topology defined as $U \subseteq \mathcal{L}_h^\top$ is open if and only if $U = \emptyset$ or else $\top \in U$. Define $f\colon \mathcal{U} \to \mathcal{L}_h^\top$ as $f(X) = \top$ if $X \notin \mathcal{L}$ and $f(X) = L$ if $X = L \in \mathcal{L}$. If $U \subseteq \mathcal{L}_h^\top$ is non-empty, then $f^{-1}(U) = (\mathcal{U} \setminus \mathcal{L}) \cup S$, where $S \subseteq \mathcal{L}_\rho$. Since $\mathcal{L}_\rho$ is a countable T_D-space, $S \in \mathbf{\Delta}^0_2(\mathcal{L}_\rho)$, so by Proposition 4.1.6 there is $W \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$ such that $S = W \cap \mathcal{L}$. $\mathcal{U} \setminus \mathcal{L}$ is in $\mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$ by assumption, thus $f^{-1}(U) = (\mathcal{U}\setminus\mathcal{L}) \cup W$ is in $\mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$, hence f is $\mathbf{\Sigma}^0_2$-measurable.

Let $\langle \mathcal{R}_\top, \rho_\top\rangle$ be a $\mathbf{\Sigma}^0_1$-admissible representation of $\mathcal{L}_h^\top$. By Theorem 4.1.43 there is a $\mathbf{\Sigma}^0_2$-measurable $F\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{R}_\top$ such that $f \circ \rho_{\mathcal{U}} = \rho_\top \circ F$.

Next define a multivalued function $g\colon \mathcal{L}_h^\top \rightrightarrows \mathcal{L}_h$ so that $g(L) = \{L\}$ for $L \in \mathcal{L}_h$ and $g(\top) = \mathcal{L}_h$. For every non-empty $U \subseteq \mathcal{L}_h$, $g^{-1}(U) = U \cup \{\top\}$, hence g is lower semicontinuous. Since $\mathcal{L}_h^\top$ is a countable T_D-space, Theorem 3.3.3 implies[3] that there is a continuous function $G\colon \mathcal{R}_\top \to \mathcal{H}$ such that $h(G(R)) \in g(\rho_\top(R))$ for every $R \in \mathcal{R}_\top$.

Define $p\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{L}_h$ by $p = h \circ G \circ F$. Then p is $\mathbf{\Sigma}^0_2$-measurable since G is continuous and F is $\mathbf{\Sigma}^0_2$-measurable. For each $R \in \mathcal{R}$ with $\rho_{\mathcal{U}}(R) = L \in \mathcal{L}$,

$$\rho_\top(F(R)) = f(\rho_{\mathcal{U}}(R)) = f(L) = L,$$

[2]The reader should note that $\prod_{L\in\mathcal{L}_h} \mathcal{S}$ is homeomorphic to $\mathcal{S}^{\mathcal{L}_h}$, so this is essentially just a special case of Lemma 3.1.19.

[3]Although Theorem 3.3.3 is phrased in terms of concept spaces with the Π-topology and texts, the fact texts form a $\mathbf{\Sigma}^0_1$-admissible representation (Theorem 3.1.11) and that every countably based T_0-space is homeomorphic to a concept space with the Π-topology shows that the result applies to all countably based T_0-spaces with texts replaced by arbitrary $\mathbf{\Sigma}^0_1$-admissible representations.

thus

$$p(R) = h(G(F(R))) \in g(\rho_\top(F(R))) = g(L) = \{L\},$$

hence $p(R) = L = \rho_{\mathcal{U}}(R)$.

Finally, since p is $\mathbf{\Sigma}^0_2$-measurable and since $\langle \mathcal{H}^\omega, h \circ \lim_{\mathcal{H}} \rangle$ is a $\mathbf{\Sigma}^0_2$-admissible representation of $\mathcal{L}_h$ by Theorem 4.3.1, there is continuous $\psi\colon \mathcal{R}_{\mathcal{U}} \to \mathcal{H}^\omega$ such that $p = h \circ \lim_{\mathcal{H}} \circ \psi$. Since p is a total function, $\lim_{\mathcal{H}} \circ \psi$ is total, thus ψ converges on every $R \in \mathcal{R}_{\mathcal{U}}$. Clearly, $\psi(R)$ converges to a hypothesis for $p(R) = \rho_{\mathcal{U}}(R)$ for each $R \in \mathcal{R}$, thus ψ confidently identifies $\mathcal{L}$ within $\mathcal{U}$. □

