Prescribed Learning of Indexed Families

Sanjay Jain*[1], Frank Stephan*[2] and Nan Ye[3]

[1] Department of Computer Science,
National University of Singapore, Singapore 117543, Republic of Singapore.
sanjay@comp.nus.edu.sg
[2] Department of Computer Science and Department of Mathematics,
National University of Singapore, Singapore 117543, Republic of Singapore.
fstephan@comp.nus.edu.sg
[3] Department of Computer Science and Department of Mathematics,
National University of Singapore, Singapore 117543, Republic of Singapore.
g0701171@nus.edu.sg

Abstract. This work extends studies of Angluin, Lange and Zeugmann on how learnability of a language class depends on the hypothesis space used by the learner. While previous studies mainly focused on the case where the learner chooses a particular hypothesis space, the goal of this work is to investigate the case where the learner has to cope with all possible hypothesis spaces. In that sense, the present work combines the approach of Angluin, Lange and Zeugmann with the question of how a learner can be synthesized. The investigation for the case of uniformly r.e. classes has been done by Jain, Stephan and Ye [6]. This paper investigates the case for indexed families and gives a special attention to the notions of conservative and non U-shaped learning.

1 Introduction

The goal of inductive inference [1, 2, 4] is to model the process of learning rigorously. Following many real-world scenarios, the learner observes more and more data which in the limit uniquely determines the concept (language) to be learnt. The learner is supposed to determine the target concept from the data it observes. Following the model of linguistics, the concept to be learnt is always considered to be an (often infinite) set of finite items which can be coded as natural numbers. The language to be learnt is chosen from a concept class $\{L_0, L_1, L_2, \ldots\}$ and the learner is using an explicit hypothesis space $\{H_0, H_1, H_2, \ldots\}$. This hypothesis space may be either the same as $\{L_0, L_1, L_2, \ldots\}$ (exact learning [1]) or chosen by the learner (class-preserving and class-comprising learning [8, 14, 15]) or imposed on the learner (prescribed and uniform learning [6]). Angluin [1] considered the important case that the concept class and hypothesis class are both given by an indexed family, that is, the class is uniformly recursive. She has given a characterization when such a class is explanatorily learnable and introduced also important variants like consistent and conservative learning.

The goal of the present work is to study prescribed and uniform learning and to contrast the results obtained for them to the well-studied cases of exact, class-preserving and class-comprising

* Supported in part by NUS grant number R252-000-212-112 and R252-000-308-112.

learning. The idea that the learner has to accept a given choice of the hypothesis class is not completely new; besides the case of exact learning (for which the results would be equivalent to the (not considered case of) class-preserving prescribed learning), it has also been considered under the framework of synthesis of learners. But the models like those considered by Zilles [17, 18] differ from the scenario in the present work. Jain, Stephan and Ye [6] have studied the more general case of uniformly r.e. concept and hypothesis spaces in a separate paper. The main difference to the setting in the r.e. case is that there it is more reasonable to consider class-preserving-uniformly and class-preserving-prescribed learning instead of uniform and prescribed learning. Furthermore, the relation between non U-shaped learning and conservative learning depends crucially on the indexed family nature of the hypothesis space.

The study of prescribed and uniform learning is done for the criteria of finite learning (Section 2), conservative learning (Section 3), non U-shaped learning (Section 4) and the various notions of monotonic learning (Section 5).

In the following, we will provide more details, but have to introduce some formal notations first. Let $\mathbb{N}$ be the set of natural numbers. Let $\langle \cdot, \cdot \rangle$ be a fixed pairing function: a recursive bijective mapping from $\mathbb{N}^2$ to $\mathbb{N}$. Let $|S|$ denote the cardinality of set S. Let $\overline{S}$ denote $\mathbb{N} - S$. Let $\min(S)$ and $\max(S)$ denote the minimum and maximum of a set respectively. Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ denote a fixed acceptable numbering of the partial recursive functions from $\mathbb{N}$ to $\mathbb{N}$. In some cases, we use φ_i as a function of two arguments. In such cases one implicitly assumes a pairing function being used to code the inputs: thus, $\varphi_i(x, y)$ means $\varphi_i(\langle x, y \rangle)$. The set W_e is the domain of φ_e. The set $\mathbb{K} = \{e : e \in W_e\}$ is the diagonal halting problem which is used as a standard example of an r.e. but nonrecursive set. Let $\mathbb{K}_t$ denote the set of elements enumerated into $\mathbb{K}$ within t steps, via some standard enumeration procedure. We assume without loss of generality that $\mathbb{K}_0 = \emptyset$. Basic formal definitions of learning are given as follows.

Definition 1. A learner is a mapping from $(\mathbb{N} \cup \{\#\})^*$ to $\mathbb{N} \cup \{?\}$. Let M be a given learner, $\{L_0, L_1, L_2, \ldots\}$ be a language class and $\{H_0, H_1, H_2, \ldots\}$ be a hypothesis space. M itself is a partial-recursive function and $\{L_0, L_1, L_2, \ldots\}, \{H_0, H_1, H_2, \ldots\}$ are indexed families of subsets of the natural numbers, that is, the mappings $e, x \mapsto L_e(x)$ and $e, x \mapsto H_e(x)$ are recursive functions from $\mathbb{N} \times \mathbb{N}$ to $\{0, 1\}$. Let σ, τ, ρ range over $(\mathbb{N} \cup \{\#\})^*$. Furthermore, let $\sigma \subseteq \tau$ denote that τ is an extension of σ as a string. Call T a text if T is a total function which maps $\mathbb{N}$ to $\mathbb{N} \cup \{\#\}$; call T a text for L_a iff the numbers occurring in T are exactly those in L_a. Given a class $\{L_0, L_1, L_2, \ldots\}$, one can uniformaly (in n) generate a text for n — such a text is called canonical text for L_n.

A learner *converges* [4] on T to b iff there is an n with $M(T[m]) = b$ for all $m \geq n$; here $T[m]$ is the finite string consisting of the first m members of T.

A learner M is *total* if $M(\sigma)$ is defined for all finite strings σ in $(\mathbb{N} \cup \{\#\})^*$. Without loss of generality, for the learning criteria in this paper, learners can be assumed to be total and this is done from now onwards.

A learner M is *finite* [4] if for every text T there is one index e such that for all n, either $M(T[n]) = ?$ or $M(T[n]) = e$.

A learner M is *confident* [10] if M is total and converges on every text T to a hypothesis.

A learner M is *conservative* [1] if for all σ, τ with $H_{M(\sigma)} \neq H_{M(\sigma\tau)}$ there is an x occurring in $\sigma\tau$ such that $x \notin H_{M(\sigma)}$.

A learner M is *non U-shaped* [3] if there are no a and $\sigma, \tau, \rho \in (L_a \cup \{\#\})^*$ such that $H_{M(\sigma)} = H_{M(\sigma\tau\rho)} = L_a$ and $H_{M(\sigma\tau)} \neq L_a$. In other words, M never changes from a correct to an incorrect and then back to a correct hypothesis.

A learner M is *decisive* [3] if there are no σ, τ, ρ such that $H_{M(\sigma\tau\rho)} = H_{M(\sigma)}$ and $H_{M(\sigma\tau)} \neq H_{M(\sigma)}$. In other words, M never returns to a once abandoned hypothesis (even semantically).

A learner M is *monotonic* [5] if for every L_a and for every $\sigma, \tau \in (L_a \cup \{\#\})^*$ the inclusion $L_a \cap H_{M(\sigma)} \subseteq L_a \cap H_{M(\sigma\tau)}$ holds. A learner M is *strong-monotonic* [5] if for all $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*$ the inclusion $H_{M(\sigma)} \subseteq H_{M(\sigma\tau)}$ holds.

Here note that ? is not considered as a conjecture and thus the constraints in conditions like conservative, monotonic, strong-monotonic, non U-shaped and decisive refer only to inputs where M makes a conjecture and does not output ?: so, more formally, a learner would be strong-monotonic iff for all σ, τ, $M(\sigma) \neq ?$ and $M(\sigma\tau) \neq ?$ implies $H_{M(\sigma)} \subseteq H_{M(\sigma\tau)}$. Similarly for the other criteria.

Finite learning is quite restrictive since the learner has to make up its mind without having viewed all of the available infinite information. Learning in the limit (or just "learning") is more powerful since the learner can revise its hypothesis a finite but arbitrary number of times. A similar observation has been made by Staiger [12] with respect to accepting ω-languages by Turing machines.

In this paper we will only be concerned about learning indexed families and using hypothesis spaces which are also indexed families. Angluin, Kapur, Lange and Zeugmann [1, 7–9, 13–15] studied how learnability of the family to be learned depends on the hypothesis space $\{H_0, H_1, H_2, \ldots\}$ used by the learner. To formalize this, they introduced the notions of exact, class-preserving and class-comprising learning. In addition to this we consider notions like uniform and prescribed learning [6]. Here I ranges over properties of the learner as defined in Definition 1, so I stands for "conservative", "finite", "monotonic" and so on.

