Refuting Learning Revisited

Wolfgang Merkle [1] and Frank Stephan [2]

Mathematisches Institut, Universität Heidelberg
69120 Heidelberg, Germany, EU

Abstract

We consider, within the framework of inductive inference, the concept of refuting learning as introduced by Mukouchi and Arikawa, where the learner is not only required to learn all concepts in a given class but also has to explicitly refute concepts outside the class.

In the first part of the paper, we consider learning from text and introduce a concept of limit-refuting learning that is intermediate between refuting learning and reliable learning. We give characterizations for these concepts and show some results about their relative strength and their relation to confident learning.

In the second part of the paper we consider learning from texts that for some k contain all positive Π_k-formulae that are valid in the standard structure determined by the set to be learned. In this model, the following results are shown. For the language with successor, any countable axiomatizable class can be limit-refuting learned from Π_1-texts. For the language with successor and order, any countable axiomatizable class can be reliably learned from Π_1-texts and can be limit-refuting learned from Π_2-texts, whereas the axiomatizable class of all finite sets cannot be limit-refuting learned from Π_1-texts. For the full language of arithmetic, which contains in addition plus and times, for any even k there is an axiomatizable class that can be limit-refuting learned from Π_{k+1}-texts but not from Π_k-texts. A similar result with $k+3$ in place of $k+1$ holds with respect to the language of Presburger's arithmetic.

Key words: Refuting learning; Inductive Inference; Axiomatizable Classes; First Order Logic; Recursion Theory.

[1] Email: `merkle@math.uni-heidelberg.de`.
[2] Email: `fstephan@math.uni-heidelberg.de`. Supported by the Deutsche Forschungsgemeinschaft (DFG) under Heisenberg grant Ste 967/1–1.

1 Introduction

Inductive Inference [12,25] studies, on an abstract level, the phenomenon of learning. Gold [7] introduced the following basic formalization of a learning situation. The objects to be learned are the sets within a given class of recursively enumerable sets. The learner has to identify each set in this class by converging to a hypothesis that describes the set uniquely while observing longer and longer prefixes of any text for this set. A learner converges if it changes its hypothesis at most finitely often, a text for a set is any sequence that contains all elements but no non-elements of the set, and usually hypotheses are indices with respect to some fixed acceptable numbering of the partial recursive functions (equivalently one could use grammars or programs enumerating the members of the set to be learned).

Gold [7] demonstrated that it is impossible to learn the class of all recursively enumerable sets. This restriction holds for topological as well as for recursion theoretical reasons.

(a) For any learner that learns all finite sets and for any infinite set A, there is a text for A on which the learner diverges.
(b) The class of all graphs of computable functions cannot be learned by a computable learner — indeed, Adleman and Blum [1] quantified the problem of learning this class by showing that learning the class requires access to an oracle of high Turing degree.

The topological and computational aspects of learning interact. Gold [7] considered models of learning where in place of arbitrary texts, the learner just receives texts that can be computed in some fixed computation model. Gold showed that a computable learner can learn all recursively enumerable sets from primitive recursive texts (by simply identifying the primitive recursive function that generates the text) while, on the other hand, the collection of all recursive texts is already so complex that a computable learner cannot learn the class of all recursively enumerable sets from recursive texts.

In the present work, the power of learners is not enlarged by restricting texts to computationally simple ones but by increasing their information content. While standard texts essentially just list the elements of the set to be learned, we consider texts that contain positive formulae that are true for the set to be learned. The consideration of such more informative texts relates to the fact that we consider a model of learning where the learner has to recognize and to explicitly refute data-sequences belonging to sets that are not learned. This model is rather restrictive in a setting of standard texts and allows just the learning of classes of finite sets. The model becomes more powerful in the setting where the texts contain formulae and in this setting, we will investigate into the question which kind of classes can be learned from what types of formulae.

Mukouchi and Arikawa [22,23] introduced the learning model where the learner has to refute data-sequences that belong to sets that are not in the class to be learned. Their model is a sharpened version of Minicozzi's reliable learning, where the learner either converges to a correct index or diverges. In the model of Mukouchi and Arikawa, instead of diverging, the learner has to give an explicit refutation signal after a finite number of steps.

Lange and Watson [17], Jantke [14], and Jain [10] considered variants of refuting learning where a learner M for a class $\mathcal{C}$ of sets is not required to refute unless the input text T satisfies both of the following conditions.

- M does not infer the concept to which T belongs.
- There is a prefix $\sigma \preceq T$ such that no data-sequence of any concept in $\mathcal{C}$ extends σ.

These restrictions where tried to overcome the observation that the original definition of "refutable learning" of Mukouchi and Arikawa [22,23] where rather restrictive. In particular, the original conditions permitted only to learn finite sets since, on the one hand, a learner cannot learn an infinite set A and all its subsets and, on the other hand, a refuting learner cannot refute any subset of a set it learns.

The present work takes an alternative approach to improve the power of a refuting learner. First, we consider a model where the learner is only required to "refute in the limit". In a context of language learning from informant, a similar concept has been introduced recently and independently by Jain, Kinber, Wiehagen and Zeugmann [11]. In a context of function-learning, already Grieser [8,9] investigated learners that refute in the limit. In his model of reflecting learning, however, a function f has only to be refuted in case it is incompatible with the class to be learned, i.e., if there is a prefix $\sigma \preceq f$ that is not extended by any function in the class to be learned. Grieser [8,9] notes that with his model in many (but not all) cases the necessity to refute can be avoided by transition to a dense superclass $\mathcal{C}'$ of the class $\mathcal{C}$ to be learned because by definition, $\mathcal{C}$ is learnable with reflection in the limit iff $\mathcal{C}'$ is learnable in the limit with respect to the standard definition of learning.

Second, more powerful variants of texts will be considered in order to overcome the restriction to classes of finite sets. This is achieved by considering a slightly altered form of the logical-based setting originally considered by Mukouchi and Arikawa [22,23]. Informally, our approach can be summarized as follows.

- The learner either has to converge to an index of the set to be learned or to the distinguished refutation symbol "?". It will be shown in Remark 3.5 that this type of learning is more restrictive than reliable learning but is less restrictive than the model used by Mukouchi and Arikawa, where the data is already refuted by outputting a single refutation symbol.
- The data-sequences are sequences of first-order sentences describing the set to be learned. In the special case where the data just contains the atomic facts that hold for the set to be learned, this is equivalent to presenting a

standard text for the set. We will, however, also consider models where the data does not just contain atomic sentences but Π_k-sentences of some given level k in the quantifier-alternation-hierarchy.

For more detailed accounts of inductive inference, inference in logic and of recursion theory in general we refer the reader to the monographs [3,12,19,24,30].

Notation. For an arbitrary set A, let A^* be the set of finite strings over A. We write $\mathbb{N}$ for the set of natural numbers. Unless explicitly stated otherwise, by the terms *set* and *class* we refer to a set of natural numbers and to a set of such sets, respectively. We fix a canonical indexing of the finite sets and we let F_i denote the finite set with canonical index i. A class $\mathcal{C}$ of finite sets is *computable* (is *recursively enumerable*) iff the set $\{i : \mathrm{F}_i \in \mathcal{C}\}$ is computable (is recursively enumerable). Observe that a non-empty class of finite sets is recursively enumerable iff it can be represented as $\{\mathrm{F}_{g(i)} : i \in \mathbb{N}\}$ for some recursive function g.

2 Learning from Standard Texts

Before we discuss refuting learning in Section 3, we shortly review some basic concepts and techniques from learning theory. Related to the task of refuting input texts, the learners considered in the following do not just output indices (i.e, natural numbers) but might also output a special refutation symbol.

We fix two distinguished symbols not in $\mathbb{N}$, the pause symbol $\#$ and the refutation symbol ?. *Texts* and *strings* are infinite and finite, respectively, sequences over $\mathbb{N} \cup \{\#\}$. The *range* of a text or string is the set of all elements appearing in it that are different from the pause symbol. We write $\mathrm{range}(\sigma)$ for the range of a string σ. A text is a *text for* a set A iff A coincides with the range of this text and hence, for example, $\#\#\ldots$ is the only text for the empty set.

Definition 2.1. An *unrestricted learner* is a mapping from strings to $\mathbb{N} \cup \{?\}$ and a *learner* is such a mapping that is computable. A learner EX-learns or, for short, *learns* a set A iff on every text for A, the learner converges to an index for A. A learner learns a class iff it learns every set in the class.

Remark 2.2. The numbers output by a learner are meant as hypothesis on the set to be learned with respect to some fixed indexing. In this connection, the usage of *computable* and of *recursively enumerable* indices is most common, i.e., the number i denotes the ith partial recursive function and the ith computational enumerable set W_i, respectively. In the sequel we will frequently consider the learning of classes of finite sets, and in this situation we might also use canonical indices as hypotheses.

Most of the results shown below will go through if we simply require that the sets to be learned can be identified by an index at all and that there is an effective mapping that maps some natural description of sets emerging from the learning algorithm to indices of these sets with respect to the indexing used. In fact, we will not presuppose more on the indexing used unless explicitly stated otherwise.

In a context of learners that always output a natural number, Osherson, Stob and Weinstein [25] considered learners that converge on all texts.

Definition 2.3. [25, Section 4.6.2] A learner is *confident* iff on any text, the learner converges to a natural number. A class is *confidently learnable* iff it is learned by a confident learner.

Standard techniques and results for confident learners extend easily to the type of learners considered here, which besides natural numbers might also output refutation symbol, if we require again that the learner has to converge – to an index or to the refutation symbol – on all texts.

Remark 2.4. Any class that is learned by a learner that converges on all texts does not contain infinite ascending chains. In particular, the class of all finite sets cannot be learned by such a learner.

For a proof by contradiction, fix a learner M that converges on all texts and consider an ascending chain $A_0, A_1, \ldots$ of sets that are all learned by M. Then one can inductively find strings σ_k such that each σ_k contains only elements from A_k and M outputs an index for A_k on input $\sigma_0\sigma_1\ldots\sigma_k$. So M outputs on the text $\sigma_0\sigma_1\ldots$ an index for each of the sets A_k and hence does not converge, neither to an index for a set nor to the refutation symbol. Note that this proof does not require that $\sigma_0\sigma_1\ldots$ is indeed a text for the set $A_0 \cup A_1 \cup \ldots$, it may also be a text for a subset.

Remark 2.5. Let M_0 be a learner that converges on all texts. Then for any set A there is a string τ that is a stabilizing sequence for M_0 and A in the sense that for all strings γ over $A \cup \{\#\}$ we have $M_0(\tau) = M_0(\tau\gamma)$. For a proof observe that, otherwise, we could construct a text for A on which M_0 diverges.

By searching for such stabilizing sequences we can construct a learner M that learns all sets that are learned by M_0 and has in addition the following properties (for details of this construction see Jain et al. [12, Proposition 5.29 on Page 102]). First, for any set A — including sets that are not learned or are not even indexed by the given indexing — the learner M converges on every text of A to the same value $M_0(\tau)$, where τ is the least stabilizing sequence for M_0 and A (with respect to some appropriate ordering on strings). Second, any text for any set A has a finite prefix that is a stabilizing sequence for M and A.

3 Refuting Learning from Standard Texts

Next we review the definitions of the concepts of refuting and reliable learning that are due to Mukouchi and Arikawa [22] and to Minicozzi [21], respectively, and we introduce the related concept of limit-refuting learning. While a refuting learner continues forever to output refutation symbols after having output a refutation symbol once, a limit-refuting learner might alternate between indices and refutation symbols in an arbitrary way before converging.

Definition 3.1. A learner *refutes* a set iff on every text for this set, the learner first outputs at most finitely many numbers (without outputting any refutation symbol) and then outputs nothing but refutation symbols. A learner is *refuting* iff for any set A, either the learner refutes A or on every text for A the learner converges to an index for A without ever outputting ?. A learner *limit-refutes* a set iff on every text for this set, the learner converges to ?, and a learner is *limit-refuting* if it either learns or limit-refutes any set.

A class $\mathcal{C}$ is *refuting learnable* iff there is a refuting learner that learns $\mathcal{C}$. A class $\mathcal{C}$ is *sharply refuting learnable* iff there is a refuting learner that learns $\mathcal{C}$ and refutes every set not in $\mathcal{C}$. The concept *limit-refuting learnable* and its sharp variant are defined likewise with refuting replaced by limit-refuting.

A learner is *reliable* iff for any set A, the learner either learns A or has infinitely many mind changes on any text for A, and a class is *reliably learnable* if it is learned by a reliable learner.

The sharply refuting learnable classes are those originally introduced by Mukouchi and Arikawa [22,23]. Observe that a class is refuting learnable iff it is a subclass of some sharply refuting learnable class. This follows because by definition any refuting class is a subclass of a sharply refuting class while, on the other hand, any sharply refuting learnable class is refuting learnable and any subclass of a refuting learnable class is again refuting learnable.

In Remarks 3.2 through 3.4, we describe some features of the types of learning described in Definition 3.1 and then, in Remark 3.5, we compare their respective strength.

Remark 3.2. For the scope of this remark, call a class *infinitely-often-refuting* learnable iff the class is learned by a learner that for any set A, either learns A or outputs infinitely many refutation symbols on any text for A. By definition, any limit-refuting learnable class is also infinitely-often-refuting learnable. Moreover, the concepts of infinitely-often-refuting learning and of reliable learning coincide. In fact, any infinitely-often-refuting learner can be transformed into a reliable learner that learns the same class by simply replacing each refutation symbol by an index that has not been output on the same text before. Conversely, every reliable learner can be transformed into an infinitely often refuting learner for the same class by outputting a refutation symbol each time before switching to a new index.

Remark 3.3. A class of sets is reliably learnable if and only if it consists only of finite sets. In particular, by Remark 3.5 below, any refuting or limit-refuting learnable class consists only of finite sets.

The restriction of reliable learning to classes of finite sets has already been observed by Osherson, Stob and Weinstein [25, Proposition 4.6.1A] and can be shown as follows. A learner that always outputs an index for the finite set seen so far learns all finite sets and is indeed reliable. Next assume that M is a reliable learner and let $a_1 a_2 \ldots$ be any text for an infinite set A. Then each set of the form $\{a_1, \ldots, a_m\}$ is either learned by M or M diverges on any text for this set and consequently there must be infinitely many sets of the former or infinitely many sets of the latter type. But in both cases, by simply repeating elements in the given text, we can construct a text for A on which M does not converge to an index for A, that is, M does not learn A.