Corollary 4.3.10 *If $\mathcal{L}_\rho$ is confidently identifiable within $\mathcal{U}_\rho$ and $\mathcal{J}_\rho \subseteq \mathcal{L}_\rho$, then $\mathcal{J}_\rho$ is confidently identifiable within $\mathcal{U}_\rho$.*

Proof: $\mathcal{J}_\rho \in \mathbf{\Pi}^0_2(\mathcal{L}_\rho)$ because $\mathcal{L}_\rho$ is a countable T_D-space, so by Proposition 4.1.6 there is $A \in \mathbf{\Pi}^0_2(\mathcal{U}_\rho)$ such that $\mathcal{J}_\rho = A \cap \mathcal{L}_\rho$. Since $\mathcal{L}_\rho \in \mathbf{\Pi}^0_2(\mathcal{U}_\rho)$, it follows that $\mathcal{J}_\rho \in \mathbf{\Pi}^0_2(\mathcal{U}_\rho)$. □

Ambainis et al. [2] proved that $\mathcal{L}$ is confidently identifiable from positive data if and only if $\mathcal{L}$ is identifiable with a mind-change bound. This result follows from Theorem 4.3.9 and Theorem 4.3.7 by taking $\mathcal{U} = \mathcal{P}(\omega)$ and $\mathcal{R}_{\mathcal{U}} = \mathcal{T}(\mathcal{P}(\omega))$.

Corollary 4.3.11 (Ambainis et al. [2]) *$\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive data if and only if $\mathcal{L}$ is scattered with respect to the* Π-*topology.* □

In the case of identification from positive and negative data, note that $\mathcal{P}(\omega)$ with the informant topology is homeomorphic to 2^ω, so the $\mathbf{\Pi}^0_2$-subsets of $\mathcal{P}(\omega)$ (with the informant topology) are precisely the zero-dimensional Polish spaces. Therefore, $\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data if and only if $\mathcal{L}$ is a countable zero-dimensional Polish space (with respect to the informant topology). Based on this observation, we can show that the above result by Ambainis et al. [2] also applies to identification from positive and negative data.

Corollary 4.3.12 *$\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data if and only if $\mathcal{L}$ is scattered with respect to the informant topology.*

Proof: If $\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data then it is a countable Polish space therefore it is scattered (see Section 6 in [29]).

For the converse, assume $\mathcal{L}$ is scattered with respect to the informant topology. Let $e\colon \mathcal{P}(\omega)_I \to \mathcal{P}(\omega)_\Pi$ be a topological embedding of $\mathcal{P}(\omega)$ with the informant topology into $\mathcal{P}(\omega)$ with the Π-topology. Then $e(\mathcal{L})$, being homeomorphic to $\mathcal{L}$, is scattered, so $e(\mathcal{L}) \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega)_\Pi)$ by Theorem 4.3.7. Therefore, $\mathcal{L} = e^{-1}(e(\mathcal{L}))$ is in $\mathbf{\Pi}^0_2(\mathcal{P}(\omega)_I)$ because e is continuous. It follows that $\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data. □

Next we give a sufficient condition for confident identifiability from positive and negative data based only on the set-theoretical structure of $\mathcal{L}$.

Corollary 4.3.13 *If $\mathcal{A}(\mathcal{L})$ is countable, then $\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data.*

Proof: Let $id: \mathcal{P}(\omega)_I \to \mathcal{P}(\omega)_\Pi$ be the identity function from $\mathcal{P}(\omega)$ with the informant topology to $\mathcal{P}(\omega)$ with the Π-topology. Clearly, id is continuous. Since $\mathcal{A}(\mathcal{L}) \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega)_\Pi)$ holds by Lemma 4.3.6, $f^{-1}(\mathcal{A}(\mathcal{L})) = \mathcal{A}(\mathcal{L}) \in \mathbf{\Pi}^0_2(\mathcal{P}(\omega)_I)$. Since $\mathcal{A}(\mathcal{L})$ is countable, it follows that $\mathcal{A}(\mathcal{L})$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data, therefore $\mathcal{L}$ is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data. □