Definition 2. In the following, let $\{L_0, L_1, L_2, \ldots\}$ and $\{H_0, H_1, H_2, \ldots\}$ be indexed families.

The class $\{L_0, L_1, L_2, \ldots\}$ is explanatory learnable [4] with hypothesis space $\{H_0, H_1, H_2, \ldots\}$ iff there is a learner M which converges on every text of a language L_a to a hypothesis b such that $H_b = L_a$.

For a property I from Definition 1, $\{L_0, L_1, L_2, \ldots\}$ is I learnable with hypothesis space $\{H_0, H_1, H_2, \ldots\}$ iff there is a learner M which explanatory learns $\{L_0, L_1, L_2, \ldots\}$ using hypothesis space $\{H_0, H_1, H_2, \ldots\}$ and furthermore satisfies the requirement I.

The class $\{L_0, L_1, L_2, \ldots\}$ is class-comprisingly I learnable iff it is I learnable with some hypothesis space $\{H_0, H_1, H_2, \ldots\}$; note that learnability automatically implies $\{L_0, L_1, L_2, \ldots\} \subseteq \{H_0, H_1, H_2, \ldots\}$.

The class $\{L_0, L_1, L_2, \ldots\}$ is class-preservingly I learnable iff it is I learnable with some hypothesis space $\{H_0, H_1, H_2, \ldots\}$ satisfying $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$.

The class $\{L_0, L_1, L_2, \ldots\}$ is exactly I learnable iff it is I learnable with $\{L_0, L_1, L_2, \ldots\}$ itself taken as hypothesis space.

The class $\{L_0, L_1, L_2, \ldots\}$ is prescribed I learnable iff it is I learnable with respect to every hypothesis space $\{H_0, H_1, H_2, \ldots\}$ such that $\{L_0, L_1, L_2, \ldots\} \subseteq \{H_0, H_1, H_2, \ldots\}$.

The class $\{L_0, L_1, L_2, \ldots\}$ is uniformly I learnable iff there is a recursive enumeration of partial-recursive functions $M_0, M_1, M_2, \ldots$ such that whenever φ_e is a decision-procedure $b, x \mapsto H_b(x)$ for an indexed family $\{H_0, H_1, H_2, \ldots\} \supseteq \{L_0, L_1, L_2, \ldots\}$ then M_e is an I learner for $\{L_0, L_1, L_2, \ldots\}$ using this hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

The class $\{L_0, L_1, L_2, \ldots\}$ is class-preserving-uniformly I learnable iff there is a recursive enumeration of partial-recursive functions $M_0, M_1, M_2, \ldots$ such that whenever φ_e is a decision-procedure $b, x \mapsto H_b(x)$ for an indexed family $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$ then M_e is an I learner for $\{L_0, L_1, L_2, \ldots\}$ using this hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

Remark 3. For the basic notion explanatory learning (= learning in the limit), all these notions are the same. This is so, as class comprising learning is the same as exact learning for explanatory learning [14]. Furthermore, given any hypothesis space $\{H_0, H_1, H_2, \ldots\}$ covering $\{L_0, L_1, L_2, \ldots\}$, for each a, one can find in the limit a b such that $L_a = H_b$.

Exact finite learning and class comprising finite learning are the same [14]. For strong-monotonic, monotonic and conservative learning, there is a proper hierarchy for learning from exact, class preserving and class comprising hypothesis spaces [14]. For every criterion I, the following implications hold:

- Every uniformly I learnable family is also class-preserving-uniformly I learnable and prescribed I learnable.
- Every class-preserving-uniformly I learnable family and every prescribed I learnable family is also exactly I learnable.
- Every exactly I learnable family is also class-preservingly I learnable.
- Every class-preservingly I learnable family is also class-comprisingly I learnable.

It depends on the actual choice of I what other implications hold (besides the transitive ones).

For example, for confident learning, the class containing all $\{x\}$ where $|W_x| < \infty$ and $\{x, y\}$ where $x \neq y$ is not class-preservingly but class-comprisingly confidently learnable. Although confident learnability becomes more general in the class-comprising case, one can show that it coincides for all other criteria from Definition 2. Suppose that N is an exact confident learner for the class $\{L_0, L_1, L_2, \ldots\}$ and e is given such that φ_e is total and the hypothesis space $\{H_0, H_1, H_2, \ldots\}$ satisfies $H_d = \{x : \varphi_e(\langle d, x\rangle) = 1\}$ for all d and $\{L_0, L_1, L_2, \ldots\} \subseteq \{H_0, H_1, H_2, \ldots\}$. Then M_e simulates the learner N as follows: if N on text T converges to a then M_e on text T converges to the least b such that $H_b = L_a$.

From the definition of uniform learning, we can easily obtain the following useful lemma.

Lemma 4. *Let $\mathcal{L}$ be a uniformly I learnable indexed family. If $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2, \ldots$ is a recursive enumeration of indexed family hypothesis spaces for $\mathcal{L}$, then there exists a recursive enumeration of learners $M_0, M_1, M_2, \ldots$ such that M_n I learns $\mathcal{L}$ with respect to $\mathcal{H}_n$.*

We will often make use of the following simple set in our proofs.

Definition 5. Define $S = \cup_{n=0,1,2,\ldots} J_n$, where J_n contains for each $e < n$ the first element, if any, of W_e enumerated from $I_n = \{2^n - 1, 2^n, 2^n + 1, \ldots, 2^{n+1} - 2\}$. Then: S is recursively enumerable; S intersects with every infinite recursively enumerable set; for every n there is an m in I_n which is not in S. In other words, S is a simple set [11]. Let S_t be the set of elements enumerated into S within t steps via some standard procedure. Here we take $S_0 = \emptyset$.

In the following sections, without loss of generality we assume for $i, j < |\{L_0, L_1, L_2, \ldots\}|$ that $L_i = L_j$ implies $i = j$.

2 Finite Learning

Finite learning or one-shot learning requires the learner to make a correct guess using only finite amount of information. So it is not a surprise that this criterion turns out to be very restrictive for prescribed and uniform learning, as shown below. The following theorem gives some characterization results and separates various notions of finite learning. It can be shown that class comprising finitely learnable classes are also exact finitely learnable [14].

Theorem 6. *Let $\{L_0, L_1, L_2, \ldots\}$ be exactly finitely learnable.*
(a) $\{L_0, L_1, L_2, \ldots\}$ *is not uniformly finitely learnable.*
(b) $\{L_0, L_1, L_2, \ldots\}$ *is class-preserving-uniformly finitely learnable.*
(c) $\{L_0, L_1, L_2, \ldots\}$ *is prescribed finitely learnable iff the class is finite and for all $i, j < |\{L_0, L_1, L_2, \ldots\}|$, either $i = j$ or $L_i \not\subset L_j$.*

Hence, the class $\{\{0\}, \{1\}, \{2\}, \ldots\}$ is exactly finitely learnable but not prescribed finitely learnable; furthermore, a class is prescribed finitely learnable iff it is finitely learnable and finite.
Proof. (a) Let $G_e = L_e$ if $e < |\{L_0, L_1, L_2, \ldots\}|$; otherwise let G_e be some recursive set outside $\{L_0, L_1, L_2, \ldots\}$. Note that the numbering $\{G_0, G_1, G_2, \ldots\}$ is introduced in order to handle finite and infinite classes uniformly (for infinite $\{L_0, L_1, L_2, \ldots\}$, note that $G_i = L_i$). Suppose $\{L_0, L_1, L_2, \ldots\}$ is uniformly finitely learnable as witnessed by the recursive enumeration of learners $M_0, M_1, M_2, \ldots$. Let F be a recursive set such that no finite variant of F is in $\{L_0, L_1, L_2, \ldots\}$. By Kleene's recursion theorem [16], there exists an e such that for every $d \in \mathbb{N}$ and $c \in \{0, 1\}$,

$$\varphi_e(2d + c, x) = \begin{cases} F(x), & \text{if } M_e \text{ outputs } 2d + c \text{ as first grammar on the} \\ & \text{canonical text } T_d \text{ for } L_d \text{ in up to } x \text{ steps;} \\ G_d(x), & \text{otherwise.} \end{cases}$$

For this e, φ_e defines an indexed family hypothesis space $\{H_0, H_1, H_2, \ldots\}$ which is a superclass of $\{L_0, L_1, L_2, \ldots\}$. By construction, M_e does not finitely learn any language in $\{L_0, L_1, L_2, \ldots\}$ with respect to the given hypothesis space.

