Learning Languages with Decidable Hypotheses*

Julian Berger[1], Maximilian Böther[1], Vanja Doskoč[2], Jonathan Gadea Harder[1], Nicolas Klodt[1], Timo Kötzing[2], Winfried Lötzsch[1], Jannik Peters[1], Leon Schiller[1], Lars Seifert[1], Armin Wells[1], and Simon Wietheger[1]

[1] Hasso Plattner Institute, University of Potsdam, Germany
firstname.lastname@student.hpi.uni-potsdam.de
[2] Hasso Plattner Institute, University of Potsdam, Germany
firstname.lastname@hpi.de

Abstract. In *language learning in the limit*, the most common type of hypothesis is to give an enumerator for a language, a W-index. These hypotheses have the drawback that even the membership problem is undecidable. In this paper, we use a different system which allows for naming arbitrary decidable languages, namely *programs for characteristic functions* (called C-indices). These indices have the drawback that it is now not decidable whether a given hypothesis is even a legal C-index.
In this first analysis of learning with C-indices, we give a structured account of the learning power of various restrictions employing C-indices, also when compared with W-indices. We establish a hierarchy of learning power depending on whether C-indices are required (a) on all outputs; (b) only on outputs relevant for the class to be learned or (c) only in the limit as final, correct hypotheses. We analyze all these questions also in relation to the mode of data presentation.
Finally, we also ask about the relation of semantic versus syntactic convergence and derive the map of pairwise relations for these two kinds of convergence coupled with various forms of data presentation.

1 Introduction

We are interested in the problem of algorithmically learning a description for a formal language (a computably enumerable subset of the set of natural numbers) when presented successively all and only the elements of that language; this is called *inductive inference*, a branch of (algorithmic) learning theory. For example, a learner h might be presented more and more even numbers. After each new number, h outputs a description for a language as its conjecture. The learner h might decide to output a program for the set of all multiples of 4, as long as all numbers presented are divisible by 4. Later, when h sees an even number not divisible by 4, it might change this guess to a program for the set of all multiples of 2.

Many criteria for determining whether a learner h is *successful* on a language L have been proposed in the literature. Gold, in his seminal paper [10], gave a first, simple learning criterion, $\mathbf{TxtGEx}$-*learning*[3], where a learner is *successful* if and only if, on

* This work was supported by DFG Grant Number KO 4635/1-1.

[3] $\mathbf{Txt}$ stands for learning from a *text* of positive examples; $\mathbf{G}$ for Gold, indicating full-information learning; $\mathbf{Ex}$ stands for *explanatory*.

every *text* for L (listing of all and only the elements of L) it eventually stops changing its conjectures, and its final conjecture is a correct description for the input language.

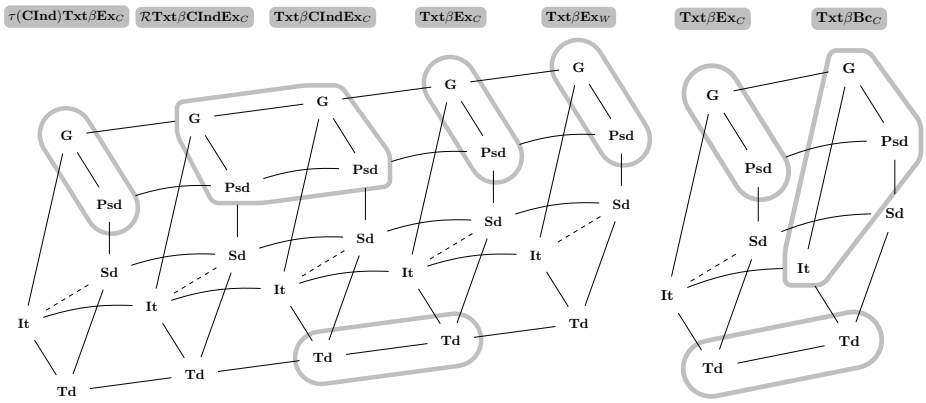
Trivially, each single, describable language L has a suitable constant function as a $\mathbf{TxtGEx}$-learner (this learner constantly outputs a description for L). Thus, we are interested in analyzing for which *classes of languages* $\mathcal{L}$ is there a *single learner* h learning *each* member of $\mathcal{L}$. This framework is also known as *language learning in the limit* and has been studied extensively, using a wide range of learning criteria similar to $\mathbf{TxtGEx}$-learning (see, for example, the textbook [11]).