The converse does not hold. For example, $\mathcal{COSINGLE} = \{\omega \setminus \{n\} \mid n \in \omega\}$ is easily seen to be confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data, but $\mathcal{A}(\mathcal{COSINGLE}) = \mathcal{P}(\omega)$ is uncountable. The concept space

$$\omega + 1 = \{\{n \mid n < m\} \mid m \in \omega\} \cup \{\omega\}$$

satisfies $\mathcal{A}(\omega + 1) = \omega + 1$, so it is confidently identifiable within $\mathcal{P}(\omega)$ from positive and negative data, even though $\omega + 1$ is not even identifiable from positive data alone.

Reliable learners

Reliable learners were introduced and investigated by Sakurai [44]. A reliable learner is opposite of a confident learner, in that it converges *only* on information presentations that are for concepts in the concept space it identifies.

Definition 4.3.14 *$\mathcal{L}$ is reliably identifiable within $\mathcal{U}$ (with respect to $\langle \mathcal{R}_\mathcal{U}, \rho_\mathcal{U} \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if there is a learner $\psi: \mathcal{R}_\mathcal{U} \to \mathcal{H}^\omega$ such that $\psi|_\mathcal{R}$ identifies $\mathcal{L}$ and for all $R \in \mathcal{R}_\mathcal{U}$, $\psi(R)$ converges if and only if $\rho_\mathcal{U}(R) \in \mathcal{L}$.* □

Equivalently, $\mathcal{L}$ is reliably identifiable within $\mathcal{U}$ if and only if there is a learner $\psi: \mathcal{R}_\mathcal{U} \to \mathcal{H}^\omega$ such that $\psi|_\mathcal{R}$ identifies $\mathcal{L}$ and $dom(\lim_\mathcal{H} \circ \psi) = \mathcal{R}$. The following theorem completely characterizes reliable identifiability.

Theorem 4.3.15 *$\mathcal{L}$ is reliably identifiable within $\mathcal{U}$ (with respect to $\langle \mathcal{R}_\mathcal{U}, \rho_\mathcal{U} \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if $\mathcal{L}_\rho$ is a countable T_D-space and $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$.*

Proof: Assume $\psi: \mathcal{R}_\mathcal{U} \to \mathcal{H}^\omega$ reliably identifies $\mathcal{L}$ within $\mathcal{U}$. Then clearly $\mathcal{L}_\rho$ is a countable T_D-space. Define $f: \mathcal{H}^\omega \to \mathcal{S}$ by $f(\xi) = \top$ if and only if $\exists m \forall n \geq m : \xi(n) = \xi(m)$. It is easily seen that f is $\mathbf{\Sigma}^0_2$-measurable, and $f \circ \psi(R) = \top$ if and only if $\psi(R)$ converges if and only if $R \in \mathcal{R}$. Therefore, $\mathcal{R} = (f \circ \psi)^{-1}(\top) \in \mathbf{\Sigma}^0_2(\mathcal{R}_\mathcal{U})$, so $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$ by Corollary 4.1.31.

For the converse, assume that $\mathcal{L}_\rho$ is a countable T_D-space and $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$. Let $\mathcal{L}_h^\perp = \mathcal{L}_h \cup \{\perp\}$ have the topology defined as $U \subseteq \mathcal{L}_h^\perp$ is open if and only if $U = \mathcal{L}_h^\perp$ or else $\perp \notin U$. Define $f: \mathcal{U} \to \mathcal{L}_h^\perp$ as $f(X) = \perp$ if $X \notin \mathcal{L}$ and $f(X) = L$ if $X = L \in \mathcal{L}$. If $U \subseteq \mathcal{L}_h^\perp$ is open and $U \neq \mathcal{L}_h^\perp$ then $\perp \notin U$ so $f^{-1}(U) \subseteq \mathcal{L}$. Thus, $f^{-1}(U) \in \mathbf{\Sigma}^0_2(\mathcal{L}_\rho)$ because $\mathcal{L}_\rho$ is a countable T_D-space, hence $f^{-1}(U) \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$ because $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$. Therefore, f is $\mathbf{\Sigma}^0_2$-measurable. It easily follows from Theorem 4.2.6 that $\lim_\perp: \mathcal{H}^\omega \to \mathcal{L}_h^\perp$, defined as