(b) Let N be an exact finite learner for $\{L_0, L_1, L_2, \ldots\}$. We define a recursive enumeration of learners $M_0, M_1, M_2, \ldots$ that class-preserving-uniformly learn $\{L_0, L_1, L_2, \ldots\}$. For $n \in \mathbb{N}$, $M_n(T[t])$ is defined as follows. If for all $k \leq t$, $N(T[k]) = ?$, then output ?. Otherwise, $N(T[k]) \neq ?$ for some $k \leq t$. Search for the minimum $i \leq t$ such that for all $j \in \mathrm{content}(T[k])$, $\varphi_n(i, j) \downarrow = 1$.

Output i if found, else output ?. It is easy to verify that M_e is a finite learner for $\{L_0, L_1, L_2, \ldots\}$ whenever φ_e defines a class-preserving hypothesis space for $\{L_0, L_1, L_2, \ldots\}$.

(c) If $\{L_0, L_1, L_2, \ldots\} = \{L_0, \ldots, L_n\}$ for some $n \in \mathbb{N}$ and $L_i \not\subseteq L_j$ for all $i, j < n+1$ with $i \neq j$, then it is prescribed finitely learnable as follows: Given a hypothesis space $\{H_0, H_1, H_2, \ldots\}$, let $i_0, \ldots, i_n$ be indices for $L_0, \ldots, L_n$ in $\{H_0, H_1, H_2, \ldots\}$ respectively and let $x_{k,l}$ be an element in $L_k - L_l$ for all $k, l \leq n$ with $k \neq l$. On input $T[t]$, search for the least k such that $x_{k,l} \in \text{content}(T[t])$ for all $l \leq n$ with $l \neq k$. If such k is found, output i_k and stop; otherwise output ?. It is easy to verify that the above learner finitely learns $\{L_0, L_1, L_2, \ldots\}$ using hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

Suppose $\{L_0, L_1, L_2, \ldots\}$ is prescribed finitely learnable but infinite. Let I_n be as defined in Definition 5. For each $m \in \mathbb{N}$, let $n \in \mathbb{N}$ be the number such that $m \in I_n$, then H_m is defined as follows:

$$H_m(x) = \begin{cases} 1 - L_{x-m-t}(x), & \text{if } x \geq m + t \text{ and } m \in S_{t+1} - S_t \text{ for some } t \in \{0, 1, \ldots, x\}; \\ L_n(x), & \text{otherwise.} \end{cases}$$

Note that $\{H_0, H_1, H_2, \ldots\} \supseteq \{L_0, L_1, L_2, \ldots\}$ and $\{H_0, H_1, H_2, \ldots\}$ is an indexed family. Let M be a finite learner for $\{L_0, L_1, L_2, \ldots\}$ with respect to $\{H_0, H_1, H_2, \ldots\}$. For each $i \in \mathbb{N}$, let $f(i)$ be the first index which M outputs on the canonical text T_i for L_i. Note that $f(i) \notin S$ and $f(i) \in I_i$; hence $f(i) \neq f(j)$ for distinct i, j. Thus $f(0), f(1), f(2), \ldots$ is a an infinite r.e. subset of $\overline{S}$, a contradiction. Hence, $\{L_0, L_1, L_2, \ldots\}$ must be finite. In addition, there do not exist i, j with $i \neq j$ and $L_i \subset L_j$ — otherwise, a σ such that content(σ) $\subseteq L_i$ on which the learner outputs a hypothesis for L_i can be extended to a text for L_j; thus the learner fails to finitely learn L_j.

3 Conservative Learning

Conservative learning is non-trivial, in the sense that there is an infinite indexed family which is uniformly conservatively learnable. This is shown in the next example.

Example 7. *Let $L_a = \mathbb{N} - \{a\}$ for all $a \in \mathbb{N}$. Then $\{L_0, L_1, L_2, \ldots\}$ is uniformly conservatively learnable.*

Proof. For $i \in \mathbb{N}$, define M_i as follows: given a text T, at time t, find the least $m \in \mathbb{N}$ such that $m \notin \text{content}(T[t])$. Find the least $j \leq t$ such that $\varphi_i(j, m) = 0$ and for all $k \leq t$ with $k \neq m$, $\varphi_i(j, k) = 1$. Output j if found; otherwise output ?. It is easy to verify that $M_0, M_1, M_2, \ldots$ witness that $\{L_0, L_1, L_2, \ldots\}$ is uniformly conservatively learnable.

The class used in above example consists of co-finite sets only. The next result shows that this is necessary for uniform conservative learning.

Theorem 8. *If $\{L_0, L_1, L_2, \ldots\}$ is uniformly conservatively learnable then every set L_a is cofinite. Moreover, there is a recursive function r bounding the non-elements of L_a for all $a < |\{L_0, L_1, L_2, \ldots\}|$.*

Proof. Let S be as in Definition 5. Furthermore, let $G_a = L_a$ if $a < |\{L_0, L_1, L_2, \ldots\}|$ and $G_a = \mathbb{N}$ otherwise.

We define a sequence of hypothesis spaces $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2, \ldots$, where for each $n \in \mathbb{N}$ the space $\mathcal{H}_n = \{H_0^n, H_1^n, H_2^n, \ldots\}$ is defined as follows:

$$H^n_{\langle i,j\rangle} = \begin{cases} G_i, & \text{if } j \notin S \text{ and } j > n; \\ G_i \cup \{t+1, t+2, t+3, \ldots\}, & \text{if } j \in S_{t+1} - S_t \text{ and } j > n; \\ \mathbb{N}, & \text{if } j \leq n. \end{cases}$$

Note that the case distinction covers all cases as $S_0 = \emptyset$. Furthermore, $\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2, \ldots$ is a recursive enumeration of indexed families. Since $\{L_0, L_1, L_2, \ldots\}$ is uniformly conservatively learnable, there exists a recursive enumeration of learners $M_0, M_1, M_2, \ldots$ such that for all n, M_n conservatively learns $\{L_0, L_1, L_2, \ldots\}$ with respect to $\mathcal{H}_n$.

For all $a < |\{L_0, L_1, L_2, \ldots\}|$ and $n \in \mathbb{N}$, let $e = \langle v(a,n), w(a,n)\rangle$ be the first number found (in some dovetailing search) such that M_n outputs e on the canonical text T_a of L_a and one of the following conditions hold:

(a) $w(a,n) \in S$ and $L_a \subseteq H^n_{\langle v(a,n), w(a,n)\rangle}$ (note that this can be verified by finding a t with $w(a,n) \in S_t$ and checking $L_a(x) \leq H^n_{\langle v(a,n), w(a,n)\rangle}(x)$ for all $x \leq t$);
(b) $v(a,n) = a$;
(c) $w(a,n) \leq n$.

Note that such $e = \langle v(a,n), w(a,n)\rangle$ exist for all n. First note that for every $a < |\{L_0, L_1, L_2, \ldots\}|$ there is an n such that either $w(a,n) \leq n$ or $w(a,n) \in S$: Otherwise the set $\{w(a,n) : n \in \mathbb{N}\}$ would be an infinite r.e. set disjoint to S, a contradiction as S is simple. Hence there is a recursive function u which searches for this n; that is, $w(a,u(a)) \leq u(a) \vee w(a,u(a)) \in S$ for all $a < |\{L_0, L_1, L_2, \ldots\}|$.

It is easy to see that there is a further recursive function r such that either $r(a) = 0 \wedge w(a,u(a)) \leq u(a)$ or $r(a) > 0 \wedge w(a,u(a)) \in S_{r(a)}$. Note that $L_a \subseteq H^{u(a)}_{\langle v(a,u(a)), w(a,u(a))\rangle}$ and $\{r(a), r(a)+1, r(a)+2, \ldots\} \subseteq H^{u(a)}_{\langle v(a,u(a)), w(a,u(a))\rangle}$. Since each M_n is conservative, it follows that $L_a = H^{u(a)}_{\langle v(a,u(a)), w(a,u(a))\rangle}$ and thus $\mathbb{N} - L_a$ contains only elements below $r(a)$ for all $a < |\{L_0, L_1, L_2, \ldots\}|$.

For prescribed conservative learning, we have a less stringent necessary condition as compared to uniform conservative learning.