In this paper, we put the focus on the possible descriptions for languages. Any computably enumerable language L has as possible descriptions any program enumerating all and only the elements of L, called a W-index (the language enumerated by program e is denoted by W_e). This system has various drawbacks; most importantly, the function which decides, given e and x, whether $x \in W_e$ is not computable. We propose to use different descriptors for languages: programs for characteristic functions (where such programs e describe the language C_e which it decides). Of course, only decidable languages have such a description, but now, given a program e for a characteristic function, $x \in C_e$ is decidable. Additionally to many questions that remain undecidable (for example, whether C-indices are for the same language or whether a C-index is for a finite language), it is not decidable whether a program e is indeed a program for a characteristic function. This leads to a new set of problems: Learners cannot be (algorithmically) checked whether their outputs are viable (in the sense of being programs for characteristic functions).

Based on this last observation, we study a range of different criteria which formalize what kind of behavior we expect from our learners. In the most relaxed setting, learners may output any number (for a program) they want, but in order to $\mathbf{Ex}$-learn, they need to converge to a correct C-index; we denote this restriction with $\mathbf{Ex}_C$. Requiring additionally to only use C-indices in order to successfully learn, we denote by $\mathbf{CIndEx}_C$; requiring C-indices on *all* inputs (not just for successful learning, but also when seeing input from no target language whatsoever) we denote by $\tau(\mathbf{CInd})\mathbf{Ex}_C$. In particular, the last restriction requires the learner to be total; in order to distinguish whether the loss of learning power is due to the totality restriction or truly due to the additional requirement of outputting C-indices, we also study $\mathcal{R}\mathbf{CIndEx}_C$, that is, the requirement $\mathbf{CIndEx}_C$ where additionally the learner is required to be total.

We note that $\tau(\mathbf{CInd})\mathbf{Ex}_C$ is similar to learning *indexable families*. Indexable families are classes of languages $\mathcal{L}$ such that there is an enumeration $(L_i)_{i\in\mathbb{N}}$ of all and only the elements of $\mathcal{L}$ for which the decision problem "$x \in L_i$" is decidable. Already for such classes of languages, we get a rich structure (see a survey of previous work [16]). For a learner h learning according to $\tau(\mathbf{CInd})\mathbf{Ex}_C$, we have that $L_x = C_{h(x)}$ gives an indexing of a family of languages, and h learns some subset thereof. We are specifically interested in the area between this setting and learning with W-indices ($\mathbf{Ex}_W$).

The criteria we analyze naturally interpolate between these two settings. We show that we have the following hierarchy: $\tau(\mathbf{CInd})\mathbf{Ex}_C$ allows for learning strictly fewer classes of languages than $\mathcal{R}\mathbf{CIndEx}_C$, which allow for learning the *same* classes as $\mathbf{CIndEx}_C$, which again are fewer than learnable by $\mathbf{Ex}_C$, which in turn renders fewer classes learnable than $\mathbf{Ex}_W$.



(a) (b)

Fig. 1: Relation of **(a)** various requirements when to output characteristic indices and **(b)** various learning criteria, both paired with various memory restrictions β. Black solid respectively dashed lines imply trivial respectively non-trivial inclusions (bottom-to-top, left-to-right). Furthermore, greyly edged areas illustrate a collapse of the enclosed learning criteria and there are no further collapses.

All these results hold for learning with full information. In order to study the dependence on the mode of information presentation, we also consider *partially set-driven* learners ($\mathbf{Psd}$, [2,19]), which only get the set of data presented so far and the iteration number as input; *set-driven* learners ($\mathbf{Sd}$, [20]), which get only the set of data presented so far; *iterative* learners ($\mathbf{It}$, [21,8]), which only get the new datum and their current hypothesis and, finally, *transductive* learners ($\mathbf{Td}$, [4,15]), which only get the current data. Note that transductive learners are mostly of interest as a proper restriction to all other modes of information presentation. In particular, we show that full-information learners can be turned into partially set-driven learners without loss of learning power and iterative learning is strictly less powerful than set-driven learning, in all settings.