Remark 3.4. Limit-refuting learnable classes do not contain infinite ascending chains. The assertion is immediate from Remark 2.4, because by definition a limit-refuting learner converges on all texts.

The following remark extends the observation of Mukouchi and Arikawa [22] that any refuting learnable class is also reliably learnable.

Remark 3.5. For any class $\mathcal{C}$,

$$\begin{aligned} \mathcal{C} \text{ refuting learnable} \quad &\Rightarrow \quad \mathcal{C} \text{ limit-refuting learnable} \\ &\Rightarrow \quad \mathcal{C} \text{ reliably learnable,} \end{aligned} \tag{1}$$

and both implications are strict. In particular, the concepts of refuting learnable, limit-refuting learnable and reliably learnable class are mutually distinct.

The first implication in (1) is immediate by definition, while the second one follows by Remark 3.2. Moreover, the first two concepts are separated by the class considered in Remark 3.9 below, while the class of all finite sets is reliably learnable but is not limit-refuting learnable, as follows by Remarks 3.3 and 3.4, respectively.

Remark 3.6. Reliable learners have been introduced by Minicozzi [21] by a slightly different formulation where on any text for a set the learner either learns the set or diverges. Minicozzi's definition is apparently less restrictive than Definition 3.1, because the former allows that a learner fails to learn a set A while it still converges to an index for A on some texts for A. Nevertheless, both definitions yield the same concept of reliably learnable class. A similar statement holds with respect to corresponding less restrictive definitions of refuting and limit-refuting learning where for example in the case of limit-refuting learning one just requires that on any text for a set the learner either converges to an index for this set or converges to the refutation symbol.

For a proof, we will show for refuting, limit-refuting, and reliable learning that given a learner M_0 that satisfies the less restrictive definition, there is a learner M that learns all sets learned by M_0 and satisfies the corresponding condition from Definition 3.1. In the case of limit-refuting learning we can simply pick a learner M according to Remark 2.5. Then by construction, M learns all sets learned by M_0 and any set is either learned or limit-refuted by M. In the case of reliable learning, by a construction similar to the one described in Remark 2.5, we obtain a learner M such that on any text for any language A, the learner M converges to an index for A in case there is a stabilizing sequence for M_0 and A, while M diverges, otherwise. In particular, M is reliable in the sense of Definition 3.1 and learns all sets that are learned by M_0. Finally, in the case of refuting learning, define a learner M by

$$M(\sigma) = \begin{cases} ? & \text{if } M_0(\tau) = ? \text{ for some } \tau \text{ with} \\ & \text{range}(\tau) \subseteq \text{range}(\sigma) \text{ and } |\tau| \leq |\sigma|, \\ M_0(\sigma) & \text{otherwise.} \end{cases}$$

Then M learns all sets learned by M and refutes all other sets, i.e., M refutes exactly the sets A such that M_0 outputs a refutation symbol on some text for A.

Remark 3.7. Minicozzi [21] showed that finite unions of reliably learnable classes are also reliably learnable. Similar assertions hold for all variants of refuting learning by the following simple principle. Given k refuting learners $M_1, M_2, \ldots, M_k$ for the classes $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$, the learner M given by

$$M(\sigma) = \begin{cases} M_l(\sigma) & \text{if } l \text{ is the least index with } M_l(\sigma) \neq ?; \\ ? & \text{otherwise;} \end{cases}$$

learns the union $\mathcal{C}$ of the classes $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ with respect to the same variant of refuting learning.

By Remarks 3.3 and 3.4, any refuting learnable class $\mathcal{C}$ contains only finite sets and any infinite set A contains some finite set $D \notin \mathcal{C}$. Mukouchi and Arikawa [23] demonstrated that in the case of unrestricted learners, the two latter properties can be extended to a characterization of the sharply refuting learnable classes, i.e., in our terms a class $\mathcal{C}$ is sharply refuting learnable by an unrestricted learner iff there are finite sets $D_0, D_1, \ldots$ such that any infinite set contains some set D_i while none of the sets D_i is contained in any set in $\mathcal{C}$. The following theorem is essentially a reformulation of the characterization of Mukouchi and Arikawa. Recall in connection with the theorem that by definition a class is refuting learnable if and only if it is contained in an sharply refuting learnable class.

Theorem 3.8. [23] *A class $\mathcal{C}$ is sharply refuting learnable iff $\mathcal{C}$ contains only finite sets and there is a recursively enumerable class $\{D_0, D_1, \ldots\}$ of finite sets such that $\mathcal{C}$ coincides with the class $\{X : D_i \not\subseteq X$ for all $i\}$.*

Theorem 3.8 implies in particular that sharply refuting learnable classes are closed under taking subsets.

Proof. First assume that $\mathcal{C}$ is sharply refuting learnable, that is, $\mathcal{C}$ contains exactly the sets that are learned by some refuting learner M. Then $\mathcal{C}$ contains only finite sets by Remark 3.2. Moreover, the class $\mathcal{D}$ of all finite sets D such that there is a string σ over D where $M(\sigma) = ?$ is obviously recursively enumerable and exactly the sets that are not learned by M contain some set in D. Conversely, given a recursively enumerable class as in the theorem and a representing function g, a refuting learner for $\mathcal{C}$ is obtained as follows. On input σ, the learner checks whether $\mathrm{F}_{g(i)} \subseteq \mathrm{range}(\sigma)$ for any $i \leq |\sigma|$. If so, the learner outputs a refutation symbol while, otherwise, it outputs an index for $\mathrm{range}(\sigma)$. ▮

Remark 3.9. Theorem 3.8 does not extend to limit-refuting learning. A counter-example is given by the class $\mathcal{C}$ of all sets of the form $\{0, 2, 4, \ldots, 2n, 2n+1\}$. The class $\mathcal{C}$ is learned by the limit-refuting learner that outputs an index for the finite set seen so far in case this set is in $\mathcal{C}$ and, otherwise, outputs a refutation symbol. However, the set of all even numbers is infinite and any of its finite subsets with maximum m can be extended to the set $\{0, 2, 4, \ldots, m, m+1\}$ in $\mathcal{C}$, i.e., there is no sequence $D_0, D_1, \ldots$ as in Theorem 3.8.

The following theorem shows that also the property of limit-refuting learnable classes shown in Remark 3.4 can be extended to a characterization.

Theorem 3.10. *A class is limit-refuting learnable iff it is contained in a recursively enumerable class of finite sets that does not contain any infinite ascending chain.*

Proof. Let $\mathcal{C}$ be a class of finite sets that is recursively enumerable with representing function g and does not contain any infinite ascending chain. Consider the learner M that on input σ outputs an index for $\mathrm{range}(\sigma)$ in case the latter set is among $\mathrm{F}_{g(0)}, \ldots, \mathrm{F}_{g(|\sigma|)}$ and, otherwise, outputs ?. Then, obviously, M is computable and learns every set in $\mathcal{C}$. Moreover, M limit-refutes any set $A \notin \mathcal{C}$. In case A is finite, this is immediate by the construction of M. So assume that A is infinite and let $a_1, a_2, \ldots$ be any text for A. As $\mathcal{C}$ contains no infinite chains, $\mathcal{C}$ contains only finitely many sets of the form $\{a_1, a_2, \ldots, a_n\}$ and hence M outputs a refutation symbol on almost all prefixes of the given text.

In order to prove the reverse direction, assume that we are given a class $\mathcal{C}_0$ that is learned by a limit-refuting learner M. Let the class $\mathcal{C}$ contain all finite sets C such that

$$\begin{aligned} &\text{for all } \tau \in (C \cup \{\#\})^* \text{ with } |\tau| \leq |C|, \\ &\qquad \text{there is } \gamma \in (C \cup \{\#\})^* \text{ with } M(\tau\gamma) \neq ?. \end{aligned} \tag{2}$$

The set of all indices i such that (2) is satisfied with C replaced by F_i is recursively enumerable, that is, $\mathcal{C}$ is a recursively enumerable class of finite sets. Moreover, by construction, $\mathcal{C}$ contains all finite sets that are learned by M, hence $\mathcal{C}_0$ is contained in $\mathcal{C}$. Assume now for a proof by contradiction that $A_0, A_1, A_2, \ldots$ is an infinite ascending chain that is contained in $\mathcal{C}$ and let A be the union of the sets A_i. Then by Remark 3.3, the learner M limit-refutes the infinite set A and hence, by Remark 2.5, there is a stabilizing sequence τ for A and M with $\mathrm{range}(\tau) \subseteq A$ and $M(\tau\gamma) = ?$ for any string γ with $\mathrm{range}(\tau) \subseteq A$.

Thus (2) is false for all finite subsets C of A where $\mathrm{range}(\tau) \subseteq C$ and $|\tau| \leq |C|$. Consequently, contrary to our assumption, almost all sets A_i are not in $\mathcal{C}$. ∎

Recall that a confident learner is a learner that always converges to a natural number, see Definition 2.3. Given any limit-refuting learner, according to Remark 3.2 this learner can be transformed into an equivalent reliable learner. Similarly, by replacing all refutation symbols by any fixed index, we can transform any limit-refuting learner into a confident learner that learns the same class, that is, any limit-refuting learnable class is also confidently learnable. Theorem 3.11 shows that the reverse implication is true for classes that are closed under taking subsets.

Theorem 3.11. *Let the class $\mathcal{C}$ be closed under taking subsets. Then the following conditions are equivalent.*
(a) *$\mathcal{C}$ is limit-refuting learnable.*
(b) *$\mathcal{C}$ is confidently learnable.*
(c) *$\mathcal{C}$ is learned by a refuting learner that may use the halting problem* K *as an oracle.*

Proof. By a standard argument, a class that contains an infinite sets and all its subsets cannot be learned at all. As a consequence, all three conditions imply that $\mathcal{C}$ contains only finite sets. Recall from the discussion preceding Theorem 3.11 that (a) implies (b).

(a) $\Rightarrow$ (c): Let M be a limit-refuting learner for $\mathcal{C}$ and by Remark 2.5 assume that any text of any set A has a finite prefix σ that is a stabilizing sequence in the sense that $M(\sigma\eta) = M(\sigma)$ for all η with $\mathrm{range}(\eta) \subseteq A$. By Remark 3.5, all sets learned by M are finite.

Consider the learner N that on input σ outputs an index for the range of σ if for all $\tau \preceq \sigma$ there is n with $M(\tau\#^n) \neq\ ?$ and, otherwise, outputs a refutation symbol. Obviously, the learner N is K-recursive and on any text, keeps the refutation symbol after having output it once. Hence it suffices to show that N is limit-refuting and learns $\mathcal{C}$. So fix any set A.

First assume that A is in $\mathcal{C}$. Then M learns A and all its subsets. Thus N outputs an index for $\mathrm{range}(\sigma)$ for every $\sigma \in (A \cup \{\#\})^*$. But A must be finite, and consequently N learns A.

Next assume that A is infinite. Then M does not learn A but converges on every text for A to the refutation symbol. By assumption on M, any given text for A has a prefix σ that is a stabilizing sequence, i.e., M outputs the refutation symbol on all extensions of σ. As a consequence, N converges to the refutation symbol on any text for A.

It remains the case that A is a finite set not belonging to $\mathcal{C}$. Then by construction, N either converges on any text for A to an index of A or N converges on any text for A to the refutation symbol.

(c) $\Rightarrow$ (a): Let N be a K-recursive refuting learner for $\mathcal{C}$ and for every s, let K_s denote the s-step approximation of K, i.e., a number is in K_s iff it is enumerated into K within s steps of some appropriate enumeration of K. Then

$\mathcal{C}$ is learned by the limit-refuting learner M that on input σ checks whether there is a string τ where

$$|\tau| \leq \max \operatorname{range}(\sigma) \quad \text{and} \quad \operatorname{range}(\tau) \subseteq \operatorname{range}(\sigma) \quad \text{and} \quad N^{\mathrm{K}_{|\sigma|}}(\tau) = ?, \quad (3)$$

and if so, outputs a refutation symbol and, otherwise, outputs an index for range(σ). The key observation in the verification of the construction is that for any fixed string τ and for almost all s the output of N on input τ is the same for oracle K and for oracle K_s. First consider a finite set D. In case there is a string τ such that (3) holds with range(σ) and $\mathrm{K}_{|\sigma|}$ replaced by D and K, respectively, then on any text for D, the learner M will converge to the refutation symbol while, otherwise, the learner converges on any text for D to an index for D. In particular, the learner learns every finite set learned by N, and hence, by the introductory remark, learns $\mathcal{C}$. Moreover, given any text T for an infinite set A, there is a string τ over A with $N(\tau) = ?$, and thus on all prefixes σ of T that are sufficiently long, M will output a refutation symbol.

(b) $\Rightarrow$ (a): Fix a confident learner M for $\mathcal{C}$. Let $\mathcal{C}_1$ be the class that contains the empty set and all non-empty finite sets C where

$$\text{for all } \tau \in C^* \text{ with } |\tau| < |C|, \text{ there is } \gamma \in C^* \text{ with } M(\tau) \neq M(\tau\gamma). \quad (4)$$

Then $\mathcal{C}_1$ is recursively enumerable and thus by Theorem 3.10 it suffices to show that $\mathcal{C}_1$ contains $\mathcal{C}$ and does not contain any infinite ascending chain.

In order to show $\mathcal{C} \subseteq \mathcal{C}_1$ fix any non-empty set $C \in \mathcal{C}$ and recall that by the introductory remark, C is finite. Thus in order to show that C is in $\mathcal{C}_1$, it suffices to show that C satisfies (4). Let τ be any string over C with $|\tau| < |C|$. By assumption, M learns the set C and its subset range(τ), and hence there are strings γ_1 and γ_2 over C such that $M(\tau\gamma_1)$ and $M(\tau\gamma_2)$ are indices for the two distinct sets C and range(τ), respectively. As a consequence, $M(\tau) \neq M(\tau\gamma_1)$ or $M(\tau) \neq M(\tau\gamma_2)$, i.e., there is a string γ as required in (4).