Theorem 9. *If $\{L_0, L_1, L_2, \ldots\}$ is prescribed conservatively learnable then almost every set in $\{L_0, L_1, L_2, \ldots\}$ is cofinite. Moreover, there is a recursive function r bounding the non-elements of the cofinite L_a.*

Proof. Let S and I_n be as in Definition 5. Define a hypothesis space $\{H_0, H_1, H_2, \ldots\}$ as follows. For $m \in \mathbb{N}$, define H_m as follows: suppose n is such that $m \in I_n$; then let

$$H_m = \begin{cases} L_n, & \text{if } m \notin S; \\ L_n \cup \{m+t, m+t+1, \ldots\}, & \text{if } m \in S_{t+1} - S_t \text{ for some } t \in \mathbb{N}. \end{cases}$$

The hypothesis space $\{H_0, H_1, H_2, \ldots\}$ is an indexed family which is a superclass of $\{L_0, L_1, L_2, \ldots\}$. Let M be a learner for $\{L_0, L_1, L_2, \ldots\}$ with respect to $\{H_0, H_1, H_2, \ldots\}$. Let m_n be the first index output by M on the canonical text T_n for L_n such that $m_n \in I_n$ (m_n may not always be defined). Then the set $\{m_n : n \in \mathbb{N}$ and m_n exists$\}$ is recursively enumerable. Let e be an index for this set. Suppose there are infinitely many coinfinite sets in $\{L_0, L_1, L_2, \ldots\}$, then there exists an $a > e$ such that L_a is coinfinite. Since L_a is coinfinite, m_a exists (as only indices in I_n can be indices for coinfinite L_n). Furthermore, $m_a \in S$ because m_a is the only element in $W_e \cap I_a$. But this implies $H_{m_a} \supset L_a$. Hence M cannot be conservative. The function r can be found by techniques similar to those in the proof of Theorem 8.

The above is not a characterization as the class of all cofinite sets is not learnable in the limit. On the other hand, one can get the following characterization: Assume that $\{L_0, L_1, L_2, \ldots\}$ is exactly conservatively learnable. (a) $\{L_0, L_1, L_2, \ldots\}$ is uniformly conservatively learnable iff there is a recursive function r such that, for all a and x, $x > r(a) \Rightarrow x \in L_a$. (b) $\{L_0, L_1, L_2, \ldots\}$ is prescribed conservatively learnable iff there is a recursive function r such that, for almost all a and for all x, $x > r(a) \Rightarrow x \in L_a$. The only if direction can be shown as in above theorems. If direction, can be proven easily using standard techniques.

The next example shows that there is also a learnable class of cofinite sets which is not prescribed conservatively learnable; this class is still class-comprising conservative learnable.

Example 10. *Suppose $\{L_0, L_1, L_2, \ldots\}$ consists of all sets of the form $\mathbb{N} - \{a\}$ and all sets of the form $\mathbb{N} - \{a, b\}$ where $a < b$, $a \in \mathbb{K} - \mathbb{K}_b$. Then $\{L_0, L_1, L_2, \ldots\}$ is class-preservingly conservatively learnable but not prescribed conservatively learnable. More precisely, $\{L_0, L_1, L_2, \ldots\}$ is not conservatively learnable with respect to the hypothesis space $\{H_0, H_1, H_2, \ldots\}$ given by $H_{2a} = \mathbb{N} - \{a\}$ and $H_1, H_3, H_5, \ldots$ being an enumeration of those sets in $\{L_0, L_1, L_2, \ldots\}$ with 2 elements in the complement.*

Proof. Without loss of generality assume that, from i, one can effectively determine $\mathbb{N} - L_i$.

We define a class-preserving hypothesis space $\{H'_0, H'_1, H'_2, \ldots\}$ by setting $H'_{2i+1} = L_i$, $H'_{2i} = \{x \in \mathbb{N} : (x \neq i) \wedge \neg(x > i$ and $i \in \mathbb{K}_{x+1} - \mathbb{K}_x)\}$, for all $i \in \mathbb{N}$. The class $\{L_0, L_1, L_2, \ldots\}$ is conservatively learnable with respect to $\{H'_0, H'_1, H'_2, \ldots\}$ as follows: Given $T[t]$ as input, learner first finds the least two elements $n_1, n_2 \notin \mathrm{content}(T[t])$, with $n_1 < n_2$. If $n_1 \notin \mathbb{K}_{j+1} - \mathbb{K}_j$, for any $j \in \mathrm{content}(T[t])$, then the learner outputs a grammar for H'_{2n_1}. Otherwise there is a $j \in \mathrm{content}(T[t])$ such that $n_1 \in \mathbb{K}_{j+1} - \mathbb{K}_j$. Now, if $n_2 \leq j$, then the learner outputs $2r + 1$, for the least r such that $L_r = \mathbb{N} - \{n_1, n_2\}$; otherwise the learner outputs $2r + 1$, for the least r such that $L_r = \mathbb{N} - \{n_1\}$. It is easy to verify that the above learner conservatively learns $\{L_0, L_1, L_2, \ldots\}$ using the class-preserving hypothesis space $\{H'_0, H'_1, \ldots\}$.

Suppose a learner M conservatively learns $\{L_0, L_1, L_2, \ldots\}$ with respect to the given hypothesis space $\{H_0, H_1, H_2, \ldots\}$. For any $n \in \mathbb{N}$, we can decide whether $n \in \mathbb{K}$ as follows: Let T be a text for $\mathbb{N} - \{n\}$. Then $M(\sigma) = 2n$ for some $\sigma \subset T$ because H_{2n} is the only correct hypothesis for $\mathbb{N} - \{n\}$. Let t be the minimum number in $\mathbb{N} - (\{0, 1, 2, \ldots, n\} \cup \mathrm{content}(\sigma))$. If $n \in \mathbb{K}$, but $n \notin \mathbb{K}_t$, then $\mathbb{N} - \{n, t\} \in \{L_0, L_1, L_2, \ldots\}$ and σ can be extended to be a text T' for $\mathbb{N} - \{n, t\}$, violating the conservativeness of M. Also, clearly if $n \in \mathbb{K}_t$, then $n \in \mathbb{K}$. Hence $n \in \mathbb{K}$ iff $n \in \mathbb{K}_t$. So we have an effective procedure to decide whether $n \in \mathbb{K}$, a contradiction.

The following theorem shows that class-preserving-uniformly conservative learnability and prescribed conservative learnability are not comparable.

Theorem 11. (a) *There exists a class $\{L_0, L_1, L_2, \ldots\}$ which is class-preserving-uniformly conservatively learnable, but not prescribed conservatively learnable.*
(b) *There exists a class $\{L_0, L_1, L_2, \ldots\}$ which is prescribed conservatively learnable but not class-preserving-uniformly conservatively learnable.*

Proof. (a) Let $L_a = \{a\}$. Then $\{L_0, L_1, L_2, \ldots\}$ is clearly class-preserving-uniformly conservatively learnable. However, by Theorem 9, $\{L_0, L_1, L_2, \ldots\}$ is not prescribed conservatively learnable.

(b) Let $L_0 = \emptyset$. Let $L_{i+1} = \{x : x \geq i\}$. Then, $\{L_0, L_1, L_2, \ldots\}$ is prescribed conservatively learnable as follows. Suppose hypothesis space $\{H_0, H_1, H_2, \ldots\}$ is given. Suppose z is such that $H_z = \emptyset$. On input $T[t]$, if content$(T[t]) = \emptyset$, then output z. If content$(T[t])$ is contained in the previous hypothesis, then repeat the previous hypothesis. Otherwise, let i be minimal such that $i \in$ content$(T[t])$. Let $j \leq t$ be minimal such that $H_j(x) = 1$, for $x \in \{i, i+1, i+2, \ldots, t\}$ and $H_j(x) = 0$, for $x < i$. If such a j exists, then output j, otherwise repeat the previous hypothesis. It is easy to verify that the above learner conservatively learns $\{L_0, L_1, L_2, \ldots\}$ using hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

Now we show that $\{L_0, L_1, L_2, \ldots\}$ is not class-preserving-uniformly conservatively learnable. Suppose by way of contradiction that $M_0, M_1, M_2, \ldots$ witnesses that $\{L_0, L_1, L_2, \ldots\}$ is class-preserving-uniformly conservatively learnable. Then by Kleene's recursion theorem [16], there exists an e such that φ_e may be defined as follows:

$$\varphi_e(2i+1, x) = L_{i+1}(x);$$
$$\varphi_e(2i, x) = \begin{cases} 1, & \text{if } 2i \text{ is the first hypothesis output by } M_e \text{ on } \#^\infty \\ & \quad \text{and this hypothesis is output by } M_e \text{ within } x \text{ steps;} \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to verify that M_e does not conservatively identify $\{L_0, L_1, L_2, \ldots\}$ using the class-preserving hypothesis space given by φ_e.

4 Non U-Shaped Learning

Every conservative learner is clearly non U-shaped. Furthermore, one can modify a conservative learner to be decisive by only changing to a new hypothesis if it is consistent with the input.

The following theorem thus shows that non U-shaped learning is equivalent to conservative learning in the case of exact, class-preserving and class-preserving-uniform learning.