$$\lim_\perp(\xi) = \begin{cases} h(\lim_\mathcal{H}(\xi)) & \text{if } \xi \text{ converges,} \\ \perp & \text{otherwise,} \end{cases}$$

is a $\mathbf{\Sigma}^0_2$-admissible representation of $\mathcal{L}_h^\perp$. Therefore, there is continuous $\psi: \mathcal{R}_\mathcal{U} \to \mathcal{H}^\omega$ such that $f \circ \rho_\mathcal{U} = \lim_\perp \circ \psi$. For each $R \in \mathcal{R}$ with $\rho_\mathcal{U}(R) = L \in \mathcal{L}$,

$f(\rho_{\mathcal{U}}(R)) = f(L) = L = \lim_{\perp}(\psi(R))$, so $\psi(R)$ converges to a hypothesis for L. If $R \notin \mathcal{R}$, then $f(\rho_{\mathcal{U}}(R)) = \perp = \lim_{\perp}(\psi(R))$, so $\psi(R)$ diverges. Therefore, ψ reliably identifies $\mathcal{L}$ within $\mathcal{U}$. $\square$

Corollary 4.3.16 *If $\mathcal{L}_\rho$ is reliably identifiable within $\mathcal{U}_\rho$ and $\mathcal{J}_\rho \subseteq \mathcal{L}_\rho$, then $\mathcal{J}_\rho$ is reliably identifiable within $\mathcal{U}_\rho$.*

Proof: $\mathcal{J}_\rho \in \mathbf{\Sigma}^0_2(\mathcal{L}_\rho)$ because $\mathcal{L}_\rho$ is a countable T_D-space, so by Proposition 4.1.6 there is $A \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$ such that $\mathcal{J}_\rho = A \cap \mathcal{L}_\rho$. Since $\mathcal{L}_\rho \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$, it follows that $\mathcal{J}_\rho \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$. $\square$

It immediately follows from Theorem 4.3.15 and Theorem 4.3.7 that if $\mathcal{L}$ is reliably identifiable within $\mathcal{P}(\omega)$ from positive data, then $\mathcal{L}$ is an Alexandrov space. Sakurai [44] gave a necessary and sufficient condition.

Corollary 4.3.17 (Sakurai [44]) *$\mathcal{L}$ is reliably identifiable within $\mathcal{P}(\omega)$ from positive data if and only if every concept in $\mathcal{L}$ is a finite set.*

Proof: If every concept in $\mathcal{L}$ is finite then $\mathcal{L}$ is clearly a countable T_D-space. It is easily seen that $\{F\} \in \mathbf{\Delta}^0_2(\mathcal{P}(\omega))$ for every finite $F \subseteq \omega$. Therefore, $\mathcal{L}$ is the countable union of $\mathbf{\Delta}^0_2$-sets, hence $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{P}(\omega))$.

For the converse, assume that $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{P}(\omega))$. Then for any $L \in \mathcal{L}$ there are Π-open sets $U, V \subseteq \mathcal{P}(\omega)$ such that $L \in U \setminus V \subseteq \mathcal{L}$. We can assume without loss of generality that $U = \uparrow_{\mathcal{P}(\omega)} F$ for some finite $F \subseteq \omega$, and that $V = \downarrow_{\mathcal{P}(\omega)} L$. If L is infinite, then for any finite $G \subseteq L$, $L' = F \cup G$ is a strict subset of L containing G, and $L' \in U \setminus V$, so G is not a finite tell-tale for L. It follows that if $\mathcal{L}$ contains an infinite concept, then $\mathcal{L}$ is not a T_D-space. $\square$

Corollary 4.3.18 *If $\mathcal{U}_\rho$ is a (possibly uncountable) T_D-space, then every countable $\mathcal{L} \subseteq \mathcal{U}$ is reliably identifiable within $\mathcal{U}$ (with respect to $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$).*

Proof: Since every singleton subset of $\mathcal{U}_\rho$ is a $\mathbf{\Delta}^0_2$-set, if $\mathcal{L}$ is countable then it is the countable union of $\mathbf{\Delta}^0_2$ sets, hence $\mathcal{L} \in \mathbf{\Sigma}^0_2(\mathcal{U}_\rho)$. $\square$

We immediately obtain Sakurai's characterization of reliable identification from positive and negative data.