Next, we show that given an infinite ascending chain $C_0, C_1, \ldots$ of sets in $\mathcal{C}_1$, contrary to our assumption on M, we could construct a text on which M diverges. The text is constructed by defining inductively prefixes of the text of increasing length. During the construction we will ensure that each such prefix is a string over one of the sets C_i. We start with the empty string and given an already constructed prefix τ in C_i^*, we choose $j > i$ with $|\tau| < |C_j|$. By assumption, C_j satisfies (4) and so we can pick an extension $\tau\gamma \in C_j^*$ with $M(\tau) \neq M(\tau\gamma)$. ▮

Theorem 3.12. *In a setting of learners that use canonical indices as hypotheses, any class is limit-refuting learnable if and only if it is confidently learnable.*

Proof. We have already seen for all indexings considered here that any limit-refuting learnable class is confidently learnable. In order to show the reverse implication, let $\mathcal{C}$ be the class of sets that are learned by a confident learner M_0

via canonical indices. By Remark 2.5, we can assume that for any given set, M_0 converges on all texts for this set to the same index. We show that the learner M defined by

$$M(\sigma) = \begin{cases} M_0(\sigma) & \text{in case } \mathrm{F}_{M_0(\sigma)} = \mathrm{range}(\sigma), \\ ? & \text{otherwise,} \end{cases}$$

is limit-refuting and learns exactly the same sets as M_0. Fix an arbitrary set A and assume that on any text for A, the confident learner M_0 converges to the canonical index i. In case M_0 learns A, that is, if i is an index for A, it is easy to see that M converges to i on any text for A, too. On the other hand, in case M_0 does not learn A, then for any text for A, almost all prefixes have range different from A and consequently M converges to ?. ∎

4 Refuting Learning in a Logical Setting

In the sequel, we consider learning in a logical setting, that is, the classes to be learned, the data, and occasionally also the hypotheses are given in terms of logical formulae. We will always work with a logical language $\mathcal{L}$ that consist of a unary predicate symbol P plus a subset of the symbols $\{\mathrm{s}, <, +, *, \overline{0}, \overline{1}, \ldots\}$. The structures considered all have domain $\mathbb{N}$ and the interpretation of the symbols other than P is always the usual one, that is, $\overline{n}$ is interpreted as number n, s is the successor function, $<$ is the usual strict order on natural numbers, and $+$ and $*$ are interpreted as addition and multiplication over $\mathbb{N}$. We will refer to such structures as *standard structures*. The aim of the learning process is then to identify the interpretation of P. The logical language $\mathcal{L}$ will be chosen among

$\mathcal{B} = \{\mathrm{P}, \overline{0}, \overline{1}, \overline{2}, \ldots\}$, the *basic language*,
$\mathcal{S} = \mathcal{B} \cup \{\mathrm{s}\}$, the *language of successor*,
$\mathcal{O} = \mathcal{B} \cup \{\mathrm{s}, <\}$, the *language of order*,
$\mathcal{P} = \mathcal{B} \cup \{\mathrm{s}, <, +\}$, the *language of Presburger's arithmetic*,
$\mathcal{A} = \mathcal{B} \cup \{\mathrm{s}, <, +, *\}$, the *language of arithmetic.*

With a language $\mathcal{L}$ understood, the standard structure determined by a set A is denoted by $\mathbf{M}(A)$. In the setting considered here, a set A and the structure $\mathbf{M}(A)$ are essentially equivalent. Accordingly, we extend the notation introduced in connection with the learning of sets to the learning of $\mathcal{L}$-structures. For example, given any $\mathcal{L}$-structure $\mathbf{M}(A)$, an $\mathcal{L}$-text for $\mathbf{M}(A)$ or, for short, a *text for* $\mathbf{M}(A)$ is simply a text for the set A and a learner learns $\mathbf{M}(A)$ if on every text for $\mathbf{M}(A)$, the learner converges to an index for A. Moreover, given a sentence Ψ and a set A, we write $\Psi[A]$ for the truth value of Ψ in $\mathbf{M}(A)$.

A class $\mathcal{C}$ of standard structures is called $\mathcal{L}$-*axiomatizable* if and only if there is a $\mathcal{L}$-sentence Ψ such that $\mathcal{C}$ contains exactly those standard structures in which the formula is true, i.e., $\mathcal{C} = \{X : \Psi[X]\}$. Before investigating into the learning of axiomatizable classes in Section 5, we review some well-known facts about such classes.

Remark 4.1. Any $\mathcal{P}$-axiomatizable class of finite sets is computable. For a proof, recall that Presburger's arithmetic, i.e., the theory of the natural numbers with addition, is decidable [26,29]. As a consequence, given an index i and a $\mathcal{P}$-formula Φ, we can effectively test whether Φ is true in the standard structure determined by the finite set $\mathrm{F}_i = \{n_1, \ldots, n_m\}$ by first replacing in Φ every subformula of the form $\mathrm{P}t$, t a term, by $t = \overline{n_1} \vee \ldots \vee t = \overline{n_m}$, then checking whether the resulting formula is true in Presburger's arithmetic.

Remark 4.2. For any $\mathcal{O}$-axiomatizable class $\mathcal{C}$, the set

$$I_{\mathcal{C}} = \{i : (\exists C \in \mathcal{C})\,[\mathrm{F}_i \subseteq C]\}$$

of canonical indices of all subsets of the sets in $\mathcal{C}$ is computable. A proof can be derived by an argument similar to the one used in Remark 4.1, using Büchi's result [2,31] that the monadic second order theory of the natural numbers with successor and order is decidable.

Remark 4.3. The Theorem of Matiyasevich (see for example Smoryński [29]) states that every recursively enumerable set is Diophantine and thus can be defined by a positive existential $\mathcal{A}$-sentence. More precisely, from an index for a recursively enumerable set W we can compute effectively a constant l and polynomials f and g in $l+1$ variables and with coefficients in $\mathbb{N}$ such that for all $x \in \mathbb{N}$,

$$x \in W \quad \text{iff} \quad (\exists z_1, \ldots, z_l \in \mathbb{N})\,[f(x, z_1, \ldots, z_l) \;=\; g(x, z_1, \ldots, z_l)]\,. \tag{5}$$

Hence the matrix of the right-hand side of (5) is an $\mathcal{A}$-formula. Furthermore, if W is computable, then its complement is recursively enumerable and so for suitable polynomials f', g' as above we have for all $x \in \mathbb{N}$,

$$x \in W \quad \text{iff} \quad (\forall z_1, \ldots, z_l \in \mathbb{N})\,[f'(x, z_1, \ldots, z_l) \;\neq\; g'(x, z_1, \ldots, z_l)]. \tag{6}$$

So we obtain a positive universal $\mathcal{A}$-sentence that defines W because the subformula $f'(\ldots) \neq g'(\ldots)$ is equivalent to $f'(\ldots) < g'(\ldots) \vee g'(\ldots) < f'(\ldots)$.

Now consider any Π_k-set A, i.e., in case k is even there is a recursive set W such that for all x,

$$x \in A \;\Leftrightarrow\; (\forall y_1 \ldots \exists y_h)\,[\langle x, y_1, \ldots, y_h\rangle \in W]$$

and in case k is odd, there is a recursive set W such that for all x,

$$x \in A \;\Leftrightarrow\; (\forall y_1 \ldots \forall y_h)\,[\langle x, y_1, \ldots, y_h\rangle \in W]\,,$$

where in both cases there are exactly $k-1$ alternations of quantifiers. Recall that the function $\langle \ldots \rangle$ is obtained by iterating the standard pairing function

$(x_1, x_2) \mapsto ((x_1 + x_2)^2 + 3x_1 + x_2)/2$, hence $2^{h+1}\langle x, y_1, \ldots, y_h\rangle$ can be written as a polynomial in x and the y_i that has coefficients in $\mathbb{N}$. Again using the representation of recursively enumerable sets by Diophantine equations, we infer that for k even or odd, the set A can be represented by

$$\begin{aligned}
x \in A \Leftrightarrow\ & (\forall y_1 \forall y_2 \ldots \exists y_h\ \exists z_1 \ldots \exists z_l) \\
& [f(x, y_1, \ldots, y_h, z_1, \ldots, z_l) = g(x, y_1, \ldots, y_h, z_1, \ldots, z_l)], \\
x \in A \Leftrightarrow\ & (\forall y_1 \forall y_2 \ldots \forall y_h\ \forall z_1 \ldots \forall z_l) \\
& [f'(x, y_1, \ldots, y_h, z_1, \ldots, z_l) \neq g'(x, y_1, \ldots, y_h, z_1, \ldots, z_l)],
\end{aligned}$$

respectively, i.e., by positive $\mathcal{A}$-sentences with the same number $k-1$ of alternations of quantifiers. Furthermore, these formulae can be effectively computed from a recursive index of the set W and an appropriate representation of the quantifier prefix for the variables $y_1, \ldots, y_h$.

Remark 4.4. Besides learners that state their hypotheses in the form of canonical, computable or recursively enumerable indices of sets, in the logical setting one can also consider learners that state their hypotheses in the form of logical formulae. We consider two ways of indexing sets by logical formulae, which might be called *coinciding indices* and *subset indices*. A formula used as a coinciding index is true for (and thus identifies) exactly one structure in the class to be learned, while for a formula used as a *subset index*, among all structures in the class to be learned that satisfy the formula there is a unique least structure (with respect to set theoretical inclusion), which is hence identified by the formula. Learning via subset indices has been considered by Martin, Sharma and Stephan [20]. Unless explicitly stated otherwise, the results shown in the sequel hold no matter whether we use canonical, computable, recursively enumerable, coinciding or subset indices, as long as all sets to be learned can be indexed at all by such indices. Accordingly, when stating these results we do not make explicit the indexing used.

The interplay between logic and learning has been considered before in several papers [4–6,13,15,16,18–20,28]. In connection with the learning of standard structures, the type of texts used above are essentially equivalent to a sequence that contains exactly the atomic $\mathcal{L}$-sentences that are true in the structure to be learned, a type of data presentation considered by Shinohara [28].

5 Learning Axiomatizable Classes From Standard Text

We show in this section for various models of learning that classes that are axiomatizable in an appropriate language and are learned by an unrestricted learner are in fact learnable. Before, we extend in Theorem 5.1 the characterizations of refuting and limit-refuting learnability stated in Theorems 3.8 and 3.10 to unrestricted learners. The proofs of the results for unrestricted learners are essentially the same as in the effective case and are omitted.

Theorem 5.1. [23] *A class $\mathcal{C}$ has an unrestricted sharply refuting learner iff $\mathcal{C}$ contains only finite sets and there is a sequence $D_0, D_1, \ldots$ of finite sets such that $\mathcal{C}$ coincides with the class $\{X : D_i \not\subseteq X$ for all $i\}$.*

A class $\mathcal{C}$ has an unrestricted limit-refuting learner iff $\mathcal{C}$ contains only finite sets and $\mathcal{C}$ does not contain any infinite ascending chain of finite sets.

Theorem 5.2. *Any $\mathcal{P}$-axiomatizable class that has an unrestricted limit-refuting learner is limit-refuting learnable.*

Proof. Let $\mathcal{C}$ be any $\mathcal{P}$-axiomatizable class that has an unrestricted limit-refuting learner. Then $\mathcal{C}$ is a computable class of finite sets according to Remark 4.1 and Theorem 5.1. We show that $\mathcal{C}$ is limit-refuting learned by the learner that on input σ outputs an index for the finite set range(σ) in case this set is in $\mathcal{C}$ and, otherwise, outputs a refutation symbol. The learner obviously learns every set in $\mathcal{C}$ and limit-refutes any finite set not in $\mathcal{C}$. Moreover, as $\mathcal{C}$ can be limit-refuting learned by an unrestricted learner and hence does not contain infinite ascending chains of finite sets according to Theorem 5.1, the learner limit-refutes all infinite sets. ∎

Theorem 5.3.
(a) *Any $\mathcal{O}$-axiomatizable class that has an unrestricted refuting learner is refuting learnable.*
(b) *Any $\mathcal{P}$-axiomatizable class that has an unrestricted refuting learner is refuting learnable by a learner that may use the halting problem K as an oracle.*

Proof. Let $\mathcal{C}$ be any class that has an unrestricted refuting learner and let $I_{\mathcal{C}}$ be defined as in Remark 4.2, i.e., $I_{\mathcal{C}}$ is the set of canonical indices of all subsets of the sets in $\mathcal{C}$. Now consider the unrestricted learner M that on input σ outputs an index for the finite set range(σ) in case the canonical index of this set is in $I_{\mathcal{C}}$ and, otherwise, outputs a refutation symbol. For $\mathcal{O}$-axiomatizable $\mathcal{C}$, by Remark 4.2, the set $I_{\mathcal{C}}$ and hence also M is computable. For $\mathcal{P}$-axiomatizable $\mathcal{C}$, by Remark 4.1 the set $I_{\mathcal{C}}$ is recursively enumerable and hence M is computable relative to oracle K.

So it suffices to show that the class $\mathcal{C}$ is indeed refuting learned by M. Observe that M learns all subsets of sets in $\mathcal{C}$ and refutes all other finite sets. Hence it remains to show that M refutes all infinite sets, where actually it suffices to show that M limit-refutes all infinite sets because by construction, whenever M outputs a refutation symbol on a string then so does M on any extension of this string. So fix any infinite set A and, by assumption on $\mathcal{C}$, fix a class $\mathcal{C}_1$ that contains $\mathcal{C}$ and is sharply refuting learned by an unrestricted learner. By Theorem 5.1, there is a finite set $D \subseteq A$ that is not contained in any set in $\mathcal{C}_1$ and hence is not contained in any set in $\mathcal{C}$. Consequently, on any text for A, the learner M outputs a refutation symbol as soon as the data contains D, hence M limit-refutes A. ∎

Theorem 5.5 below shows that the assertion on $\mathcal{P}$-axiomatizable classes in Theorem 5.3 cannot be strengthened to computable learners in place of learners that may use the halting problem. In the proof of Theorem 5.5, we exploit that given a predicate for the square numbers, we can define multiplication in $\mathcal{P}$. This technique, which is due to Putnam [27], is described in Remark 5.4.