Altogether we analyze 25 different criteria and show how each pair relates. All these results are summarized in Figure 1(a) as one big map stating all pairwise relations of the learning criteria mentioned, giving 300 pairwise relations in one diagram, proven with 13 theorems in Section 3. Note that the results comparing learning criteria with W-indices were previously known, and some proofs could be extended to also cover learning with C-indices. For the proofs, please consider the full version of this paper [1].

In Section 4, we derive a similar map considering a possible relaxation on $\mathbf{Ex}_C$-learning: While $\mathbf{Ex}_C$ requires syntactic convergence to one single correct C-index, we consider *behaviorally correct* learning ($\mathbf{Bc}_C$, [6,17]) where the learner only has to semantically converge to correct C-indices (but may use infinitely many different such indices). We again consider the different modes of data presentation and determine all pairwise relations in Figure 1(b). The proofs are again deferred to the full version [1].

2 Preliminaries

2.1 Mathematical Notations and Learning Criteria

In this section, we discuss the used notation as well as the system for learning criteria [15] we follow. Unintroduced notation follows the textbook [18].

With $\mathbb{N}$ we denote the set of all natural numbers, namely $\{0, 1, 2, \ldots\}$. We denote the subset and proper subset relation between two sets with $\subseteq$ and $\subsetneq$, respectively. We use $\emptyset$ and ε to denote the empty set and empty sequence, respectively. The set of all computable functions is denoted by $\mathcal{P}$, the subset of all total computable functions by $\mathcal{R}$. If a function f is (not) defined on some argument $x \in \mathbb{N}$, we say that f converges (diverges) on x, denoting this fact with $f(x)\downarrow$ ($f(x)\uparrow$). We fix an effective numbering $\{\varphi_e\}_{e\in\mathbb{N}}$ of $\mathcal{P}$. For any $e \in \mathbb{N}$, we let W_e denote the domain of φ_e and call e a W-index of W_e. This set we call the *e-th computably enumerable set*. We call $e \in \mathbb{N}$ a *C-index* (*characteristic index*) if and only if φ_e is a total function such that for all $x \in \mathbb{N}$ we have $\varphi_e(x) \in \{0, 1\}$. Furthermore, we let $C_e = \{x \in \mathbb{N} \mid \varphi_e(x) = 1\}$. For a computably enumerable set L, if some $e \in \mathbb{N}$ is a C-Index with $C_e = L$, we write $\varphi_e = \chi_L$. Note that, if a set has a C-index, it is *recursive*. The set of all recursive sets is denoted by **REC**. For a finite set $D \subseteq \mathbb{N}$, we let $\mathrm{ind}(D)$ be a C-index for D. Note that $\mathrm{ind} \in \mathcal{R}$. Furthermore, we fix a Blum complexity measure Φ associated with φ, that is, for all $e, x \in \mathbb{N}$, $\Phi_e(x)$ is the number of steps the function φ_e takes on input x to converge [3]. The padding function $\mathrm{pad} \in \mathcal{R}$ is an injective function such that, for all $e, n \in \mathbb{N}$, we have $\varphi_e = \varphi_{\mathrm{pad}(e,n)}$. We use $\langle\cdot,\cdot\rangle$ as a computable, bijective function that codes a pair of natural numbers into a single one. We use π_1 and π_2 as computable decoding functions for the first and section component, i.e., for all $x, y \in \mathbb{N}$ we have $\pi_1(\langle x, y\rangle) = x$ and $\pi_2(\langle x, y\rangle) = y$.