Corollary 4.3.19 (Sakurai [44]) *$\mathcal{L}$ is reliably identifiable within $\mathcal{P}(\omega)$ from positive and negative data if and only if $\mathcal{L}$ is countable.* $\square$

Refuting learners

Refutable identification was introduced by Mukouchi and Arikawa [41]. In this case, the learner is required to identify $\mathcal{L}$ in the limit, but in addition, if given an information presentation of a concept not in $\mathcal{L}$, then the learner must halt within finite time and display an error. This means that the learner must be able to observe the property that $R \in \mathcal{R}_{\mathcal{U}} \setminus \mathcal{R}$ within finite time, which means that $\mathcal{R}$ must be a closed subset of $\mathcal{R}_{\mathcal{U}}$, hence $\mathcal{L}$ must be a closed subset of $\mathcal{U}$.

Theorem 4.3.20 *$\mathcal{L}$ is refutably identifiable within $\mathcal{U}$ (with respect to $\langle \mathcal{R}_{\mathcal{U}}, \rho_{\mathcal{U}} \rangle$ and $\langle \mathcal{H}, h \rangle$) if and only if $\mathcal{L}_\rho$ is a countable T_D-space and $\mathcal{L}$ is a closed subset of $\mathcal{U}_\rho$.* $\square$

The following example shows how the results in this section apply to non-traditional concept spaces.

Example 4.3.21 Let $\mathbb{Q}$ be the set of rational numbers viewed as a subspace of the real numbers, $\mathbb{R}$, with the Euclidean topology. Since $\mathbb{R}$ is a metric space, it is a T_D-space, hence $\mathbb{Q}$ is a countable T_D-space. It is well known that $\mathbb{Q}$ is in $\mathbf{\Sigma}^0_2(\mathbb{R}) \setminus \mathbf{\Pi}^0_2(\mathbb{R})$. Therefore, with respect to any $\mathbf{\Sigma}^0_1$-admissible representation of $\mathbb{R}$, $\mathbb{Q}$ is reliably identifiable within $\mathbb{R}$ but neither confidently identifiable nor refutably identifiable within $\mathbb{R}$. Note that this also applies if we replace $\mathbb{R}$ with any perfect Polish space and $\mathbb{Q}$ with any countable dense subset (see Exercise 8.7 in [29]). The natural numbers, $\mathbb{N}$, are a closed subset of $\mathbb{R}$, so they are refutably identifiable within $\mathbb{R}$. $\square$

Chapter 5

Conclusion

We have explored many algebraic and topological aspects of algorithmic learning theory.

In Chapter 2 we introduced an algebraic closure operator on concept spaces which we used to give a characterization of the mind-change complexity of unbounded unions of restricted pattern languages.

In Chapter 3 we showed many connections between well known structural properties of concept spaces and well known topological properties. We also gave a complete topological characterization of strong and weak reducibility between concept spaces.

In Chapter 4 we introduced a very general class of representations of spaces and characterized the functions between spaces that can be realized. We then used these results to place several variations of identification in the limit as well as classification in the limit within a common topological framework.

Some important future work would be to model scenarios where the learner interacts with the environment and to introduce probability into our framework.

An example of an interactive learner is a *query learner* (see [5]). A query learner asks a teacher questions about an unknown concept, and produces hypotheses based on the answers it receives. Such a learner is essentially doing two tasks at once: querying and inferring. The querying aspect can be modeled as a function $q\colon (\mathcal{Q} \times \mathcal{A})^{<\omega} \to \mathcal{Q}$ that chooses a query from $\mathcal{Q}$ based on the finite sequences of answers (elements of $\mathcal{A}$) that it received from previous queries. If q's queries are answered with respect to a particular concept $L \in \mathcal{L}$, then q determines an infinite sequence in $(\mathcal{Q} \times \mathcal{A})^{\omega}$ which can be interpreted as a particular representation of L. In this way, q determines a representation of $\mathcal{L}$ in terms of infinite sequences of query-answer pairs. The inferring aspect of a query learner can then be interpreted as a learner ψ in our sense with respect to the representation generated by q. The main issue then is whether or not a querier q exists that can generate a representation in a way that a learner ψ can identify $\mathcal{L}$.