Remark 5.4. If we add to $\mathcal{P}$ a unary predicate Q that is meant to be always interpreted by the set $\{0, 1, 4, \ldots\}$ of squares, we can define a formula μ such that $\mu(x, y, z)$ is true (in the standard model of the natural numbers) if and only if z is the product of x and y. For a start, we use the fact that the distance between successive squares m^2 and $(m + 1)^2$ is just $2m + 1$ in order to define the following formula due to Putnam [27, Equation 21]

$$\tau(x, u) \equiv \mathrm{Q}u \wedge \mathrm{Q}(u + x + x + \overline{1}) \wedge (\forall t < x + x)\,[\neg \mathrm{Q}(u + t + \overline{1})],$$

where $\tau(x, u)$ is true if and only if u is the square of x. Using this formula, we can simply let

$$\begin{aligned}&\mu(x, y, z) \equiv \\ &(\exists u, v, w)[\tau(x, u) \wedge \tau(y, v) \wedge \tau(x + y, w) \wedge w = u + z + z + v].\end{aligned} \qquad (7)$$

Theorem 5.5. *There is a $\mathcal{P}$-axiomatizable class that has an unrestricted refuting learner but has no refuting learner.*

Proof. For the scope of this proof, let for all e and m,

$$S(e, m) \;=\; \{e, e + 1, e + 4, \ldots, e + m^2\}.$$

Recall from Remark 4.3 that by the theorem of Matiyasevich we can choose a Diophantine representation of the halting problem K, i.e., there are polynomials f and g with coefficients in $\mathbb{N}$ such that for all e,

$$e \in \mathrm{K} \quad \text{iff} \quad (\exists z_1, \ldots, z_l \in \mathbb{N})\,[f(e, z_1, \ldots, z_l) \;=\; g(e, z_1, \ldots, z_l)]. \qquad (8)$$

The formula on the right hand side of (8) can be written as an $\mathcal{A}$-formula and, by replacing multiplication according to Remark 5.4 and introducing sufficiently many intermediate variables $z_{l+1}, \ldots, z_k$, the formula can be written as a formula $\Psi(e)$ over $\mathcal{P} \cup \{\mathrm{Q}\}$ that has the form

$$(\exists z_1 \in \mathbb{N} \ldots \exists z_k \in \mathbb{N})\,[\Delta_0(e, z_1, \ldots, z_k)]$$

where the matrix $\Delta_0(e, z_1, \ldots, z_k)$ is just the conjunction over terms of the form $y_0 = y_1 + y_2$ and $\mathrm{Q}(y_0)$ with $y_i = e$ or $y_i = z_j$. Furthermore, there is a formula $\Psi(e, m)$ over $\mathcal{P} \cup \{\mathrm{Q}\}$ expressing that $\Psi(e)$ is true via z_1 through z_k such that all the z_j satisfy in addition $z_j \leq m$. For further use, observe that given e and m we can check effectively whether $\Psi(e, m)$ is true by searching through the finitely many tuples $(z_1, \ldots, z_k)$ where $z_j \leq m$ for $j = 1, \ldots, k$. By definition of Ψ, we have

$$e \in K \quad \text{if and only if} \quad (\exists m)\,[\Psi(e, m)]. \qquad (9)$$

For any e in K, let m_e be minimum over all m such that $\Psi(e,m)$ is true. Let

$$A_e = S(e, m_e) = \{e, e+1, e+4, \ldots, e+m_e^2\} \qquad \text{and} \qquad \mathcal{C} = \{A_e : e \in K\}.$$

We show that $\mathcal{C}$ is the class we are locking for. First, we define an unrestricted refuting learner that learns $\mathcal{C}$. In case the data seen so far contains nothing but pause symbols, N outputs an index for the empty set. Otherwise let D be the non-empty set of all numbers seen so far. In this case N checks whether there is an index $e \leq \min D$ with $A_e \in \mathcal{C}$ and $D \subseteq A_e$. If yes, N outputs an index for D, if no, N outputs a refutation symbol. By construction, N learns all sets contained in any set in $\mathcal{C}$ and hence learns $\mathcal{C}$. It remains to show that N refutes all sets that are not contained in a set in $\mathcal{C}$, where actually it suffices to show that N limit-refutes all such sets because by construction, if N outputs a refutation symbol on some input, then it does so on all extensions of this input. So fix any set A that is not contained in any set in $\mathcal{C}$ and for all e in $\mathrm{K} \cap \{0, \ldots, \min A\}$, let v_e be the least number in $A \setminus A_e$. Then on any text for A, the learner N will output nothing but refutation symbols after having seen $\min A$ and all the v_e.

Second, assume that there were a refuting learner M for $\mathcal{C}$. Then by Remark 3.4, the class of sets learned by M does not contain infinite ascending chains and in particular for any index e, some of the sets $S(e,0) \subseteq S(e,1) \subseteq S(e,2), \ldots$ are refuted by M. In fact, we can pick a computable function $e \mapsto r_e$ such that for all e, the learner M refutes the set $S(e, r_e)$ because given e, we can find a suitable value r_e by simulating M in parallel on appropriate texts for the sets $S(e,j)$, $j \in \mathbb{N}$. By the remark made after Theorem 3.8, the class of sets learned by M is closed under taking subsets and so for any e in K, the refuted set $S(e, r_e)$ cannot be contained in the set $S(e, m_e)$, which is learned, i.e., m_e must be less than r_e. Hence for e in K, $\Psi(e, r_e)$ is true, whereas for e not in K, $\Psi(e,m)$ is false for all m. But as the truth value of $\Psi(e, r_e)$ can be decided effectively in e, we obtain as a contradiction that K is computable.

Third, we show that the class $\mathcal{C}$ is $\mathcal{P}$-axiomatizable via a formula $\Delta_1 \wedge \Delta_2$. The formula Δ_1 is simply an axiomatization of the class of all sets $S(e,m)$ with $e, m \in \mathbb{N}$ that is obtained by requiring that the set under consideration is finite, the two least elements have distance 1, and the subsequent distances between successive elements always increase by 2. In order to define Δ_2, let $\Psi'(e,m)$ be the formula obtained by replacing in $\Psi(e,m)$ all subformulae of the form $\mathrm{Q}w$ by the formula $\mathrm{P}(w + \min \mathrm{P})$. Then we have for all e and m,

$$\Psi(e,m) \text{ is true} \qquad \text{if and only if} \qquad \Psi'(e,m) \text{ is true in } \mathbf{M}(S(e,m)). \quad (10)$$

For a proof, observe that by convention, while evaluating the formulae in (10), Q is interpreted by the set of squares and P is interpreted by the set $S(e,m)$. In this situation, both formulae are essentially equivalent up to the fact that, intuitively speaking, Ψ uses the set of all squares whereas Ψ' uses just the subset $\{0, 1, \ldots, m^2\}$ of all squares that are encoded into the set $S(e,m)$. This difference, however, cannot result in different truth values because also with Ψ, just squares of size at most m^2 occur as witnesses.

Then let Δ_2 be a formula that expresses that the two greatest elements of the set under consideration have distance $2m-1$ and that $\Psi'(e, m')$ is true for $m' = m$ but is false for all $m' < m$. Observe that for any set $S(e, m)$ of size at least 2, the distance between the greatest element $e + m^2$ and the next element is indeed $2m - 1$.

It remains to show that $\Delta_1 \wedge \Delta_2$ axiomatizes $\mathcal{C}$. A set that is not of the form $S(e, m)$ cannot be in $\mathcal{C}$ and does not satisfy Δ_1. So consider any set $S(e, m)$. In case e is not in K, the set is not in $\mathcal{C}$, but also does not satisfy Δ_2 because $\Psi'(e, m)$ is false for all m according to (9) and (10). Next assume that e is in K. In case m differ from m_e, the set is not in $\mathcal{C}$ and Δ_2 is false by definition of m_e and (10). In case $m = m_e$, however, the set is in $\mathcal{C}$ and $\Delta_1 \wedge \Delta_2$ is satisfied. ∎

In the proof of Theorem 5.5, we have constructed a $\mathcal{P}$-axiomatizable class $\mathcal{C}$ that has no refuting learner but has an unrestricted refuting learner. The unrestricted learner constructed does not learn the class $\mathcal{C}$ sharply refuting, i.e., it learns a proper superclass of $\mathcal{C}$. The following argument shows that this is necessarily so . A $\mathcal{P}$-axiomatizable class $\mathcal{C}$ that is sharply refuting learned by an unrestricted learner must be closed under taking subsets. Then any set F is contained in a set in $\mathcal{C}$ iff F itself is in $\mathcal{C}$ and consequently, an argument similar to the one used in the proof of Theorem 5.3 shows that the set $I_{\mathcal{C}}$ defined there is computable and that we can obtain a refuting learner for $\mathcal{C}$.

6 Learning $\mathcal{O}$-Axiomatizable Classes From Π_k-Texts

In this section we consider learning models where the information given in the data is not just a listing of all elements of the set A to be learned but in addition contains all formulae of a certain type that are true in $\mathbf{M}(A)$. Similar settings have been considered before, e.g., by Gasarch and Smith [5], Martin and Osherson [18], Martin, Sharma and Stephan [20], and Shinohara [28].

Recall that a Π_0-$\mathcal{L}$-formula is an $\mathcal{L}$-formula without quantifiers while for all $k \geq 1$, a *Π_k-$\mathcal{L}$-formula* is an $\mathcal{L}$-formula that consists of a quantifier prefix followed by a quantifier-free formula where the prefix starts with a universal quantifier and has at most $k-1$ alternations between universal and existential quantifiers (e.g., for a quantifier-free $\mathcal{L}$-formula Φ, the formula $(\forall x_1 \forall x_2 \exists x_3)\,[\Phi(x_1, x_2, x_3)]$ is a Π_2-$\mathcal{L}$-formula). The concept of a Σ_k-$\mathcal{L}$-formula is defined almost literally the same except that the quantifier prefix of such a formula starts with an existential quantifier. Recall further that a Π_k-$\mathcal{L}$-sentence is a Π_k-$\mathcal{L}$-formula that does not contain free variables and that an $\mathcal{L}$-formula is *positive* iff it does not contain logical connectives other than $\vee$ and $\wedge$.

Definition 6.1. A Π_k-$\mathcal{L}$-*text* for a set A is a sequence that, besides pause symbols, contains exactly all the positive Π_k-sentences that are valid in $\mathbf{M}(A)$.

For any set A, a text and a Π_0-$\mathcal{L}$-text for A provide essentially the same information, whereas we will see below that for $k > 0$, in general more classes can be learned from Π_k-$\mathcal{L}$-texts than just from texts.

Remark 6.2. As already observed by Martin, Sharma and Stephan [20], there is no need to define Σ_k-texts because the amount of information provided by a Π_k-$\mathcal{L}$-text and by a Σ_{k+1}-$\mathcal{L}$-text is exactly the same. For a proof, observe that for any Π_k-formula $\Psi(x_1, \ldots, x_m)$, in any standard structure the Σ_{k+1}-formula $(\exists x_1 \ldots x_m)\,[\Psi(x_1 \ldots x_m)]$ is true if and only if for some $n_1, \ldots, n_m$, the Π_k-formula $\Psi(\overline{n_1}, \ldots, \overline{n_m})$ is true.

Remarks 3.3 and 3.5 imply that for any language $\mathcal{L}$, just classes of finite sets can be refuting, limit-refuting, or reliably learned from Π_0-$\mathcal{L}$-text. In contrast to this, Theorems 6.3 and 6.5 show that Π_2-$\mathcal{O}$-texts permit limit-refuting and reliably learning, respectively, of all countable $\mathcal{O}$-axiomatizable classes or, equivalently, of all classes that for some constant n, contain only sets that are ultimately periodic with period n.

Theorem 6.3. *Any countable $\mathcal{O}$-axiomatizable class is limit-refuting learnable from Π_2-$\mathcal{O}$-texts.*

Proof. Fix any countable $\mathcal{O}$-axiomatizable class $\mathcal{C}$. First we use a result of Büchi [2] in order to show that there is a natural number n such that $\mathcal{C}$ is contained in the class $\mathcal{C}_n$ of all sets that are ultimately periodic with period n, i.e., that satisfy the formula

$$\Phi_n \equiv (\exists x)\,(\forall y > x)\,[\mathrm{P}(y + \overline{n}) = \mathrm{P}y].$$

It then suffices to show that the classes $\mathcal{C}_n$ can be learned as required. For further use note that Φ_n can be rewritten as a $\mathcal{S}$-formula by replacing $y + \overline{n}$ by the term $\mathrm{s} \ldots \mathrm{s}y$ that corresponds to an n-fold application of the successor operation.

In order to show that $\mathcal{C}$ is contained is some class $\mathcal{C}_n$, observe that the $\mathcal{O}$-axiomatizable class $\mathcal{C}$ is a fortiori axiomatizable in the monadic second order theory over $\mathcal{S}$ (in which $<$ can be defined). As a consequence, there is a Büchi automaton that accepts exactly the sets in $\mathcal{C}$, i.e., $A \in \mathcal{C}$ iff there is an accepting state of the automaton that is reached infinitely often while scanning A (see Thomas [31] for details). Let $q_1, \ldots, q_m$ be the accepting states of the automaton that can be reached at all and for $i = 1, \ldots, m$, consider the set of all binary strings ω such that when reading ω with initial state q_i, the automaton ends in state q_i, and let W_i contain the strings in this set that are minimal with respect to the prefix relation. Then the class $\mathcal{C}$ contains exactly the sets that can be written in the form $\mu_0\mu_1\mu_2 \ldots$ where for some i, the string μ_0 leads from the initial state to q_i and the strings $\mu_1, \mu_2, \ldots$ are all in W_i. Now assume that for some i, the set W_i contains two distinct strings ω

and ω'. Then $\gamma_0 = \omega\omega'$ and $\gamma_1 = \omega'\omega$ are distinct strings of the same length and hence for some string μ, the mapping $X \mapsto \mu\gamma_{X(0)}\gamma_{X(1)}\ldots$ is a one-to-one mapping from sets to the class $\mathcal{C}$, i.e., contrary to assumption, $\mathcal{C}$ cannot be countable. Consequently all W_i are singletons, say $W_i = \{\omega_i\}$, and any set in $\mathcal{C}$, for some i, can be written as $\mu\omega_i\omega_i\omega_i\ldots$ and is hence ultimately periodic with period $|\omega_i|$. As a consequence, all sets in $\mathcal{C}$ are ultimately periodic with period $n = |\omega_1| \cdot \ldots \cdot |\omega_m|$.