We learn computably enumerable sets L, called *languages*. We fix a *pause* symbol $\#$, and let, for any set S, $S_\# \coloneqq S \cup \{\#\}$. Information about languages is given from *text*, that is, total functions $T\colon \mathbb{N} \to \mathbb{N} \cup \{\#\}$. A text T is of a certain language L if its *content* is exactly L, that is, $\mathrm{content}(T) \coloneqq \mathrm{range}(T) \setminus \{\#\}$ is exactly L. We denote the set of all texts as $\mathbf{Txt}$ and the set of all texts of a language L as $\mathbf{Txt}(L)$. For any $n \in \mathbb{N}$, we denote with $T[n]$ the initial sequence of the text T of length n, that is, $T[0] \coloneqq \varepsilon$ and $T[n] \coloneqq (T(0), \ldots, T(n-1))$. Given a language L and $t \in \mathbb{N}$, the set of sequences consisting of elements of $L \cup \{\#\}$ that are at most t long is denoted by $L_\#^{\leq t}$. Furthermore, we denote with $\mathbb{S}\mathrm{eq}$ all finite sequences over $\mathbb{N}_\#$ and define the *content* of such sequences analogous to the content of texts. The concatenation of two sequences $\sigma, \tau \in \mathbb{S}\mathrm{eq}$ is denoted by $\sigma\tau$ or, more emphasizing, $\sigma^\frown\tau$. Furthermore, we write $\subseteq$ for the *extension relation* on sequences and fix a order $\leq$ on $\mathbb{S}\mathrm{eq}$ interpreted as natural numbers.

Now, we formalize learning criteria using the following system [15]. A *learner* is a partial function $h \in \mathcal{P}$. An *interaction operator* β is an operator that takes a learner $h \in \mathcal{P}$ and a text $T \in \mathbf{Txt}$ as input and outputs a (possibly partial) function p. Intuitively, β defines which information is available to the learner for making its hypothesis. We consider *Gold-style* or *full-information* learning [10], denoted by $\mathbf{G}$, *partially set-driven* learning ($\mathbf{Psd}$, [2,19]), *set-driven* learning ($\mathbf{Sd}$, [20]), *iterative* learning ($\mathbf{It}$, [21,8]) and *transductive* learning ($\mathbf{Td}$, [4,15]). To define the latter formally, we introduce a symbol

"?" for the learner to signalize that the information given is insufficient. Formally, for all learners $h \in \mathcal{P}$, texts $T \in \mathbf{Txt}$ and all $i \in \mathbb{N}$, define

$$\begin{aligned}
\mathbf{G}(h,T)(i) &= h(T[i]);\\
\mathbf{Psd}(h,T)(i) &= h(\mathrm{content}(T[i]),i);\\
\mathbf{Sd}(h,T)(i) &= h(\mathrm{content}(T[i]));\\
\mathbf{It}(h,T)(i) &= \begin{cases} h(\varepsilon), & \text{if } i=0;\\ h(\mathbf{It}(h,T)(i-1),T(i-1)), & \text{otherwise;}\end{cases}\\
\mathbf{Td}(h,T)(i) &= \begin{cases} ?, & \text{if } i=0;\\ \mathbf{Td}(h,T)(i-1), & \text{else, if } h(T(i-1))=?;\\ h(T(i-1)), & \text{otherwise.}\end{cases}
\end{aligned}$$

For any of the named interaction operators β, given a β-learner h, we let h^* (the *starred* learner) denote a $\mathbf{G}$-learner simulating h, i.e., for all $T \in \mathbf{Txt}$, we have $\beta(h,T) = \mathbf{G}(h^*,T)$. For example, let h be a $\mathbf{Sd}$-learner. Then, intuitively, h^* ignores all information but the content of the input, simulating h with this information, i.e., for all finite sequences σ, we have $h^*(\sigma) = h(\mathrm{content}(\sigma))$.