Theorem 12. *Assume that the class $\{L_0, L_1, L_2, \ldots\}$ is class-preserving-uniformly non U-shaped learnable. Then $\{L_0, L_1, L_2, \ldots\}$ is already class-preserving-uniformly conservatively learnable by the same learner. The same applies for exact and class-preserving learning.*

Proof. We show that if M non U-shaped learns $\{L_0, L_1, L_2, \ldots\}$ with respect to a class-preserving hypothesis space $\{H_0, H_1, H_2, \ldots\}$, then M conservatively learns $\{L_0, L_1, L_2, \ldots\}$ with respect to $\{H_0, H_1, H_2, \ldots\}$. Assume M is not conservative, then there exist τ, σ such that $H_{M(\tau)} \neq H_{M(\tau\sigma)}$, but $\text{content}(\tau\sigma) \subseteq H_{M(\tau)}$. Since $\{H_0, H_1, H_2, \ldots\}$ is class-preserving, there exists an n such that $L_n = H_{M(\tau)}$. Let T be a text for L_n, then $\tau\sigma T$ is a text for L_n. However, M is not non U-shaped on $\tau\sigma T$, as it first outputs a correct hypothesis $H_{M(\tau)} = L_n$ and then abandons it. The same argument applies for an exact hypothesis space.

However, for class-comprising learning, non U-shaped learning is more powerful than conservative learning, as shown by the following theorem.

Theorem 13. *Assume that $\{L_0, L_1, L_2, \ldots\}$ contains all sets $\{x, x+1, x+2, \ldots\}$ and all finite sets D such that there is an s with $\min(D) \in \mathbb{K}_{s+1} - \mathbb{K}_s$ and $0 < |D| < s$. Then $\{L_0, L_1, L_2, \ldots\}$ has a non U-shaped class-comprising learner but not a conservative class-comprising learner.*

Proof. Let the hypotheses space $\{H_0, H_1, H_2, \ldots\}$ be such that $H_{3x} = \{x + t : x \notin \mathbb{K}_t, t \in \mathbb{N}\}$, $H_{3x+1} = \{x, x+1, x+2, \ldots\}$ and $H_{3x+2} = D_x$, where $D_0, D_1, D_2, \ldots$ is the canonical numbering of all finite sets. Then a non U-shaped learner M for $\{L_0, L_1, L_2, \ldots\}$ with respect to $\{H_0, H_1, H_2, \ldots\}$ is defined as follows. Given any text T, at time t, find the smallest element $x \in \text{content}(T[t])$ and the largest s such that $s \leq t \wedge x \notin \mathbb{K}_s$; without loss of generality it is assumed that $\mathbb{K}_0 = \emptyset$ and thus s always exists. If $x \in \mathbb{K}_{s+1}$ and $|\text{content}(T[t])| < s$, then output the hypothesis H_{3y+2} with $H_{3y+2} = \text{content}(T[t])$; if $x \in \mathbb{K}_{s+1}$ and $|\text{content}(T[t])| \geq s$, then output H_{3x+1}; if $x \notin \mathbb{K}_{s+1}$, then output H_{3x}. Note that $H_{3x} \notin \{L_0, L_1, L_2, \ldots\}$ whenever $x \in \mathbb{K}$; hence it can be verified easily that the above learner is non U-shaped.

Suppose that $\{L_0, L_1, L_2, \ldots\}$ has a conservative class-comprising learner M. Then for any $x \in \mathbb{N}$, on the canonical text T for $\{x, x+1, x+2, \ldots\}$, there exists some t such that $H_{M(T[t])} \supset \text{content}(T[t])$. If $x \in \mathbb{K}$, then x must be in $\mathbb{K}_{|\text{content}(T[t])|+1}$, otherwise we can extend $T[t]$ to be a text for $\text{content}(T[t])$, which is a language in $\{L_0, L_1, L_2, \ldots\}$, thus violating the conservativeness of M. Clearly $x \in \mathbb{K}_{|\text{content}(T[t])|+1}$ implies $x \in \mathbb{K}$. Thus, we have $x \in \mathbb{K}$ iff $x \in \mathbb{K}_{|\text{content}(T[t])|+1}$. Hence, we have an effective procedure to decide whether $x \in \mathbb{K}$, a contradiction.

The following theorem gives a sufficient condition for uniform non U-shaped learnability. Furthermore, this condition helps us to separate uniform non U-shaped learnability from prescribed conservative learnability.

Theorem 14. *If the class $\{L_0, L_1, L_2, \ldots\}$ is exactly finitely learnable then $\{L_0, L_1, L_2, \ldots\}$ is uniformly non U-shaped learnable. In particular, there are classes which are uniformly non U-shaped learnable but not prescribed conservatively learnable.*

Proof. Let M be an exact finite learner for $\{L_0, L_1, L_2, \ldots\}$, we define a recursive enumeration of non U-shaped learners $M_0, M_1, M_2, \ldots$ which uniformly learn $\{L_0, L_1, L_2, \ldots\}$. For each $i \in \mathbb{N}$, $M_i(T[t])$ is defined as follows: if $M(T[t]) = ?$, then output ?; if $M(T[t]) = e$, then for each $j \leq t$, define $r_j = \min\{x : x > t \text{ or } L_e(x) \neq \varphi_i(j, x)\}$. Output the minimal j which maximizes r_j. It can be easily verified that the above learners witness that $\{L_0, L_1, L_2, \ldots\}$ is uniformly non U-shaped learnable.

Take any exactly finitely learnable language collection with infinitely many coinfinite languages, then it is uniformly non U-shaped learnable but not prescribed conservatively learnable. An example is the language collection $\{L_0, L_1, L_2, \ldots\}$ where $L_n = \{0, 2, 4, 6, \ldots\} \cup \{2n + 1\}$.

With the following result we can see that non U-shaped learning and decisive learning are equivalent for prescribed learning and uniform learning.

Theorem 15. *If $\{L_0, L_1, L_2, \ldots\}$ is prescribed non U-shaped learnable then $\{L_0, L_1, L_2, \ldots\}$ is also prescribed decisively learnable. If $\{L_0, L_1, L_2, \ldots\}$ is uniformly non U-shaped learnable then $\{L_0, L_1, L_2, \ldots\}$ is also uniformly decisively learnable.*

Proof. It suffices to show that if M non U-shaped learns $\{L_0, L_1, L_2, \ldots\}$ with respect to a given hypothesis space $\{H_0, H_1, H_2, \ldots\}$, then we can effectively build another learner M' which decisively learns $\{L_0, L_1, L_2, \ldots\}$ with respect to $\{H_0, H_1, H_2, \ldots\}$. The desired M' can be defined as follows. Given a text T, let $M'(T[0]) = M(T[0])$. For $t > 0$, $M'(T[t]) = M(T[t])$ if for all $t' < t$, for some $x \leq t$, $H_{M(T[t])}(x) \neq H_{M'(T[t'])}(x)$; and $M'(T[t]) = M'(T[t-1])$ otherwise. It can be easily verified that M' is decisively learning $\{L_0, L_1, L_2, \ldots\}$ using $\{H_0, H_1, H_2, \ldots\}$.

As in the case of conservative learning, class-preserving-uniform non U-shaped learnability and prescribed non U-shaped learnability are not comparable as well.

Theorem 16. (a) *There exists an $\{L_0, L_1, L_2, \ldots\}$ which is class-preserving-uniformly non U-shaped learnable but not prescribed non U-shaped learnable.*
(b) *There exists an $\{L_0, L_1, L_2, \ldots\}$ which is prescribed non U-shaped learnable but not class-preserving-uniformly non U-shaped learnable.*

Proof. (a) Let $\{L_0, L_1, L_2, \ldots\}$ be a recursive enumeration of all sets $\{2x\}$ with $x \notin \mathbb{K}$ and $\{2x, 2y + 1\}$ with $x \in \mathbb{K} \wedge y \in \mathbb{N}$. It is easy to see that such an enumeration exists.

To see that $\{L_0, L_1, L_2, \ldots\}$ is class-preserving-uniformly non U-shaped learnable consider the following recursive sequence $M_0, M_1, \ldots$ of learners.

M_e on input $T[t]$ does the following. If content$(T[t])$ is empty, then output ?. Otherwise output the least $j \leq t$ such that content$(T[t]) \subseteq \{x : \varphi_e(j, x) = 1\}$, if any. Otherwise repeat the previous hypothesis. It is easy to verify that M_e conservatively learns $\{L_0, L_1, L_2, \ldots\}$ using hypothesis space provided by φ_e, if this φ_e is recursive and defines a class-preserving hypothesis space for $\{L_0, L_1, L_2, \ldots\}$.

We now show that $\{L_0, L_1, L_2, \ldots\}$ is not prescribed non U-shaped learnable. Let $c_{\mathbb{K}}$ be the convergence modulus of the halting problem $\mathbb{K}$, that is, $c_{\mathbb{K}}(i) = \min\{t : \forall j \leq i\, [\mathbb{K}_t(j) = \mathbb{K}(j)]\}$.