It remains to show that for any n, the class $\mathcal{C}_n$ is limit-refuting learnable from Π_2-$\mathcal{O}$-texts. For a formula $\Gamma(x)$ and a term t, let $\Gamma(t)$ be obtained by replacing in $\Gamma(x)$ all occurrences of x with t. For all numbers n and m, let $\Gamma_{m,n}(x)$ be a formula such that for all numbers k, $\Gamma_{m,n}(\overline{k})$ is true in $\mathbf{M}(A)$ if and only if A contains at least m of the n numbers $k+1, k+2, \ldots, k+n$. We can choose the formulae $\Gamma_{m,n}(x)$ as quantifier-free, positive $\mathcal{S}$-formulae, for example, we let

$$\Gamma_{2,3}(x) \equiv (\mathrm{Ps}x \wedge \mathrm{Pss}x) \vee (\mathrm{Ps}x \wedge \mathrm{Psss}x) \vee (\mathrm{Pss}x \wedge \mathrm{Psss}x).$$

Then a set A is ultimately periodic with period n if and only if for almost all numbers k, the formula $\Gamma_{m,n}(\overline{k})$ is true in $\mathbf{M}(A)$ while the formula $\Gamma_{m+1,n}(\overline{k})$ is false. The learner for $\mathcal{C}_n$ now works as follows.

Algorithm. First check if there is $m \leq n$ and k_0 such that the data seen so far contains the first but not the second formula in

$$(\forall x > \overline{k_0})\,[\Gamma_{m,n}(x)], \qquad (\forall y)\,(\exists x)\,[y < x \wedge \Gamma_{m+1,n}(x)]. \tag{11}$$

In case there is no such m, output a refutation symbol. Otherwise, fix minimal such m and for this m, fix minimal such k_0. Let $k_1 > k_0$ be minimal such that for all $k \geq k_1$, the data does not contain $\Gamma_{m+1,n}(\overline{k})$. Then output an index for the set A that agrees with the data seen so far on the interval $\{0, \ldots, k_1+n-1\}$ and that from k_1 on has period n in the sense that for all $k \geq k_1$, we have $A(k) = A(k+n)$.

Verification. Fix any set A and let m_0 be maximal such that in $\mathbf{M}(A)$ for almost all k, the formula $\Gamma_{m_0,n}(\overline{k})$ is true. Then the first formula in (11) will appear in the data (for appropriate values of k_0) exactly for all $m \leq m_0$, while the second formula appears at least for all $m \leq m_0$.

In case $\Gamma_{m_0+1,n}(\overline{k})$ holds in $\mathbf{M}(A)$ just for finitely many k, the set A is ultimately periodic with period m_0 and is hence in $\mathcal{C}_n$. In this case for $m = m_0$, the second formula in (11) will never appear in the data and hence the learner converges to an index for A.

In case $\Gamma_{m_0+1,n}(\overline{k})$ holds in $\mathbf{M}(A)$ for infinitely many k, we infer by the maximality of m_0 that the formula cannot hold for all k and that hence A does not have period m_0 and is not in $\mathcal{C}_n$. In this case, the data will eventually contain the second formula in (11) with $m = m_0$, and hence the learner correctly refutes the input-text. ∎

Theorem 6.4. *Any countable $\mathcal{O}$-axiomatizable class is reliably learnable from Π_1-$\mathcal{O}$-texts.*

Proof. By the discussion in the proof of Theorem 6.3 it suffices to show that the class of all ultimately periodic sets can be learned as required. Fix an appropriate enumeration $\Psi_0, \Psi_1, \ldots$ of all formulae of the form

$$\bigwedge_{k \in A \cap \{0,\ldots,k_0+n+1\}} \mathrm{P}\overline{k} \quad \wedge \quad (\forall x > k_0)\,[\Gamma_{m,n}(x)]$$

where A is an arbitrary set and where $\Gamma_{m,n}$ with $m \leq n$ is the formula defined in the proof of Theorem 6.3. For all i, let A_i denote the (unique) minimum set such that Ψ_i is true in $\mathbf{M}(A_i)$ and observe that the class of ultimately periodic sets coincides with the class $\{A_0, A_1, \ldots\}$.

Algorithm. If there is an index i such that the data seen so far contains Ψ_i but contains no formula $\mathrm{P}\overline{k}$ with $k \notin A_i$, let i_0 be the least such index and output an index for A_{i_0}. Otherwise, output an index for the finite set of all k such that the data contains $\mathrm{P}\overline{k}$.

Verification. Fix an arbitrary set A. Then for any index i such that Ψ_i is true in $\mathbf{M}(A)$ but A differs from A_i, the set A_i must be strictly contained in A and hence on any text for A, the learner eventually sees a formula $\mathrm{P}\overline{k}$ with $k \notin A_i$ and does not output an index for A_i afterwards.

If A is ultimately periodic, let i_0 be the least index such that A is equal to A_{i_0}. Then by the preceding discussion, the learner will at most finitely often output an index for a set A_i with $i < i_0$ and thus will converge to an index for A. If, on the other hand, A is not ultimately periodic, then in particular A is not finite and hence the learner diverges on any text for A on which it branches to the otherwise case infinitely often. But on all other texts the learner infinitely often outputs an index for one of the sets A_i, while by the introductory remark each such index is output at most finitely often, i.e., in this case the learner diverges, too. ∎

The class of all finite sets is easily seen to be $\mathcal{O}$-axiomatizable and hence it can be limit-refuting learned from Π_2-$\mathcal{O}$-texts according to Theorem 6.3. In fact, this class is sharply refuting learned from Π_2-$\mathcal{O}$-texts by the learner that always outputs an index for the finite set seen so far as long it has not yet seen the formula $(\forall y)\,(\exists x)\,[y < x \wedge \mathrm{P}x]$, in which case it outputs a refutation symbol. On the other hand, the class of finite sets is an example for an $\mathcal{O}$-axiomatizable class that cannot be limit-refuting learned from Π_1-$\mathcal{O}$-texts. This shows in particular that Theorem 6.3 cannot be improved from Π_2-$\mathcal{O}$-texts to Π_1-$\mathcal{O}$-texts.

Theorem 6.5. *There is an $\mathcal{O}$-axiomatizable class, namely the class of all finite sets, that cannot be limit-refuting learned from Π_1-$\mathcal{O}$-texts.*

Proof. For a proof by contradiction, assume that there is a limit-refuting learner M that learns the class of all finite sets from Π_1-$\mathcal{O}$-texts. Let E_0 be equal to $\{j^2 : j \in \mathbb{N}\}$. As there are only countably many indices, the learner M does not learn and hence must refute some infinite subset E of E_0. Thus there is a stabilizing sequence for E and M in the sense that there are finitely many Π_1-$\mathcal{O}$-sentences $\Psi_1, \Psi_2, \ldots, \Psi_k$ that are all true in $\mathbf{M}(E)$ and such that on having seen this sentences the learner M will output nothing but refutation symbols unless some sentence comes up that is not true in $\mathbf{M}(E)$. We now argue that M will not learn any finite subset D of E such that all the Ψ_i are true in $\mathbf{M}(D)$. In fact for such D, we can extend the sequence $\Psi_1, \Psi_2, \ldots, \Psi_k$ to a Π_1-$\mathcal{O}$-text for D. All the formulae appearing in this text are positive and hence hold not only in the set D but also in its superset E, i.e., on this text the learner M converges to the refutation symbol and not to an index for D.

Recall that a ground term is a term that does not contain variables and observe that the non-ground terms over $\mathcal{O}$ have the form of zero or more successor symbols that are followed by a variable. Let D be the finite set of all numbers in E that are denoted by a ground term in some Ψ_i. By the preceding discussion, we are done if we can show that all the Ψ_i are true in $\mathbf{M}(D)$. So fix an index i. Without loss of generality assume that Ψ_i is in prenex-normal form and let Φ be the part after the quantifiers, that is, Φ is a quantifier-free $\mathcal{O}$-formula with Ψ_i being equivalent to $(\forall x_1 \ldots x_n)\,[\Phi(x_1 \ldots x_n)]$. Note that Φ is also a positive formula. Then for any numbers $z_1, \ldots, z_n$, we have to show that $\Phi_i[z_1, \ldots, z_n]$ is true in $\mathbf{M}(D)$. Given such z_i, we can find k such that for any non-ground term t in Ψ_i, $t[z_1 + k, \ldots, z_n + k]$ is not in E_0, i.e, is not a square. By assumption, Ψ_i is true in $\mathbf{M}(E)$ and hence so is $\Phi[z_1 + k, \ldots, z_n + k]$. We now argue that for any atomic subformula Δ of Φ, if $\Delta[z_1{+}k, \ldots, z_n{+}k]$ is true in $\mathbf{M}(E)$, then $\Delta[z_1, \ldots, z_n]$ is true in $\mathbf{M}(D)$. In case Δ is of the form $t_1 = t_2$ or $t_1 < t_2$, this follows easily by the special form of $\mathcal{O}$-terms. Next assume that Δ is of the form $\mathrm{P}t$. Then if t is a ground term, the assertion follows by definition of D whereas, otherwise, the assertion is vacuously satisfied because by choice of k, $t[z_1 + k, \ldots, z_n + k]$ is not in E. As a consequence and because Ψ_i does not contain negation symbols, Ψ_i is true in $\mathbf{M}(D)$. ▮

7 Learning $\mathcal{S}$-Axiomatizable Classes From Π_k-Texts

By a result of Thomas [32,33], membership of a set L in an $\mathcal{S}$-axiomatizable class $\mathcal{L}$ can be checked by counting, up to some threshold value m, for all strings of length less than or equal to some number n how often they appear as substring in the characteristic sequence of L.

Fact 7.1. *Let* $\mathrm{occ}_m(L, \eta)$ *be the minimum of* m *and the number of substrings of* L *that are equal to* η*. Say two languages* L *and* H *are* (m, n)*-equivalent if they satisfy*

$$(\forall k \le n)\,[L(k) = H(k)] \;\wedge$$
$$(\forall \eta \in \{0,1\}^*, |\eta| \le n)\,[\mathrm{occ}_m(L, \eta) = \mathrm{occ}_m(H, \eta)].$$

For every $\mathcal{S}$-axiomatizable class $\mathcal{C}$ there are constants m, n such that whenever two languages L, H are (m, n)-equivalent then $L \in \mathcal{L}$ if and only if $H \in \mathcal{L}$.

By Fact 7.1, any $\mathcal{S}$-axiomatizable class is equal to a union of equivalence classes of (m, n)-equivalence. As a consequence, one can determine membership of a set L in such a class by just analyzing $L(0)L(1)\ldots L(n)$ and the numbers $\mathrm{occ}_m(L, \eta)$ for all strings η up to length n. For example, consider the $\mathcal{S}$-axiomatizable class $0^+1^+0^\infty$ of all sets that have a characteristic function of the form $0^i1^j0^\infty$ for some $i, j > 0$. For $n = 2$ and $m = 2$ we then have

$$L \in 0^+1^+0^\infty \Leftrightarrow L(0) = 0 \wedge (\mathrm{occ}_m(L, 01) = 1 \wedge \mathrm{occ}_m(L, 10) = 1).$$

Theorem 7.2. *Every countable $\mathcal{S}$-axiomatizable class can be limit-refuting learned from Π_1-$\mathcal{S}$-text.*

Proof. Fix any countable $\mathcal{S}$-axiomatizable class $\mathcal{C}$. Like in the proof of Theorem 6.3 one can derive from the result of Büchi [2] that there is some number n_1 such that every set in $\mathcal{C}$ is ultimately periodic with period n_1, and hence with period kn_1 for any natural number $k \geq 1$. Furthermore, by Fact 7.1, there are m and n_0 such that the class $\mathcal{C}$ is equal to the union of finitely many equivalence classes of (m, n_0)-equivalence. Now fix a natural number $n \geq n_0$ that is a multiple of n_1. By definition, (m, n)-equivalence implies (m, n_0)-equivalence, hence $\mathcal{C}$ is the union of finitely many equivalence classes of (m, n)-equivalence. By Remark 3.7 it is sufficient to show that each such equivalence class can be limit-refuting learned from Π_1-$\mathcal{S}$-text. So fix such an equivalence class $\mathcal{H}$. By construction, any set in the subclass $\mathcal{H}$ of $\mathcal{C}$ can be written in the form $\delta\sigma^\infty$ for some string δ and a string σ of length n. Again by Remark 3.7 it suffices to show that for any given σ of length n, the subclass of $\mathcal{H}$ of all sets of the form $\delta\sigma^\infty$ can be learned as required. So fix an arbitrary string σ of length n and let $\{a_1, \ldots, a_i\}$ be the set of positions that σ maps to 1.

For any string δ, let the set $L(\delta)$ be equal to $\delta\sigma^\infty$. For the scope of this proof, call δ *admissible* if $L(\delta)$ is in $\mathcal{H}$ and call δ *consistent* with a given $\mathcal{S}$-sentence Ψ if Ψ is true in $\mathbf{M}(L(\delta))$, i.e., if Ψ is true when the predicate symbol P is interpreted by the set $L(\delta)$. The hypotheses output by the learner constructed below will always be of the form $L(\delta)$ for appropriate strings δ. It is easy to see that the learner can rule out all hypotheses $L(\delta)$ where the string δ is not admissible or is not consistent with the data seen so far. The following claim shows that these two properties of δ can be checked by a computable learner.

Claim 1. Given a string δ and an $\mathcal{S}$-sentence Ψ, it can be effectively checked whether δ is admissible and whether δ is consistent with Ψ.

It can be effectively checked whether δ is admissible because by the special form of the set $L(\delta) = \delta\sigma^\infty$ we can compute the length n prefix of $L(\delta)$ and the values $\mathrm{occ}_m(L(\delta), \eta)$ for all strings η of length at most n.

Next we show that we can effectively decide whether Ψ is true in $\mathbf{M}(L(\delta))$.

Similar to the numbers $a_1, \ldots, a_i$ defined above, let $\{b_1, \ldots, b_j\}$ be the sets of positions that δ maps to 1. Replace in Ψ each subformula of the form Pt, t a term, by the formula

$$t = \overline{a_1} \vee \ldots \vee t = \overline{a_i} \vee (\exists y)\,[t = |\delta| + \overline{n}y + b_1 \vee \ldots \vee t = |\delta| + \overline{n}y + b_j],$$

where $\overline{n}y$ stands for n times adding y. By eliminating the predicate symbol P this way, we obtain a $\mathcal{P}$-sentence that is true (in the standard model of the natural numbers) if and only if the original formula Ψ is true in $\mathbf{M}(L(\delta))$. Now we are done because the set of all true $\mathcal{P}$-sentences that do not contain the predicate symbol P, i.e., the set of true sentences of Presburger's arithmetic, is decidable.