If we assume that $\mathcal{Q}$ and $\mathcal{A}$ are countable, and that for each concept $L \in \mathcal{L}$ and query $Q \in \mathcal{Q}$ there is a unique answer $A \in \mathcal{A}$ to the query Q about L, then each $L \in \mathcal{L}$ can be viewed as a function $L\colon \mathcal{Q} \to \mathcal{A}$, hence $\mathcal{L}$ is naturally interpreted as a subspace of the exponential object $\mathcal{A}^{\mathcal{Q}}$ (where $\mathcal{Q}$ and $\mathcal{A}$ have discrete topologies). If q is defined to always ask every possible query in $\mathcal{Q}$, then q will determine a representation of $\mathcal{L}$ that is admissible with respect to the topology $\mathcal{L}$ inherits as a subspace of $\mathcal{A}^{\mathcal{Q}}$. Therefore, all of our results will apply to identifying $\mathcal{L}$ in this situtation.

Probability is an important tool in learning theory that we have not addressed. One application is to investigate identification with high probabilites. For example, we might assume that each $L \in \mathcal{L}$ determines a probabilistic measure μ_L on the Borel subsets of $\mathcal{T}(\mathcal{L})$ determining the probability that particular texts T will be presented given L is the concept to be identified. Then given any learner $\psi\colon \mathcal{T}(\mathcal{L}) \to \mathcal{H}^\omega$, and assuming L is the concept to be identified, the probability that ψ converges to a hypothesis for L is $\mu_L((h \circ \lim_{\mathcal{H}} \circ \psi)^{-1}(L))$. We could then investigate which scenarios would allow a ψ to exist that could identify each concept with some given probability. It would be interesting to see how allowing some probability of error affects ψ's ability to realize non-continuous functions of varying complexity.

Another application is to investigate learners that must identify or in some sense work with probability distributions. Some investigation into $\mathbf{\Sigma}^0_1$-admissible representations of probability measures in terms of probabilistic processes has been done by [50] and [49]. $\mathbf{\Sigma}^0_2$-admissible representations may be useful for representing probability distributions by random samples. For example, define $\delta\colon\subseteq 2^\omega \to [0,1]$ by

$$\delta(\xi) = p \iff \lim_{n\to\infty} \Big(\frac{1}{n}\sum_{i=0}^{n} \xi(i)\Big) = p$$

Assume $X_p^{(0)}, X_p^{(1)}, \dots$ is a countable sequence of independently drawn random samples with $X_p \in \{0,1\}$, $P(X_p = 1) = p$, and $P(X_p = 0) = 1 - p$. By the strong law of large numbers, we have that $\delta(\langle X_p^{(0)}, X_p^{(1)}, \dots\rangle) = p$ almost surely (i.e. with probability 1). This justifies our interpretation of δ as a representation of probability distributions by samples. Although δ is *not* a $\mathbf{\Sigma}^0_1$-admissible representation, it is a $\mathbf{\Sigma}^0_2$-admissible representation of $[0,1]$ with the Euclidean topology.

Bibliography

[1] S. Abramsky, *Domain theory in logical form*, Annals of Pure and Applied Logic **51** (1991), 1–77.

[2] A. Ambainis, S. Jain, and A. Sharma, *Ordinal mind change complexity of language identification*, Theoretical Computer Science **220** (1999), 323–343.

[3] D. Angluin, *Inductive inference of formal languages from positive data*, Information and Control **45** (1980), 117–135.

[4] ______, *Inference of reversible languages*, Journal of the ACM **29** (1982), 741–765.

[5] D. Angluin, *Queries and concept learning*, Machine Learning **2** (1988), 319–342.

[6] C. E. Aull and W. J. Thron, *Separation axioms between T_0 and T_1*, Indag. Math. **24** (1963), 26–37.

[7] I. Battenfeld, M. Schröder, and A. Simpson, *A convenient category of domains*, Electr. Notes Theor. Comput. Sci. **172** (2007), 69–99.