Furthermore, $c_{\mathbb{K},k}(i) = \min\{t : \forall j \leq i\, [\mathbb{K}_t(j) = \mathbb{K}_k(j)]\}$ is the k-th approximation to $c_{\mathbb{K}}$; clearly $c_{\mathbb{K},k}(i) \leq k$ for all i. Now consider the following superclass of $\{L_0, L_1, L_2, \ldots\}$:

$$H_{\langle i,j,k\rangle} = \begin{cases} \{2i, 2j+1\} & \text{if } k = 0; \\ \{2i\} & \text{if } k > 0 \wedge c_{\mathbb{K},k}(i) = c_{\mathbb{K}}(i); \\ \{2i, 2t+1\} & \text{otherwise, where } t = \min\{s : c_{\mathbb{K},s}(i) > c_{\mathbb{K},k}(i)\}. \end{cases}$$

$\{H_0, H_1, H_2, \ldots\}$ is an indexed family and it contains $\{L_0, L_1, L_2, \ldots\}$. Now assume that N is a recursive learner for $\{L_0, L_1, L_2, \ldots\}$ with hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

Let f be the partial-recursive function such that $f(i)$ is the first s such that $N((2i)^s)$ is an index for either $\{2i\}$ or some set $\{2i, 2j+1\}$. Note that $f(i)$ is defined for all $i \notin \mathbb{K}$. Due to the fast growth-rate of $c_{\mathbb{K}}$, for almost all the i in the domain of f, it holds that $N((2i)^{f(i)})$ is an index for some set $\{2i, 2j_i+1\}$ with j_i depending on $i, f(i)$. As N learns $\{2i\}$ on the text $(2i)^\infty$ for all $i \notin \mathbb{K}$, there exists a further partial-recursive function g with the following properties: g is defined on almost all elements in $\mathbb{N}-\mathbb{K}$; for all i in the domain of g, $g(i) > f(i)$ and $N((2i)^{g(i)})$ is an index for a set containing $2i$ but not $2j_i+1$. As $\mathbb{N}-\mathbb{K}$ is not recursively enumerable, $f(i), g(i)$ are defined for infinitely many $i \in \mathbb{K}$ as well. So there is some $i \in \mathbb{K}$ with $N((2i)^{f(i)})$ being an index for $\{2i, 2j_i+1\}$ and $N((2i)^{g(i)})$ being an index of some other set. It follows that N is not U-shaped on the text $(2i)^{g(i)}\,(2j_i+1)^\infty$.

(b) The class $\{L : |L| \leq 1\}$ is easily seen to be prescribed non U-shaped learnable. Let $\{H_0, H_1, H_2, \ldots\}$ contain the class to be learnt. If $\text{content}(T[t]) = \emptyset$ then the learner outputs the least index for $\emptyset$. Otherwise, let x be the least member of $\text{content}(T[t])$; the learner outputs the least number e with $\{0, 1, 2, \ldots, x+t\} \cap H_e = \{x\}$.

Now it is shown that the above class is not class-preserving-uniformly non U-shaped learnable. The reason is that one cannot figure out the index of the empty set in a given indexing; indeed one can make an indexing $\mathcal{H}_e$ of $\{L : |L| \leq 1\}$ for which the index of $\emptyset$ is larger than the convergence modulus $c_{\mathbb{K}}(e)$. Let M_e be a uniformly obtained learner for $\{L : |L| \leq 1\}$ using hypothesis space $\mathcal{H}_e$; such a learner exists, but it will be shown that for some e the learner M_e cannot be non U-shaped. Let $f(e)$ be the index of the first hypothesis output by M_e on $\#^\infty$. As $f(e) < c_{\mathbb{K}}(e)$ for infinitely many e, there is an e such that the first index output on $\#^\infty$ by M_e is for some set $\{x\}$. As M_e learns $\#^\infty$ there is some later index output for $\emptyset$ after having seen $\#^s$ for some s. It follows that M_e is U-shaped on $\#^s x^\infty$.

5 Monotonic Learning

The prescribed and uniform versions of strong-monotonic and monotonic learning are very restrictive.

Theorem 17. (a) $\{L_0, L_1, L_2, \ldots\}$ *is prescribed strong-monotonically learnable iff* $\{L_0, L_1, L_2, \ldots\}$ *is finite.*
(b) $\{L_0, L_1, L_2, \ldots\}$ *cannot be uniformly strong-monotonically learnable.*

Proof. (a) If $\{L_0, L_1, L_2, \ldots\}$ is finite, then it is easily seen to be prescribed strong-monotonically learnable.

Now assume that $\{L_0, L_1, L_2, \ldots\}$ is infinite. Let $\texttt{odd}(x) = 1$ for odd x and $\texttt{odd}(x) = 0$ for even x. Furthermore $\texttt{even}(x) = 1 - \texttt{odd}(x)$. Let $\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \ldots$ be a fixed enumeration of all learners. Suppose $\{L_0, L_1, L_2, \ldots\}$ is infinite. We define an indexed family hypothesis space $\{H_0, H_1, H_2, \ldots\}$ such that $\{L_0, L_1, L_2, \ldots\}$ is not strong-monotonically learnable with respect to $\{H_0, H_1, H_2, \ldots\}$. Let F be a recursive set such that F differs from each set in $\{L_0, L_1, L_2, \ldots\}$ on infinitely many even and infinitely many odd places. Let T_i denote a standard text for L_i, obtained effectively from i.

$$H_{\langle i,j\rangle}(x) = \begin{cases} \max\{\texttt{even}(x), F(x)\}, & \text{if } \langle i,j\rangle \text{ is the first index, with the} \\ & \text{first component being } i, \text{ output by } \mathcal{M}_i \\ & \text{on } T_i \text{ within } x \text{ steps;} \\ \min\{\texttt{odd}(x), F(x)\}, & \text{if } \langle i,j\rangle \text{ is the second distinct index, with the} \\ & \text{first component being } i, \text{ output by } \mathcal{M}_i \\ & \text{on } T_i \text{ within } x \text{ steps;} \\ L_i(x), & \text{otherwise.} \end{cases}$$

$\{H_0, H_1, H_2, \ldots\}$ is an indexed family hypothesis space for $\{L_0, L_1, L_2, \ldots\}$. For $i \in \mathbb{N}$, consider the behaviour of $\mathcal{M}_i$ on the canonical text T_i for L_i:

1. If $\mathcal{M}_i$ does not output an index of the form $\langle i,j\rangle$, then M_i fails to learn L_i because from the definition of $\{H_0, H_1, H_2, \ldots\}$, only indices of such form can be indices for L_i.
2. If $\mathcal{M}_i$ outputs only one such index, then from the definition of $\{H_0, H_1, H_2, \ldots\}$, the index is not for any $L \in \{L_0, L_1, L_2, \ldots\}$, thus not for L_i.
3. If $\mathcal{M}_i$ outputs two different such indices, say $\langle i, j_1\rangle$ and $\langle i, j_2\rangle$ being the first and second one respectively, then from the definition of $\{H_0, H_1, H_2, \ldots\}$, $H_{\langle i,j_1\rangle} \not\subseteq H_{\langle i,j_2\rangle}$, because $H_{\langle i,j_1\rangle}$ contains all even numbers larger than x while $H_{\langle i,j_2\rangle}$ does not, where x is the number of steps needed for $\mathcal{M}_i$ to output $\langle i, j_2\rangle$.

Hence, $\mathcal{M}_i$ fails to learn L_i strong-monotonically from T_i. Thus, no learner learns $\{L_0, L_1, L_2, \ldots\}$ strong-monotonically with respect to $\{H_0, H_1, H_2, \ldots\}$, a contradiction. Hence, $\{L_0, L_1, L_2, \ldots\}$ must be finite.

(b) Let $G_e = L_e$ if $e < |\{L_0, L_1, L_2, \ldots\}|$ and let G_e be some recursive set outside $\{L_0, L_1, L_2, \ldots\}$ otherwise. To see that $\{L_0, L_1, L_2, \ldots\}$ is not uniformly strong-monotonically learnable, suppose by way of contradiction that there exists a recursive enumeration of learners $M_0, M_1, \ldots$ such that whenever φ_i defines a hypothesis space $\{H_0, H_1, H_2, \ldots\}$ which contains $\{L_0, L_1, L_2, \ldots\}$, then M_i learns $\{L_0, L_1, L_2, \ldots\}$ strong-monotonically with respect to $\{H_0, H_1, H_2, \ldots\}$. Let F be a recursive set such that F differs from each set in $\{L_0, L_1, L_2, \ldots\}$ on infinitely many even and infinitely many odd places. Let T_i denote a standard text for L_i, obtained effectively from i. By

Kleene's recursion theorem [16], there exists an e such that:

$$\varphi_e(\langle i,j\rangle,x) = \begin{cases} \max\{\texttt{even}(x),F(x)\}, & \text{if } \langle i,j\rangle \text{ is the first index, with the} \\ & \text{first component being } i\text{, output by } M_e \\ & \text{on } T_i \text{ within } x \text{ steps;} \\ \min\{\texttt{odd}(x),F(x)\}, & \text{if } \langle i,j\rangle \text{ is the second distinct index, with the} \\ & \text{first component being } i\text{, output by } M_e \\ & \text{on } T_i \text{ within } x \text{ steps;} \\ G_i(x), & \text{otherwise.} \end{cases}$$

It can be verified that M_e does not strong-monotonically learn $\{L_0, L_1, L_2, \ldots\}$ with respect to the hypothesis space defined by φ_e in a way similar to part (a).