For any string δ, we define a positive Π_1-$\mathcal{S}$-sentence Φ_δ that is true in case P is interpreted by $L(\delta)$ but is false in case P is interpreted by any proper subset of $L(\delta)$. Given δ, as in the proof of Claim 1, let $\{b_1, \ldots, b_j\}$ be the set of positions that δ maps to 1. Let

$$\begin{aligned}\Phi_\delta = {} & P\overline{a_0} \wedge P\overline{a_1} \wedge \ldots \wedge P\overline{a_i} \\ & \wedge \quad (\forall x \geq |\delta| - 1)\,(\exists k < n)\,[P(x + k + b_1) \wedge \ldots \wedge P(x + k + b_j)] \\ & \wedge \quad (\forall x \geq |\delta| - 1)\,[\Gamma_{j,n}(x)],\end{aligned}$$

where the formula $\Gamma_{j,n}(x)$ is defined as in the proof of Theorem 6.3. If $i = 0$ or $j = 0$, the corresponding parts of Φ_δ are replaced by a tautology, say $\mathrm{s}\overline{0} = \overline{1}$, in order to make the structure of the formula unique. Observe in connection with the definition of Φ_δ that any existential quantifier that is bounded by a constant can be written as a disjunction and that for any term c that is a sum of constants, the term $x + c$ stands for $\mathrm{s}^c x$.

Algorithm. (Let σ, m, n, j be the string and the numbers defined above.)

- Let $\delta_1, \delta_2, \ldots, \delta_h$ be the sequence of all strings δ such that the formula Φ_δ has appeared in the input text. The strings are ordered according their "date of appearance", i.e., Φ_{δ_i} has appeared before $\Phi_{\delta_{i+1}}$.
- If there is an index $k \leq h$ such that δ_k is consistent with the current data (i.e., all sentences contained in the data so far are true in $\mathbf{M}(L(\delta_k))$), then for the least such k, select the string $\delta = \delta_k$.
- If we have just selected a string δ and this string is admissible, output an index for $L(\delta)$, else output the refutation symbol.

Verification. By Claim 1 it is immediate that the constructed learner is indeed computable. So it remains to show that the learner learns all sets in $\mathcal{H}$ that are of the form $L(\delta)$ and limit-refutes all other sets. The proof of this assertion is split into several claims.

Claim 2. Let δ be a string and let L be a proper subset of $L(\delta)$. Then Φ_δ is false in $\mathbf{M}(L)$.

Fix a number d in $L(\delta)$ that is not in L. In case d is in the domain of δ, the formula Φ_δ contains a conjunct equivalent to $P\overline{d}$ and hence is false in $\mathbf{M}(L)$. In case d is not in the domain of δ, there is a number h such that d is in the domain of $\delta\sigma^{h+1}$ but not in the domain of $\delta\sigma^h$. If we let $x = |\delta_k\sigma^h|$, the interval $\{x, x+1, \ldots, x+n-1\}$ contains exactly j elements of $L(\delta)$ and one of these elements is d, hence the subset L of $L(\delta)$ contains at most $j-1$ elements in this interval. As a consequence, $\Gamma_{j,n}(x)$ and hence also Φ_δ is false in $\mathbf{M}(L)$.

Claim 3. For any string δ and on any text for a set that differs from $L(\delta)$, the learner selects δ at most finitely often.

For given δ, fix any set L that differs from $L(\delta)$. In case L is a proper subset of $L(\delta)$, then by Claim 2 the formula Φ_δ will not appear in any text for L and hence the learner will never select δ. In case there is an element d of L that is not in $L(\delta)$, as soon as the data contains the formula $\mathrm{P}\overline{d}$, the string δ is no longer consistent with the data seen so far and hence from that time on, the learner will not select δ.

Claim 4. Let L be a set of the form $L(\delta)$. Then the learner learns L if L is in $\mathcal{H}$ and limit-refutes L if L is not in $\mathcal{H}$.

Fix any Π_1-$\mathcal{S}$-text for $L = L(\delta)$. Let $\delta_1, \delta_2, \ldots$ be the sequence of all strings δ such that Φ_δ appears in this text, ordered by appearance of the formulae Φ_{δ_l}. Let k be the least index such that L is equal to L_{δ_k}. Such an index exists because any Π_1-$\mathcal{S}$-text for $L(\delta)$ contains the formula Φ_δ. By Claim 3, any index $l < k$ will be selected at most finitely often. Furthermore, the formula Φ_{δ_k} eventually appears in the input and is always consistent with the current data. So the learner will almost always select the string δ_k. But by definition, δ_k is admissible if and only if $L(\delta_k)$ is in $\mathcal{H}$, hence the learner outputs almost always an index for $L = L(\delta_k)$ in case this set is indeed in $\mathcal{H}$, and, otherwise, outputs almost always a refutation symbol.

Claim 5. Let L be a set that is not of the form $L(\delta)$ and where the characteristic function of L contains the substring σ infinitely often. Then the learner limit-refutes L.

Let $\tau_1\sigma^{e_1}\tau_2\sigma^{e_2}\ldots$ be a representation of L such that, for all k, $e_k \geq 1$ and $\sigma \not\preceq \tau_{k+1}\sigma$. Let

$$\alpha_k = \tau_1\sigma^{e_1}\ldots\tau_k\sigma^{e_k}$$

Let $\mathrm{occ}_m(\alpha, \eta)$ the number (up to m) of occurrences of η in α. Then, since α_k is a prefix of α_{k+1}, we have for all k and η,

$$\mathrm{occ}_m(\alpha_k, \eta) \leq \mathrm{occ}_m(\alpha_{k+1}, \eta) \leq m.$$

As a consequence, there is a number $k > 1$ such that, for all η of length at most n,

$$\mathrm{occ}_m(\alpha_k, \eta) = \mathrm{occ}_m(\alpha_{k+1}, \eta).$$

From this equation one infers $\mathrm{occ}_m(\alpha_k, \eta) = m$ for every substring η of $\sigma\tau_{k+1}\sigma$ that has length at most n. If there is now a string ϑ such that the set $\alpha_k\vartheta\sigma^\infty$

is in $\mathcal{H}$, then also all sets where τ_{k+1} is inserted arbitrarily often between two occurrences of σ are in $\mathcal{H}$, i.e., in particular, $\mathcal{H}$ contains all sets of the form

$$\alpha_k \vartheta \sigma \mu_1 \mu_2 \ldots, \qquad \mu_i \in \{\sigma\sigma\tau_{k+1}, \sigma\tau_{k+1}\sigma\}.$$

By assumption, σ is not a prefix of $\tau_{k+1}\sigma$ and hence $\gamma_0 = \sigma\sigma\tau_{k+1}$ and $\gamma_1 = \sigma\tau_{k+1}\sigma$ are distinct strings of equal length. Like in the proof of Theorem 6.3, this then implies that $X \mapsto \alpha_k\delta\gamma_{X(0)}\gamma_{X(1)}\ldots$ is a one-to-one mapping from sets to $\mathcal{H}$, thus contradicting the assumption that $\mathcal{C}$ is countable. So α_k cannot be extended to a set of the form $L(\delta)$ in $\mathcal{H}$.

Let D_L be the sets of all elements of L that are in the domain of α_k. Fix any text for L and assume that the learning process has reached a stage where for all c in D_L the formula $P\overline{c}$ has already appeared in the data. Consider an arbitrary string δ of length at least $|\alpha_k|$ and let D_δ be the sets of all places in the domain of α_k that δ maps to 1. In case $D_\delta \setminus D_L$ is non-empty, the formula Φ_δ is false in $M(L)$ and will not appear in any text for L, hence δ is never selected. In case $D_L \setminus D_\delta$ is non-empty, say contains a number d, then δ is not consistent with the formula $\mathrm{P}\overline{d}$, which by assumption has already appeared in the input, hence δ will never be selected. Finally, in case the sets D_δ and D_L are the same, α_k is a prefix of δ. Hence by the discussion in the preceding paragraph $L(\delta)$ is not in $\mathcal{H}$ and thus δ is not admissible. In summary, after having seen the formulae $P\overline{c}$ for all elements of L in the domain of α_k, if a string δ of length at least $|\alpha_k|$ is selected at all, this string will not be admissible. Now by assumption, L is not of the form $L(\delta)$, hence by Claim 3 each of the finitely many strings δ of length at most $|\alpha_k|$ will be selected at most finitely often. So the learner outputs almost always a refutation symbol.

Claim 6. If L contains σ only finitely often as a substring, then the learner limit-refutes L.

If for all strings δ, the formula Φ_δ is false in $\mathbf{M}(L)$, then on any text for L the learner outputs nothing but refutation symbols and we are done. So assume that Φ_δ is true in $\mathbf{M}(L)$ for some δ. Then in particular for all x there is a constant $k < n$ such that the j numbers $x + k + b_1$ through $x + k + b_j$ are all contained in L. On the other hand, the set L does not coincide with σ on the interval $\{x+k, \ldots, x+k+n-1\}$ for almost all x and k. Hence for almost all x there is $k < n$ such that this interval contains at least $n+1$ elements of L. As a consequence, the following sentence is true in $\mathbf{M}(L)$

$$(\forall x \geq c)\,(\exists k < n)\,[\Gamma_{j+1,n}(x+k)]$$

and an equivalent sentence shows up in the Π_1-$\mathcal{S}$-text of L. No language of the form $L(\delta)$ is consistent with this formula since every length n substring of σ^∞ contains exactly j times a 1. Hence after having seen this formula, the learner always outputs the refutation symbol.

By Claim 4, the learner works correctly for all sets of the form $L(\delta)$, while Claims 5 and 6 establish that the learner limit-refutes any set that is not of this form. This completes the proof that the given learning algorithm is a limit-refuting learner for the class of all sets of the form $L(\delta)$ in $\mathcal{H}$. The class $\mathcal{C}$ is a finite union of classes of the latter type and hence, by Remark 3.7, also $\mathcal{C}$ has a limit-refuting learner. ∎

Remark 7.3. Similarly it holds that any countable $\mathcal{B}$-axiomatizable class is limit-refuting learnable from Π_1-$\mathcal{B}$-texts. The class $\{A : |\overline{A}| \leq 1\}$ is countable and $\mathcal{B}$-axiomatizable via the formula $(\exists x)\,(\forall y)\,[x = y \vee Py]$. But this class is not learnable from standard text at all. Since, for all languages $\mathcal{L}$ considered here, Π_0-$\mathcal{L}$-texts are equivalent to standard texts, it follows that results like Theorem 7.2 cannot be improved to learnability from Π_0-$\mathcal{L}$-texts.

8 Learning $\mathcal{A}$-Axiomatizable Classes From Π_k-Texts

In this section we show — in contrast to $\mathcal{O}$-axiomatizable classes considered in the last section — that in the case of reliable and limit-refuting learning of $\mathcal{A}$-axiomatizable classes, for increasing k we can learn more and more classes from Π_k-$\mathcal{A}$-text.

Theorem 8.1. *For every k, the class of all $\emptyset^k$-recursive sets is not reliably learnable from Π_k-$\mathcal{A}$-text but is reliably learnable from Π_{k+1}-$\mathcal{A}$-text. In particular, for every k there is a countable $\mathcal{A}$-axiomatizable class that is not reliably learnable from Π_k-$\mathcal{A}$-texts.*

We show the theorems of this section for learners that use subset indices but the results holds by virtually the same proof for learners that use coinciding indices.

Proof. Fix k and let $\mathcal{C}$ be the class of all $\emptyset^k$-recursive sets. Recall that a set A is in $\mathcal{C}$ iff A and its complement are both Σ_{k+1}-sets or, equivalently, iff A and its complement are both Π_{k+1}-sets. Furthermore, given an index e of an oracle Turing machine that computes A relative to oracle $\emptyset^k$, we can compute representations of A and its complement as Σ_{k+1}- and Π_{k+1}-sets and, by Remark 4.3, from these representations we can compute positive Π_{k+1}-$\mathcal{A}$-formulae Θ_0^e and Θ_1^e such that for all n,

$$(n \notin A \text{ iff } \Theta_0^e(\overline{n})[A]) \quad \text{and} \quad (n \in A \text{ iff } \Theta_1^e(\overline{n})[A]) \tag{12}$$

in case A is computed by the e-th oracle Turing machine relative to oracle $\emptyset^k$.

The reliable learner. Now we are in position to define a reliable learner N that learns the class $\mathcal{C}$ from Π_{k+1}-$\mathcal{A}$-text. The learner N essentially works by syntactically analyzing the data, i.e., it checks whether certain formulae containing Θ_0^e and Θ_1^e have already appeared in the input text. Recall that the enumeration $\langle(\Theta_0^e(x), \Theta_1^e(x))\rangle_{e\in\mathbb{N}}$ of the Π_k-$\mathcal{A}$-formulae describing the $\emptyset^k$-recursive sets is computable but in case the eth oracle Turing machine with oracle $\emptyset^k$ is not total, the formulae $\Theta_0^e(x)$ and $\Theta_1^e(x)$ might fail to describe any set uniquely. Note that these formulae do not contain the predicate symbol P.

Algorithm. At any stage of the learning process, say that an index e qualifies if both of the formulae

$$\text{(a)} \quad (\forall x)\,[\Theta_0^e(x) \vee Px], \qquad \text{(b)} \quad (\forall x)\,[\Theta_0^e(x) \vee \Theta_1^e(x)] \tag{13}$$

have appeared in the data seen so far but no formula of the form

$$\text{(c)} \quad \Theta_0^e(\overline{n}) \wedge \mathrm{P}\overline{n}, \qquad \text{(d)} \quad \Theta_0^e(\overline{n}) \wedge \Theta_1^e(\overline{n}). \tag{14}$$

The learner N outputs the subset index given by sentence (a) for the least index e that currently qualifies and, otherwise, if no index e qualifies, N signals divergence by outputting a formula different from the previous one.