[8] V. Brattka, *Effective Borel measurability and reducibility of functions*, Mathematical Logic Quarterly **51** (2005), 19–44.

[9] V. Brattka and M. Makananise, *Limit computable functions and subsets*, (to appear).

[10] D. J. Brown and R. Suszko, *Abstract logics*, Dissertationes Mathematicae **102** (1973), 9–41.

[11] S. Burris and H. P. Sankappanavar, *A course in universal algebra*, Springer-Verlag, 1981.

[12] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms*, second ed., Springer-Verlag, 1996.

[13] M. de Brecht, M. Kobayashi, H. Tokunaga, and A. Yamamoto, *Inferability of closed set sytems from positive data*, Joint JSAI 2006 Workshop Post-Proceedings, Lecture Notes in Artificial Intelligence, vol. 4384, Springer, 2007, pp. 265–275.

[14] M. de Brecht and A. Yamamoto, *Mind change complexity of inferring unbounded unions of pattern languages from positive data*, Proc. ALT '06, Lecture Notes in Artificial Intelligence, vol. 4262, Springer, 2006, pp. 124–138.

[15] ______, *Topological properties of concept spaces*, Proc. ALT '08, Lecture Notes in Artificial Intelligence, vol. 5254, Springer, 2008, pp. 374–388.

[16] ______, *$\mathbf{\Sigma}^0_\alpha$-admissible representations*, Sixth International Conference on Computability and Complexity in Analysis (CCA 2009), 2009.

[17] R. Engelking, *General topology*, Heldermann, 1989.

[18] R. Freivalds and C. H. Smith, *On the role of procrastination for machine learning*, Information and Computation **107** (1993), 237–271.

[19] J. H. Gallier, *What's so special about Kruskal's theorem and the ordinal γ_0? a survey of some results in proof theory*, Annals of Pure and Applied Logic **53** (1991), 199–260.

[20] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. W. Mislove, and D. W. Scott, *Continuous lattices and domains*, Cambridge University Press, 2003.

[21] E. M. Gold, *Language identification in the limit*, Information and Control **10** (1967), 447–474.

[22] R. Hasegawa, *Well-ordering of algebras and Kruskal's theorem*, Logic, Language and Computation, Lecture Notes in Computer Science, vol. 792, Springer, 1994, pp. 133–172.

[23] G. Higman, *Ordering by divisibility in abstract algebras*, Proceedings of the London Mathematical Society, Third Series, vol. 2, 1952, pp. 326–336.

[24] S. Jain, E. Kinber, and R. Wiehagen, *Language learning from texts: Degrees of intrinsic complexity and their characterizations*, Journal of Computer and System Sciences **63** (2001), 305–354.

[25] S. Jain, D. Osherson, J. Royer, and A. Sharma, *Systems that learn: An introduction to learning theory*, second ed., MIT Press, 1999.

[26] S. Jain and A. Sharma, *Elementary formal systems, intrinsic complexity, and procrastination*, Proceedings of COLT '96, 1996, pp. 181–192.

[27] S. Jain and A. Sharma, *The intrinsic complexity of language identification*, Journal of Computer and System Sciences **52** (1996), 393–402.

[28] P. T. Johnstone, *Stone spaces*, Cambridge University Press, 1982.

[29] A. Kechris, *Classical descriptive set theory*, Springer-Verlag, 1995.

[30] J. Kelley, *General topology*, Springer-Verlag, 1975.

[31] K. Kelly, *The logic of reliable inquiry*, Oxford University Press, 1996.

[32] S. Kobayashi, *Approximate identification, finite elasticity and lattice structure of hypothesis space*, Technical Report, CSIM 96-04, Dept. of Compt. Sci. and Inform. Math., Univ. of Electro- Communications, 1996.

[33] K. Kunen, *Set theory: An introduction to independence proofs*, North Holland Publishing Company, 1980.

[34] W. Luo and O. Schulte, *Mind change efficient learning*, Information and Computation **204** (2006), 989–1011.

[35] S. Mac Lane, *Categories for the working mathematician*, Springer-Verlag, 1971.