As in the case of uniform conservative learning, there also exists an infinite class which is uniformly monotonically learnable.

Example 18. *Let $L_a = \{a\}$. Then $\{L_0, L_1, L_2, \ldots\}$ is an infinite class which is uniformly monotonically learnable.*

Proof. For each $e \in \mathbb{N}$, define $M_e(T[t])$ as follows. If $\text{content}(T[t]) = \emptyset$ then output ?. Otherwise, let $a \in \text{content}(T[t])$. Let $j \le t$ be minimal (if any) such that $\varphi_e(j,a) = 1$ and for all $n \le t$ with $n \ne a$, $\varphi_e(j,n) = 0$. If such j is found, output j; otherwise output ?. It is easy to verify that $M_0, M_1, M_2, \ldots$ uniformly monotonically learn $\{L_0, L_1, L_2, \ldots\}$.

The following result shows that in fact it is necessary for a class to contain only finite sets in order to be uniformly monotonically learnable.

Theorem 19. *If $\{L_0, L_1, L_2, \ldots\}$ is uniformly monotonically learnable, then $\{L_0, L_1, L_2, \ldots\}$ contains only finite sets.*

Proof. Let I_n be as in Definition 5. Let $G_e = L_e$ if $e < |\{L_0, L_1, L_2, \ldots\}|$ and let G_e be some recursive set outside $\{L_0, L_1, L_2, \ldots\}$ otherwise. Let $X = \{x : \exists n \exists i \le n\ [x = \min(L_i \cap I_n)]\} \cup \{x : \exists n \exists i \le n\ [x = \min(\overline{L_i} \cap I_n)]\}$. Then X is recursive and for any finite variant Y of X and all d, $Y \ne L_d$ and $Y \cap L_d$ is infinite whenever L_d is infinite.

If $\{L_0, L_1, L_2, \ldots\}$ is uniformly monotonically learnable, then there exists a recursive enumeration of learners $M_0, M_1, M_2, \ldots$ such that whenever φ_i defines a hypothesis space $\{H_0, H_1, H_2, \ldots\}$ which contains $\{L_0, L_1, L_2, \ldots\}$, then M_i learns $\{L_0, L_1, L_2, \ldots\}$ monotonically with respect to $\{H_0, H_1, H_2, \ldots\}$. Let T_i denote a standard text for L_i, obtained effectively from i. By Kleene's recursion theorem [16], there exists an e such that:

$$\varphi_e(\langle i,j\rangle,x) = \begin{cases} X(x), & \text{if } \langle i,j\rangle \text{ is the first index, with the} \\ & \text{first component being } i\text{, output by } M_e \\ & \text{on } T_i \text{ within } x \text{ steps;} \\ 0, & \text{if } \langle i,j\rangle \text{ is the second distinct index, with the} \\ & \text{first component being } i\text{, output by } M_e \\ & \text{on } T_i \text{ within } x \text{ steps;} \\ G_i(x), & \text{otherwise.} \end{cases}$$

Let $\{H_0, H_1, H_2, \ldots\}$ be the hypotheses defined by φ_e. Then clearly $\{H_0, H_1, H_2, \ldots\}$ is an indexed family and $\{H_0, H_1, H_2, \ldots\} \supseteq \{L_0, L_1, L_2, \ldots\}$. We show that if there exists any infinite L_n with $n < |\{L_0, L_1, L_2, \ldots\}|$, then M_e does not monotonically learn L_n with respect to $\{H_0, H_1, H_2, \ldots\}$, thus M_e fails to learn $\{L_0, L_1, L_2, \ldots\}$ monotonically with respect to $\{H_0, H_1, H_2, \ldots\}$.

1. If M_e does not output any index of the form $\langle n, j\rangle$ on T_n, then M_e fails to learn L_n, as only indices of the form $\langle n, j\rangle$ can be indices for L_n.
2. If M_e only output one distinct index of the form $\langle n, j\rangle$, say, $\langle n, j_0\rangle$, then by the definition of φ_e, $H_{\langle n,j_0\rangle} \neq L_n$, as $H_{\langle n,j_0\rangle}$ is a finite variant of X.
3. If M_e outputs at least two distinct indices of the form $\langle n, j\rangle$, then let $\langle n, j_0\rangle$ and $\langle n, j_1\rangle$ be the first and second such distinct indices respectively. By the definition of φ_e, $H_{\langle n,j_0\rangle} \cap L_n$ is infinite while $H_{\langle n,j_1\rangle} \cap L_n$ is finite (as $H_{\langle n,j_1\rangle}$ is finite). Hence M_e is not monotonic.

This completes the proof.

For prescribed monotonic learning, finitely many sets in the language class can violate the above necessary condition for uniform monotonic learning.

Theorem 20. *If $\{L_0, L_1, L_2, \ldots\}$ is prescribed monotonically learnable, then $\{L_0, L_1, L_2, \ldots\}$ contains only finitely many infinite sets.*

Proof. Let $A = \{\langle i, j\rangle : \exists e \leq i[\varphi_e(i, 0) \downarrow= \langle i, j\rangle]\}$, $B = \{\langle i, j\rangle : \exists e \leq i \exists k \leq i + 2[\varphi_e(i, k) \downarrow= \langle i, j\rangle]\}$. Then A and B are both r.e. sets. Now suppose $\{L_0, L_1, L_2, \ldots\}$ is infinite and uniformly recursive. Let X be as in the proof for Theorem 19. Let A_x and B_x be the sets of elements enumerated into A and B in x steps respectively. Let $\{H_0, H_1, H_2, \ldots\}$ be as follows:

$$H_{\langle i,j\rangle}(x) = \begin{cases} X(x), & \text{if } \langle i, j\rangle \in A_x; \\ 0, & \text{if } \langle i, j\rangle \in B_x - A_x; \\ L_i(x), & \text{otherwise.} \end{cases}$$

Then $\{H_0, H_1, H_2, \ldots\}$ is a uniformly recursive hypothesis space containing $\{L_0, L_1, L_2, \ldots\}$. Since $\{L_0, L_1, L_2, \ldots\}$ is prescribed monotonically learnable, there exists a learner M which monotonically learns $\{L_0, L_1, L_2, \ldots\}$ with respect to $\{H_0, H_1, H_2, \ldots\}$. Define $f(i, j)$ =the j^{th} (starting from 0) distinct index of the form $\langle i, k\rangle(k \in \mathbb{N})$ output by M on the canonical text T_i for L_i. Then $f(i, j)$ is partial recursive, and thus there exists n such that $\varphi_n(i, j) = f(i, j)$.

Suppose to the contrary that $\{L_0, L_1, L_2, \ldots\}$ contains infinitely many infinite languages, then there exists $m > n$ such that L_m is infinite. By the definition of $\{H_0, H_1, H_2, \ldots\}$, the indices for L_m can only be of the form $\langle m, k\rangle$ for some $k \in \mathbb{N}$, thus $\langle m, j_0\rangle = \varphi_n(m, 0)$ is defined and $\langle m, j_0\rangle \in A$. By the definition of $\{H_0, H_1, H_2, \ldots\}$, $H_{\langle m,j_0\rangle} \neq L_m$, and $H_{\langle m,j_0\rangle} \cap L_n$ is an infinite set. This implies $\langle m, j_1\rangle = \varphi_n(m, 1)$ is defined and $\langle m, j_1\rangle \in B$. If $\langle m, j_1\rangle \notin A$, then by the definition of $\{H_0, H_1, H_2, \ldots\}$, $H_{\langle m,j_1\rangle} \cap L_m$ is finite. Thus $H_{\langle m,j_0\rangle} \cap L_m \nsubseteq H_{\langle m,j_1\rangle} \cap L_m$ and we are done. Otherwise, we can see that as long as $\langle m, j_k\rangle = \varphi_n(m, k) \in A$, then $\langle m, j_{k+1}\rangle = \varphi_n(m, k + 1)$ is defined. However, A contains at most $m + 1$ indices of the form $\langle m, j\rangle$, thus there exists

$k \leq m$ such that $\langle m, j_{k+1}\rangle \in B - A$. At this point, using same argument as above, we can see monotonicity of M is violated, a contradiction. Hence $\{L_0, L_1, L_2, \ldots\}$ contains only finitely many infinite languages.