Verification. As the formula Θ_i^e can be obtained effectively in i and e, the learner N is indeed computable. For all e and for $i = 0, 1$, let the set T_i^e contain exactly the natural numbers n such that $\Theta_i^e(\overline{n})$ is true in the standard model. We next show that if the learner receives any Π_{k+1}-$\mathcal{A}$-text for a set A then for any index e the following three conditions are equivalent

(I) the index e qualifies almost always,
(II) the index e qualifies infinitely often,
(III) the set A and its complement coincide with T_1^e and T_0^e, respectively.

For a proof, fix any index e. The first condition trivially implies the second one. If the second condition holds, then T_0^e and T_1^e must be a partition of $\mathbb{N}$ because the formula (b) eventually shows up but no formula of the form (d). Moreover, T_1^e is a subset of A by appearance of (a) and T_0^e and A are disjoint by non-appearance of any formula of the form (c). In summary, this then implies the third condition. Finally, in case e satisfies the third condition, it is easy to see that the formulae (a) and (b) eventually show up but no formula of the form (c) or (d), i.e., the index e qualifies almost always.

In order to show that N is reliable, fix any text for a set A. First assume that there is an index that satisfies the three equivalent conditions (I) through (III) and let e be the least such index. Then by construction, the learner N converges to formula (a), which by (III) is indeed a subset index for A. Next assume that no index qualifies infinitely often. Then either there are only finitely many indices that qualify at all and N diverges by construction or there are infinitely many indices that qualify and N diverges by outputting infinitely many of them. Finally, N learns $\mathcal{C}$ because by the discussion preceding (12) for any set in $\mathcal{C}$ there are formulae as in (12) and hence with any text for this set some index e will qualify infinitely often.

Non-Learnability from $\mathbf{\Pi_k}$-$\mathcal{A}$-text. It remains to show that $\mathcal{C}$ is not reliably learnable from Π_k-$\mathcal{A}$-text. In order to do so, we fix any learner M that is reliable (in the sense that for any set Y, either M converges on any

Π_k-$\mathcal{A}$-text for Y to a subset index for Y or M diverges on any Π_k-$\mathcal{A}$-text for Y) and show that there is a $\emptyset^k$-recursive set R, i.e., $R \in \mathcal{C}$, such that M does not learn R.

Let $\Psi_0, \Psi_1, \ldots$ be an enumeration of all positive Π_k-$\mathcal{A}$-sentences where, in order to simplify notation, we assume that Ψ_0 is true for all sets. Recall that these sentences are monotone in the sense that whenever $A \subseteq B$ and $\Psi_n[A]$ is true, then so is $\Psi_n[B]$. We define inductively for all $\alpha \in \{0,1\}^*$ computable sets A_α, B_α, C_α and corresponding intervals

$$I_\alpha \quad = \quad \{X : A_\alpha \subseteq X \subseteq B_\alpha\} \quad .$$

The intervals are chosen such that whenever β extends α, then I_β is contained in I_α. Moreover, we will ensure for all α,

$$\Psi_i[A_\alpha] = \Psi_i[B_\alpha] \qquad \text{for all } i \le |\alpha|, \tag{15}$$

and hence, by monotonicity of the Ψ_i, all sets in I_α agree with respect to the predicates Ψ_0 through $\Psi_{|\alpha|}$.

The inductive definition starts with the interval I_λ bounded by $A_\lambda = \emptyset$ and $B_\lambda = \mathbb{N}$. Then given A_α and B_α, in order to define $A_{\alpha 0}$, $A_{\alpha 1}$, $B_{\alpha 0}$ and $B_{\alpha 1}$, we proceed as follows.

- Let $n = |\alpha|$. Let C_α be the union of A_α and the set containing every second element in $B_\alpha \setminus A_\alpha$, i.e., if $c_1, c_2, \ldots$ is a strictly ascending enumeration of the elements of $B_\alpha - A_\alpha$, then

  $$C_\alpha = A_\alpha \cup \{c_2, c_4, \ldots\}.$$

  So we have $A_\alpha \subset C_\alpha \subset B_\alpha$ and in this chain of inclusions, any set contains infinitely many more numbers than its predecessors.
- Let
  $$(D_\alpha, E_\alpha) \quad = \quad \begin{cases} (C_\alpha, B_\alpha) & \text{in case } \Psi_n[C_\alpha] \text{ holds,} \\ (A_\alpha, C_\alpha) & \text{otherwise.} \end{cases}$$
  (By monotonicity of Ψ_n, if the first case applies then Ψ_n is true for all sets between C_α and B_α, if the second case applies then Ψ_n is false for all sets between A_α and C_α. Hence by construction, $\{X : D_\alpha \subseteq X \subseteq E_\alpha\}$ is an infinite subinterval of I_α such that all sets in this interval agree with respect to Ψ_n.)
- Let x_α be the first element in $E_\alpha \setminus D_\alpha$. Define two infinite and disjoint subintervals $I_{\alpha 0}$ and $I_{\alpha 1}$ of I_α by letting $A_{\alpha 0} = D_\alpha$, $B_{\alpha 0} = E_\alpha \setminus \{x_\alpha\}$, and $A_{\alpha 1} = D_\alpha \cup \{x_\alpha\}$, $B_{\alpha 1} = E_\alpha$.

By construction, for all α the sets A_α, B_α and C_α are computable. Furthermore, for any given α, the inductive definition of these sets is computable relative to the oracle $\emptyset^k$, hence with access to this oracle we can compute programs that do not use an oracle and decide the sets A_α, B_α, and C_α. For a proof it suffices to observe that if we have already obtained a program that computes C_α, then by Remark 4.3 we can compute a Π_k-$\mathcal{A}$-formula that does not contain the predicate P and which is true if and only if $\Psi_n[C_\alpha]$ is. Furthermore, the truth value of this formula can be computed relative to the oracle $\emptyset^k$.

By construction, for any set F there is a unique set X_F that is contained in the intersection of all classes I_α such that α is a prefix of the characteristic function of F. If we let

$$\Psi_n^F = \begin{cases} \Psi_n & \text{if } \Psi_n[X] \text{ for all } X \in I_{F(0)F(1)\ldots F(n)}, \\ \# & \text{otherwise (that is, } \neg\Psi_n[X] \text{ for all } X \in I_{F(0)F(1)\ldots F(n)}) \end{cases},$$

where "#" is the pause symbol, then $\Psi_0^F, \Psi_1^F, \ldots$ is a Π_k-$\mathcal{A}$-text for X_F. Note that Ψ_n^F is just the value of the formula $\Psi_n[C_\alpha]$ where α is equal to $F(0)F(1)\ldots F(n-1)$. For any α, let $t_\alpha = \Psi_0^F\Psi_1^F\ldots\Psi_n^F$ where F is any extension α and the actual choice of F does not affect the value of t_α. Then t_α can be computed from α relative to $\emptyset^k$.

We show now that given any reliable learner M, there is a set F such that F and X_F are both computable relative to oracle $\emptyset^k$, but M does not learn the set X_F. Given any string α, the learner M cannot learn all of the uncountably many sets X_F with $F \succeq \alpha$ and so, as M is reliable, we can pick an extension $\beta \succeq \alpha$ with $M(t_\alpha) \neq M(t_\beta)$. This way, by successively extending an arbitrary initial string, we obtain a set F and a text $\Psi_0^F, \Psi_1^F, \ldots$ for X_F such that on this text, M changes its mind infinitely often. As this search algorithm is computable relative to $\emptyset^k$, the resulting sequence F and the set X_F are both computable relative to $\emptyset^k$. ∎

Theorem 8.2. *For every k, there is a countable $\mathcal{A}$-axiomatizable class that is limit-refuting learnable from Π_{k+1}-$\mathcal{A}$-text but not from Π_k-$\mathcal{A}$-text. In particular, for every k there is a countable $\mathcal{A}$-axiomatizable class that is not limit-refuting learnable from Π_k-$\mathcal{A}$-texts.*

Proof. Fix any k. A class $\mathcal{R}$ as in the theorem can be chosen as a suitable subset of the class $\mathcal{C}$ of all set that are computable relative to oracle $\emptyset^k$. The proof works by constructing for any learner M_e a set $R_e \in \mathcal{C}$ that on the one hand witnesses that M_e does not reliably and hence does not limit-refuting learn R_e from Π_k-$\mathcal{A}$-text and that on the other hand can easily be decoded from any Π_{k+1}-$\mathcal{A}$-text.

In the proof of Theorem 8.1, for a given reliable learner M we have constructed a set X_F that could not be learned by M from Π_k-$\mathcal{A}$-text. The construction was computable relative to the oracle $\emptyset^{k+1}$ and started with an initial interval bounded by $A_\lambda = \emptyset$ and $B_\lambda = \mathbb{N}$. Now assume that for given e, we apply this construction to any learner M_e where the initial interval is bounded by $A_\lambda^e = \{x : x < e\}$ and $B_\lambda^e = \{x : x \neq e\}$ (i.e., the initial interval contains exactly the sets where the least non-element is e). Then the construction remains computable relative to oracle $\emptyset^k$. Furthermore, in case M_e is reliable, we obtain a set that is contained in the interval between A_λ^e and B_λ^e and is not learned by M_e. In case M_e is not reliable, the construction might get stuck in the sense that it does not yield the characteristic function of a set but a partial $\{0,1\}$-valued function. Applying the construction this way to a given index e can be made uniform in the sense that there is a partial function ϕ with the following properties.

- The partial function ϕ is partial recursive relative to the oracle $\emptyset^k$.
- For all e, the partial function $\phi(e,.)$ is defined for all arguments less than or equal to e where

$$\phi(e,0) = 1, \ldots, \phi(e,e-1) = 1 \quad \text{and} \quad \phi(e,e) = 0\ .$$

  In particular, the least non-element of the set $R_e = \{x : \phi(e,x)\!\downarrow= 1\}$ is e.
- Whenever the learner M_e is reliable then $\phi(e,x)$ is defined for all x and the set R_e is not learned by M_e from Π_k-$\mathcal{A}$-text.

Then the class $\mathcal{C} = \{R_e : \phi(e,.) \text{ is total}\}$ is $\mathcal{A}$-axiomatizable. Furthermore, the class $\mathcal{C}$ is not reliably learnable and hence not limit-refuting learnable from Π_k-$\mathcal{A}$-text because for any reliable learner M_e, the set R_e is in $\mathcal{C}$ but by construction M_e fails to learn R_e from Π_k-$\mathcal{A}$-text.

It remains to show that $\mathcal{C}$ is limit-refuting learnable from Π_{k+1}-$\mathcal{A}$-text. Fix any oracle Turing machine that computes ϕ relative to oracle $\emptyset^k$. Let V contain all tuples (e,i,x,s) such that on input x, this machine either outputs i within s computation steps or does not terminate within s computation steps. Then V is computable relative to oracle $\emptyset^k$, hence there is a positive Π_{k+1}-$\mathcal{A}$-formula Δ that does not contain the predicate P such that

$$(e,i,x,s) \in V \quad \Leftrightarrow \quad \Delta(e,i,x,s).$$

Fix such a formula Δ and let

$$\widetilde{\Theta}^e_i(x) \quad \equiv \quad (\forall s)\,[\Delta(e,i,x,s)]$$

Note that $\widetilde{\Theta}^e_1(x)$ is equivalent to $x \in R_e$ whenever $\phi(e,.)$ is total. Since Δ is a Π_{k+1}-$\mathcal{A}$-formula, so are all the formulae $\widetilde{\Theta}^e_i$.

Then the following limit-refuting learner N learns all sets R_e from Π_{k+1}-texts. The learner outputs nothing but refutation symbols as soon as it has seen the formula $\forall x Px$, i.e., N correctly limit-refutes the set $\mathbb{N}$. On texts for any other set, we can assume that in the limit N is able to correctly infer the least non-element of the set to be learned. Thus it suffices to fix e and to show that under the assumption that N knows e, N behaves correctly on the class $\mathcal{D}_e$ of all sets that contain all $j < e$ but not e.

Let N output a refutation symbol in case it has seen, for some n, the formulae $\widetilde{\Theta}^e_0(\overline{n})$ and $\widetilde{\Theta}^e_1(\overline{n})$. As a consequence, N correctly limit-refutes all sets in $\mathcal{D}_e$ in case $\phi(e,.)$ is not total. Otherwise, let N output the formula (e) in (16) as long as N has seen formula (e) but has not seen any formula of the form (f).

$$\text{(e)} \quad (\forall x)\,[\widetilde{\Theta}^e_0(x) \vee \mathrm{P}x], \qquad \text{(f)} \quad \widetilde{\Theta}^e_0(\overline{n}) \wedge \mathrm{P}\overline{n} \tag{16}$$

By $\phi(e,.)$ being total, it follows that $\widetilde{\Theta}^e_0(\overline{n})$ is true iff $n \notin R_e$ and hence (e) is a subset index for R_e. As a consequence, N learns the set R_e and correctly limit-refutes all other sets in $\mathcal{D}_e$. We leave the details of the verification, which is similar to the corresponding argument in the proof of Theorem 8.1, to the reader. ∎

Recall from Remark 5.4 that in Presburger's arithmetic augmented with a predicate for the set Q of square numbers, we can define multiplication. By using this trick, from Theorems 8.1 and 8.2 on $\mathcal{A}$-axiomatizable classes we easily obtain corresponding, but slightly weaker results for $\mathcal{P}$-axiomatizable classes.

Corollary 8.3. *For any k there is a countable $\mathcal{P}$-axiomatizable class that is limit-refuting learnable from Π_{k+3}-$\mathcal{P}$-text but is not reliably learnable from Π_k-$\mathcal{P}$-text.*

Proof. We show that for any odd k there is a countable $\mathcal{P}$-axiomatizable classes $\mathcal{C}'$ that is limit-refuting learnable from Π_{k+2}-$\mathcal{P}$-text but is not reliably learnable from Π_k-$\mathcal{P}$-text. The separation result for $k+3$ versus k for even k is then immediate. So fix any odd k.