[36] E. Martin, A. Sharma, and F. Stephan, *Unifying logic, topology and learning in parametric logic*, Theoretical Computer Science **350** (2006), 103–124.

[37] T. Motoki, T. Shinohara, and K. Wright, *The correct definition of finite elasticity: Corrigendum to identification of unions*, Proc. COLT '91, 1991, p. 375.

[38] Y. Mukouchi, *Characterization of finite identification*, Proceedings of the International Workshop on Analogical and Inductive Inference, Lecture Notes in Computer Science, vol. 642, Springer, 1992, pp. 260–267.

[39] ———, *Containment problems for pattern languages*, IEICE Trans. Inform. Systems **E75-D** (1992), 420–425.

[40] ———, *Inductive inference of an approximate concept from positive data*, Proc. ALT '94, Lecture Notes in Artificial Intelligence, vol. 872, Springer, 1994, pp. 484–499.

[41] Y. Mukouchi and S. Arikawa, *Inductive inference machines that can refute hypothesis spaces*, Proc. ALT '93, Lecture Notes in Artificial Intelligence, vol. 744, Springer, 1993, pp. 123–137.

[42] G. Plotkin, *Domains*, The "Pisa notes", 1983.

[43] J. Saint Raymond, *Preservation of the Borel class under countable-compact-covering mappings*, Topology and its Applications **154** (2007), 1714–1725.

[44] A. Sakurai, *Inductive inference of formal languages from positive data enumerated primitive-recursively*, Proc. 2nd Int. Workshop on Algorithmic Learning Theory, IOS Press, 1992, pp. 73–83.

[45] M. Sato and T. Moriyama, *Inductive inference of length bounded EFS's from positive data*, Technical Report DMSIS-RR-94-2, Department of Mathematical Sciences and Information Sciences, University of Osaka, 1994.

[46] M. Sato, Y. Mukouchi, and D. Zheng, *Characteristic sets for unions of regular pattern languages and compactness*, Proc. ALT '98, Lecture Notes in Artificial Intelligence, vol. 1501, Springer, 1998, pp. 220–233.

[47] M. Sato and K. Umayahara, *Inductive inferability for formal languages from positive data*, IEICE Trans. Inf. and Syst. **E75-D(4)** (1992), 415–419.

[48] M. Schröder, *Extended admissibility*, Theoretical Computer Science **284** (2002), 519–538.

[49] ______, *Admissible representations of probability measures*, Electronic Notes in Theoretical Computer Science **167** (2007), 61–78.

[50] M. Schröder and A. Simpson, *Representing probability measures using probabilistic processes*, Journal of Complexity **22** (2006), 768–782.

[51] D. Scott, *Domains for denotational semantics*, Automata, Languages and Programming, 1982, Lecture Notes in Computer Science, vol. 140, Springer, 1982, pp. 577–613.

[52] V. Selivanov, *Towards a descriptive set theory for domain-like structures*, Theoretical Computer Science **365** (2006), 258–282.

[53] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), 379–423, 623–656.

[54] T. Shinohara and H. Arimura, *Inductive inference of unbounded unions of pattern languages from positive data*, Theoretical Computer Science **241** (2000), 191–209.

[55] S. G. Simpson, *Ordinal numbers and the Hilbert basis theorem*, Journal of Symbolic Logic **53** (1988), 961–974.

[56] M. Smyth, *Topology*, Handbook of Logic in Computer Science, vol. 1, Oxford University Press, 1992.

[57] F. Stephan and Y. Ventsov, *Learning algebraic structures from text*, Theoretical Computer Science **268** (2001), 221–273.

[58] A. Tang, *Chain properties in $P(\omega)$*, Theoretical Computer Science **9** (1979), 153–172.

[59] K. Weihrauch, *Computable analysis*, Springer-Verlag, 2000.

[60] K. Weihrauch and C. Kreitz, *Representations of the real numbers and of the open subsets of the set of real numbers*, Annals of Pure and Applied Logic **35** (1987), 247–260.

[61] K. Wright, *Identification of unions of languages drawn from an identifiable class*, Proceedings of COLT '89, 1989, pp. 328–333.

[62] M. Ziegler, *Revising type-2 computation and degrees of discontinuity*, Electronic Notes in Theoretical Computer Science **167** (2007), 255–274.

Index