The following example shows that the condition is a necessary, but not sufficient condition for prescribed monotonic learnability.

Example 21. *Let $\{L_0, L_1, L_2, \ldots\}$ contain all sets with one or two elements plus perhaps other non-empty sets. Then $\{L_0, L_1, L_2, \ldots\}$ is not prescribed monotonically learnable.*

Proof. We define an indexed family hypothesis space $\{H_0, H_1, H_2, \ldots\}$ as follows:

$$H_{\langle i,j\rangle} = \begin{cases} L_j, & \text{if } j < i \text{ and } |L_j \cap \{0,1,\ldots,i\}| \geq 2; \\ \{j, i\}, & \text{if } j < i \text{ and } |L_j \cap \{0,1,\ldots,i\}| < 2; \\ \{i\}, & \text{if } j \geq i \text{ and } j \notin S; \\ \{i, j+t+1\}, & \text{if } j \geq i \text{ and } j \in S_{t+1} - S_t. \end{cases}$$

The class $\{L_0, L_1, L_2, \ldots\}$ is not monotonically learnable with respect to $\{H_0, H_1, H_2, \ldots\}$ as shown below. Suppose M learns $\{L_0, L_1, L_2, \ldots\}$ monotonically with respect to $\{H_0, H_1, H_2, \ldots\}$, then for all $i \in \mathbb{N}$, on the text i^∞, M must output an index of the form $\langle i, j\rangle$ with $j \geq i$. Let $\langle i, n_i\rangle$ be the first such index. If for all $i \in \mathbb{N}$, $H_{\langle i,n_i\rangle} = \{i\}$, then $n_i \notin S$, and thus $\{n_0, n_1, \ldots\}$ is an infinite r.e. subset of $\overline{S}$, a contradiction. Hence for some $i \in \mathbb{N}$, $n_i \in S$ and $H_{\langle i,n_i\rangle} = \{i, t+1\}$ for some $t \geq i$. Suppose, $M(i^{t_1}) = \langle i, n_i\rangle$ and for some $t_2 > t_1$, $M(i^{t_2}) = \langle i, j\rangle$, where $j > i$ and $j \notin S$, then we can extend i^{t_2} to a text $i^{t_2}(t+1)^\infty$ for $\{i, t+1\}$. On this text M does not monotonically learn $\{i, t+1\}$ because for t_1, t_2 with $t_1 < t_2$, $H_{M(i^{t_1})} \cap \{i, t+1\} = \{i, t+1\} \not\subseteq H_{M(i^{t_2})} \cap \{i, t+1\} = \{i\}$.

Theorem 22. (a) *There exists a class $\{L_0, L_1, L_2, \ldots\}$ which is class-preserving-uniformly strong-monotonically learnable but not prescribed monotonically learnable.*
(b) *There exists a class $\{L_0, L_1, L_2, \ldots\}$ which is prescribed monotonically learnable but not class-preserving-uniformly monotonically learnable.*
(c) *Every prescribed strong-monotonically learnable class is also class-preserving-uniformly strong-monotonically learnable.*

Proof. (a) Consider the class $\{L_0, L_1, L_2, \ldots\}$ with $L_i = \{\langle i, j\rangle : j \in \mathbb{N}\}$. This is easily seen to be class-preserving-uniformly finitely (and thus strong-monotonically) learnable. However as $\{L_0, L_1, L_2, \ldots\}$ contains infinitely many infinite-coinfinite languages, by Theorem 20, it is not prescribed monotonically learnable.

(b) Consider the class consisting of the empty set and all singleton sets $\{x\}$. It is easily seen to be prescribed monotonically learnable. Using an argument similar to the proof of Theorem 16 (b), we can see that $\{L_0, L_1, L_2, \ldots\}$ is not class-preserving-uniformly monotonically learnable.

(c) Any prescribed strong-monotonically learnable class is finite, and a finite class is easily seen to be class-preserving-uniformly strong-monotonically learnable.

Remark 23. A class is *dual strong-monotonically* learnable iff there is a learner such that every subsequent hypothesis is for a subset of the previous one. A class is *dual monotonically* learnable iff there is a learner such that for every set L in the class, if the learner outputs H_j after H_i on input text for L, then it holds that $H_i \cup L \supseteq H_j \cup L$.

One can obtain results similar to those for strong-monotonic and monotonic learning. However, there is one difference: Dual strong-monotonically learnable classes have to be inclusion-free, that is, there are no sets L_i, L_j in the class with $L_i \subset L_j$. Hence there is a close connection between finite learning and dual strong-monotonic learning: A class is exactly dual strong-monotonically learnable iff it is exactly finitely learnable [14]; a class is class-preservingly dual strong-monotonically learnable iff it is class-preservingly finitely learnable [14]; a class is prescribed dual strong-monotonically learnable iff it is finite and inclusion-free; no class is uniformly dual strong-monotonically learnable. But there is a difference for class-comprising learning as there is a class which is class-comprisingly dual strong-monotonically learnable but not finitely learnable [14]. The class of all sets $\{x, x+1, x+2, \ldots\}$ is uniformly dual monotonically learnable. Any class which is uniformly dual monotonically learnable contains only cofinite sets. If a class is prescribed dual monotonically learnable then it contains only finitely many coinfinite sets. Every prescribed dual strong-monotonically learnable class is also class-preserving-uniformly dual strong-monotonically learnable. There is a class which is class-preserving-uniformly dual strong-monotonically learnable but not prescribed dual monotonically learnable; this class consists of all singleton sets $\{x\}$. There is a class which is prescribed dual monotonically learnable but not class-preservingly dual monotonically learnable; this class consists of the empty set, the set of even numbers and all sets of the form $\{0, 2, 4, \ldots, 2x\} \cup \{2x+2, 2x+3, 2x+4, \ldots\}$.

6 Acknowledgements

We thank the anonymous referees for several helpful comments.

References

1. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
2. Lenore Blum and Manuel Blum. Towards a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.
3. Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan and Rolf Wiehagen. When unlearning helps. *Information and Computation*, to appear.
4. E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
5. Klaus-Peter Jantke. Monotonic and Non-monotonic Inductive Inference. *New Generation Computing*, 8:349–360, 1991.

6. Sanjay Jain, Frank Stephan and Nan Ye. Prescribed Learning of R.E. Classes. In M. Hutter, R. Servedio and E. Takimoto, editors, *Algorithmic Learning Theory, 18th International Conference*, ALT 2007, *Springer Lecture Notes in Artificial Intelligence* 4754:64–78, 2007.
7. Steffen Lange. *Algorithmic Learning of Recursive Languages.* Habilitationsschrift, der Fakultät für Mathematik und Informatik der Universität Leipzig eingereichte, Mensch and Buch Verlag, Berlin, 2000.
8. Steffen Lange and Thomas Zeugmann. Language learning in dependence on the space of hypotheses. *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, Santa Cruz, California, United States, pages 127–136, 1993.
9. Steffen Lange, Thomas Zeugmann and Shyam Kapur. Monotonic and dual monotonic language learning. *Theoretical Computer Science*, 155:365–410, 1996.
10. 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.
11. Emil Post. Recursively enumerable sets of positive integers and their decision problems, *Bulletin of the American Mathematical Society*, 50:284–316, 1944.
12. Ludwig Staiger. On the power of reading the whole infinite input tape. *Grammars*, 3:247–257, 1999.
13. Thomas Zeugmann. *Algorithmisches Lernen von Funktionen und Sprachen.* Habilitationsschrift, Technische Hochschule Darmstadt, 1993.
14. Thomas Zeugmann and Steffen Lange. A guided tour across the boundaries of learning recursive languages. *Algorithmic Learning for Knowledge-Based Systems*, final report on research project Gosler, edited by Klaus P. Jantke and Steffen Lange, *Springer Lecture Notes in Artificial Intelligence* 961:193–262, 1995.
15. Thomas Zeugmann, Steffen Lange and Shyam Kapur. Characterizations of monotonic and dual monotonic language learning, *Information and Computation*, 120:155–173, 1995.
16. Hartley Rogers. *Theory of Recursive Functions and Effective Computability*, McGraw-Hill, 1967. Reprinted by MIT Press in 1987.
17. Sandra Zilles. Separation of uniform learning classes. *Theoretical Computer Science*, 313:229–265, 2004.
18. Sandra Zilles. Increasing the power of uniform inductive learners. *Journal of Computer and System Sciences*, 70:510–538, 2005.