By Theorem 8.2 we can fix a countable $\mathcal{A}$-axiomatizable class $\mathcal{C}$ that can be limit-refuting learned from Π_{k+1}-$\mathcal{A}$-text but not from Π_k-$\mathcal{A}$-text. Let Q be the set of square numbers and for any set A let

$$A_Q = \{3n : n \in A\} \;\cup\; \{3n+1 : n \in Q\} \;\cup\; \{3n+2 : n \notin Q\}. \qquad (17)$$

We demonstrate that the class $\mathcal{C}_Q = \{A_Q : A \in \mathcal{C}\}$ witnesses the assertion of Corollary 8.3. The class $\mathcal{C}_Q$ is countable because $\mathcal{C}$ is. Given an $\mathcal{A}$-sentence Ψ_1 that axiomatizes $\mathcal{C}$, we obtain a $\mathcal{P}$-sentence $\Psi_0 \wedge \Psi'_1$ that axiomatizes $\mathcal{C}_Q$ as follows. First, let Ψ_0 be a formula of Presburger's arithmetic similar to Putnam's formula in Remark 5.4 that is true exactly for sets of the form A_Q, i.e., for sets that indeed have the set of squares and its complement encoded in them according to (17). Second, by the technique described in Remark 5.4 and using a predicate Q for the set of squares, transform the sentence Ψ_1 into an equivalent sentence over the language $\mathcal{P} \cup \{\mathrm{Q}\}$. Obtain Ψ'_1 by replacing in the new sentence each subformula $\mathrm{P}t$ and $\mathrm{Q}t$, t a term, by $\mathrm{P}(t+t+t)$ and by $\mathrm{P}(t+t+t+\overline{1})$, respectively. Then, by construction, for any set A the formula Ψ_1 is true in $\mathbf{M}(A)$ if and only if Ψ'_1 is true in $\mathbf{M}(A_Q)$. In summary, the $\mathcal{P}$ sentence $\Psi_0 \wedge \Psi'_1$ axiomatizes the class $\mathcal{C}_Q$ because its first conjunct is true exactly for sets of the form A_Q, while for any set of this form the second conjunct is true if and only if A is in $\mathcal{C}$, i.e., if A_Q is in $\mathcal{C}_Q$.

Next we show that given a learner M that limit-refuting learns the class $\mathcal{C}$ from Π_{k+1}-$\mathcal{A}$-text we can construct a learner M_Q that limit-refuting learns the class $\mathcal{C}_Q$ from Π_{k+2}-$\mathcal{P}$-text. For any positive Π_{k+1}-$\mathcal{A}$-formula Δ there is a positive Π_{k+2}-$\mathcal{P}$-formula Δ_Q such that for any set A the original formula Δ is true in $\mathbf{M}(A)$ if and only if Δ_Q is true in $\mathbf{M}(A_Q)$. Again, first we transform Δ into an equivalent $\mathcal{P} \cup \{\mathrm{Q}\}$-formula and then, second, in order to obtain the formula Δ_Q, for any term t and any subformula of one of the following forms, we replace in the formula obtained in the first step

$$\begin{aligned} \mathrm{P}t &\text{ by } \mathrm{P}(t+t+t), \\ \mathrm{Q}t &\text{ by } \mathrm{P}(t+t+t+1), \\ \neg\mathrm{Q}t &\text{ by } \mathrm{P}(t+t+t+2). \end{aligned} \tag{18}$$

The first transformation changes the matrix of the formula Δ into a Σ_2-formula while the second transformation does not add any quantifiers, so the original Π_{k+1}-formula Δ is changed into the Π_{k+2}-formula Δ_Q. Now any Π_{k+2}-$\mathcal{P}$-text for a set of the form A_Q contains the formulae Δ_Q for all Π_{k+1}-$\mathcal{A}$-formulae Δ that are true in $\mathbf{M}(A)$. Thus on any such text, the learner M_Q can recover a Π_{k+2}-$\mathcal{A}$-text for the set A and, by simulating M on the latter text, can correctly learn or limit-refute any set of the form A_Q. Furthermore, the learner M_Q limit-refutes any set that is not of the form A_Q if we require in addition that M_Q outputs a refutation symbol as long as it has not seen the formula $(\forall x)\,[\mathrm{P}(x+x+x+\overline{1}) \vee \mathrm{P}(x+x+x+\overline{2})]$ and as soon it has seen a formula $\mathrm{P}(\overline{3d+1})$ for a non-square d or a formula $\mathrm{P}(\overline{3d+2})$ for a square d. As a consequence, the learner M_Q limit-refuting learns the class $\mathcal{C}_Q$.

Finally, in order to show that the class $\mathcal{C}_Q$ cannot be reliably learned from Π_k-$\mathcal{P}$-text, we argue that any Π_k-$\mathcal{A}$-text for a set A can be effectively transformed into a Π_k-$\mathcal{P}$-text for the set A_Q. Then if there were a learner N_Q that limit-refuting learns $\mathcal{C}_Q$ from Π_k-$\mathcal{P}$-text, contrary to assumption we could construct a learner N that limit-refuting learns the class $\mathcal{C}$ from Π_k-$\mathcal{A}$-text. By definition, A is in $\mathcal{C}$ if and only if A_Q is in $\mathcal{C}_Q$, hence the learner N can simply transform any given text for a set A into a Π_k-$\mathcal{P}$-text for A_Q, then simulating N_Q on the new text.

In order to obtain a transformation of Π_k-$\mathcal{A}$-texts into Π_k-$\mathcal{P}$-texts as required, we show that there is a computable transformation $\Gamma \mapsto \Gamma'$ from Π_k-$\mathcal{P}$-sentences to Π_k-$\mathcal{A}$-sentences such that positive sentences are mapped to positive sentences and for any set X and any Γ, the truth values of $\Gamma[X_Q]$ and $\Gamma'[X]$ are the same. The translation of texts, which goes in the reverse direction, then can be done by enumerating all positive Π_k-$\mathcal{P}$-sentences Γ and placing exactly those Γ into the Π_k-$\mathcal{P}$-text for A_Q where Γ' eventually occurs in the given Π_k-$\mathcal{A}$-text for A. Note that the same transformation can be used to transform the indices output by the simulated Π_k-$\mathcal{P}$-learner N_Q for $\mathcal{C}_Q$ into indices of the $\mathcal{A}$-learner N for $\mathcal{C}$.

The transformation from Γ to Γ' works by picking two variables x_1 and x_2 that do not appear in Γ, then replacing in Γ every subformula of the form $\mathrm{P}t$, t a term, by the formula

$$\begin{aligned} (\forall x_1, x_2)\,[\,&(\overline{3}\cdot x_1 \neq t \vee \mathrm{P}x_1) \\ &\wedge (\overline{3}\cdot x_1 + \overline{1} \neq t \vee (x_1 \leq x_2 \cdot x_2 \vee (x_2+\overline{1})\cdot(x_2+\overline{1}) \leq x_1)) \\ &\wedge (\overline{3}\cdot x_1 + \overline{2} \neq t \vee x_1 \neq x_2 \cdot x_2)\,], \end{aligned}$$

while leaving the formula Γ unchanged otherwise. The subformula substituted for $\mathrm{P}t$ expresses that $x_1 \in A$ if $t = 3x_1$, that x_1 is a square (i.e., is not contained in the open interval bounded by two adjacent squares $(x_2)^2$ and $(x_2+1)^2$) if $t = 3x_1 + 1$ and that x_1 is not equal to any square $(x_2)^2$ if $t = 3x_1 + 2$. By construction, the mapping from Γ to Γ' preserves truth values as required and maps positive sentences to positive sentences. Furthermore, since k is odd, in case Γ is a Π_k-sentence, its last quantifier is universal or there are less than $k-1$ alternations of quantifiers, hence Γ' is again a Π_k-sentence.

The transformation from Γ to Γ' completes the reduction of the problem of learning $\mathcal{C}$ from Π_k-$\mathcal{A}$-texts to the problem of learning $\mathcal{C}_Q$ from Π_k-$\mathcal{P}$-texts. Since $\mathcal{C}$ cannot be learned reliably from Π_k-$\mathcal{A}$-texts, a corresponding assertion on non-learnability holds for $\mathcal{C}_Q$. ∎

Remark 8.4. We conjecture that the separation in Corollary 8.3 can be sharpened to the assertion that for any k, there is a countable $\mathcal{P}$-axiomatizable class that is limit-refuting learnable from Π_{k+1}-$\mathcal{P}$-text but is not reliably learnable from Π_k-$\mathcal{P}$-text.

9 Acknowledgements

We like to thank Thomas Wilke for very helpful discussion about his work on $\mathcal{S}$-axiomatizable classes. Furthermore, we are grateful to two anonymous referees of *Theoretical Computer Science* for their comments and corrections.

References

[1] Leonard M. Adleman and Manuel Blum: Inductive inference and unsolvability. *Journal of Symbolic Logic*, 56:891–900, 1991.

[2] J. Richard Büchi: On a decision method in restricted second order arithmetic. *Proceedings of the International Congress on Logic, Methodology and Philosophy of Science*, Stanford University Press, Stanford, California, 1960.

[3] Heinz-Dieter Ebbinghaus, Jörg Flum and Wolfgang Thomas: *Mathematical Logic*, Springer, 1994.

[4] William I. Gasarch, Mark G. Pleszkoch and Robert Solovay: Learning via queries in $[+, <]$. *The Journal of Symbolic Logic*, 57:53–81, 1992.

[5] William I. Gasarch and Carl H. Smith: Learning via queries, *Journal of the Association of Computing Machinery*, 39:649–674, 1992.

[6] Clark Glymour: Inductive Inference in the limit. *Erkenntnis* 22:23–31, 1985.

[7] E. Mark Gold: Language identification in the limit. *Information and Control*, 10: 447–474, 1967.

[8] Gunter Grieser: *Reflexion in der Induktiven Inferenz.* Diploma Thesis, Technische Hochschule Leipzig, 1996.

[9] Gunter Grieser: Reflecting Inductive Inference Machines and its Improvement by Therapy. *Proceedings of the Seventh Annual International Workshop on Algorithmic Learning Theory* (ALT), *Lecture Notes in Artificial Intelligence* 1160:325–336, Springer, 1996.

[10] Sanjay Jain: Learning with refutation. *Journal of Computer and Systems Sciences*, 57:356–365, 1998.

[11] Sanjay Jain, Efim Kinber, Rolf Wiehagen, Thomas Zeugmann: *On learning of functions refutably*, Manuscript, 2000.

[12] Sanjay Jain, Daniel Osherson, James Royer and Arun Sharma: *Systems that Learn*, revised edition of [25]. The MIT Press, Cambridge, Massachusetts, 1999.

[13] Sanjay Jain and Arun Sharma: Elementary formal systems, intrinsic complexity and procrastination. *Information and Computation*, 132:65–84, 1997.

[14] Klaus-Peter Jantke: Reflecting and self-confident inductive inference machines. *Proceedings of the Sixth Annual International Workshop on Algorithmic Learning Theory* (ALT), *Lecture Notes in Artificial Intelligence* 997:282–297, Springer, 1995.

[15] Kevin T. Kelly: *The Logic of Reliable Inquiry.* Oxford University Press, New York, 1996.

[16] Kevin T. Kelly and Clark Glymour: Inductive inference from theory-laden data. *Journal of Philosophical Logic*, 21:391–444, 1992.

[17] Steffen Lange and Phil Watson: Machine discovery in the presence of incomplete or ambiguous data. *Joint Proceedings of the Fourth International Workshop on Analogical and Inductive Inference* (AII) *and of the Fifth Workshop on Algorithmic Learning Theory* (ALT), *Lecture Notes in Artificial Intelligence* 872:438–452, Springer, 1994.

[18] Eric Martin and Daniel N. Osherson: Scientific discovery based on belief revision. *Journal of Symbolic Logic*, 62:1352-1370, 1997.

[19] Eric Martin and Daniel N. Osherson: *Elements of Scientific Inquiry.* The MIT Press, Cambridge, Massachusetts, 1998.

[20] Eric Martin, Arun Sharma and Frank Stephan: *Learning Power and Language Expressiveness. Forschungsbericht Mathematische Logik* 49, Mathematisches Institut, Universität Heidelberg, Heidelberg, 2000.

[21] Eliana Minicozzi. Some natural properties of strong-identification in inductive inference. *Theoretical Computer Science*, 2:345–360, 1976.

[22] Yasuhito Mukouchi and Setsuo Arikawa: Inductive inference machines that can refute hypothesis spaces. *Proceedings of the Fourth Annual International Workshop on Algorithmic Learning Theory* (ALT), *Lecture Notes in Artificial Intelligence* 744:123–136, Springer, 1993.

[23] Yasuhito Mukouchi and Setsuo Arikawa: Towards a mathematical theory of machine discovery from facts. *Theoretical Computer Science*, 137:53–84, 1995.

[24] Piergiorgio Odifreddi: *Classical Recursion Theory, Volumes* I *and* II. North Holland and Elsevier, Amsterdam, 1989 and 1999.

[25] Daniel Osherson, Michael Stob and Scott Weinstein: *Systems That Learn. An Introduction to Learning Theory for Cognitive and Computer Scientists.* Bradford — The MIT Press, Cambridge, Massachusetts, 1986.

[26] Mojzesz Presburger: Über die Vollständigkeit eines gewissen Systems der Arithmetik der ganzen Zahlen in welchem die Addition als einzige Operation hervortritt. *C. R. 1^er Congrès des Mathématiciens des Pays Slaves* (Warsaw) 92–101, 395, 1930.

[27] Hilary Putnam: Decidability and essential undecidability. *Journal of Symbolic Logic*, 22:39–54, 1957.

[28] Takeshi Shinohara: Inductive inference of monotonic formal systems from positive data. *New Generation Computing*, 8:371–384, 1991.

[29] Craig Smoryński: *Logical Number Theory I. An Introduction.* Springer, 1991.

[30] Robert I. Soare: *Recursively Enumerable Sets and Degrees. A Study of Computable Functions and Computably Generated Sets.* Springer, 1987.

[31] Wolfgang Thomas: Automata on Infinite Objects. *Handbook of Theoretical Computer Science*, Vol. B, edited by Jan van Leeuwen, p. 133–191, Elsevier, Amsterdam, 1990.

[32] Wolfgang Thomas: Classifying regular events in symbolic logic, *Journal of Computer and System Sciences*, 25:360–376, 1982.

[33] Thomas Wilke: Locally threshold testable languages of infinite words. *Proceedings of the Tenth Annual Symposium on Theoretical Aspects of Computer Science* (STACS), *Lecture Notes in Computer Science* 665:607–616, Springer, 1993.