For a learner to successfully identify a language, we may oppose constraints on the hypotheses the learner makes. These are called *learning restrictions*. As a first, famous example, we required the learner to be *explanatory* [10], i.e., the learner must converge to a *single*, correct hypothesis for the target language. We hereby distinguish whether the final hypothesis is interpreted as a C-index ($\mathbf{Ex}_C$) or as a W-index ($\mathbf{Ex}_W$). Formally, for any sequence of hypotheses p and text $T \in \mathbf{Txt}$, we have

$$\begin{aligned}
\mathbf{Ex}_C(p,T) &\Leftrightarrow \exists n_0 \colon \forall n \geq n_0 \colon p(n) = p(n_0) \wedge \varphi_{p(n_0)} = \chi_{\mathrm{content}(T)};\\
\mathbf{Ex}_W(p,T) &\Leftrightarrow \exists n_0 \colon \forall n \geq n_0 \colon p(n) = p(n_0) \wedge W_{p(n_0)} = \mathrm{content}(T).
\end{aligned}$$

We say that explanatory learning requires *syntactic* convergence. If there exists a C-index (or W-index) for a language, then there exist infinitely many. This motivates to not require syntactic but only *semantic* convergence, i.e., the learner may make mind changes, but it has to, eventually, only output correct hypotheses. This is called *behaviorally correct* learning ($\mathbf{Bc}_C$ or $\mathbf{Bc}_W$, [6,17]). Formally, let p be a sequence of hypotheses and let $T \in \mathbf{Txt}$, then

$$\begin{aligned}
\mathbf{Bc}_C(p,T) &\Leftrightarrow \exists n_0 \colon \forall n \geq n_0 \colon \varphi_{p(n)} = \chi_{\mathrm{content}(T)};\\
\mathbf{Bc}_W(p,T) &\Leftrightarrow \exists n_0 \colon \forall n \geq n_0 \colon W_{p(n)} = \mathrm{content}(T).
\end{aligned}$$

In this paper, we consider learning with C-indices. It is, thus, natural to require the hypotheses to consist solely of C-indices, called *C-index learning*, and denoted by $\mathbf{CInd}$. Formally, for a sequence of hypotheses p and a text T, we have

$$\mathbf{CInd}(p,T) \Leftrightarrow \forall i,x \colon \varphi_{p(i)}(x) \in \{0,1\}\,.$$

For two learning restrictions δ and δ', their combination is their intersection, denoted by their juxtaposition $\delta\delta'$. We let $\mathbf{T}$ denote the learning restriction that is always true, which is interpreted as the absence of a learning restriction.

A *learning criterion* is a tuple $(\alpha, \mathcal{C}, \beta, \delta)$, where $\mathcal{C}$ is the set of admissible learners, usually $\mathcal{P}$ or $\mathcal{R}$, β is an interaction operator and α and δ are learning restrictions. We denote this criterion with $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$, omitting $\mathcal{C}$ if $\mathcal{C} = \mathcal{P}$, and a learning restriction if it equals $\mathbf{T}$. We say that an admissible learner $h \in \mathcal{C}$ $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learns a language L if and only if, for arbitrary texts $T \in \mathbf{Txt}$, we have $\alpha(\beta(h, T), T)$ and for all texts $T \in \mathbf{Txt}(L)$ we have $\delta(\beta(h, T), T)$. The set of languages $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learned by $h \in \mathcal{C}$ is denoted by $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta(h)$. With $[\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta]$ we denote the set of all classes $\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta$-learnable by some learner in $\mathcal{C}$. Moreover, to compare learning with W- and C-indices, these classes may only contain recursive languages, which we denote as $[\tau(\alpha)\mathcal{C}\mathbf{Txt}\beta\delta]_{\mathbf{REC}}$.

2.2 Normal Forms

When studying language learning in the limit, there are certain properties of learner that are useful, e.g., if we can assume a learner to be total. Cases where learners may be assumed total have been studied in the literature [13,14]. Importantly, this is the case for explanatory Gold-style learners obeying delayable learning restrictions and for behaviorally correct learners obeying delayable restrictions. Intuitively, a learning restriction is *delayable* if it allows hypotheses to be arbitrarily, but not indefinitely postponed without violating the restriction. Formally, a learning restriction δ is delayable, if and only if for all non-decreasing, unbounded functions $r\colon \mathbb{N} \to \mathbb{N}$, texts $T, T' \in \mathbf{Txt}$ and learning sequences p such that for all $n \in \mathbb{N}$, $\mathrm{content}(T[r(n)]) \subseteq \mathrm{content}(T'[n])$ and $\mathrm{content}(T) = \mathrm{content}(T')$, we have, if $\delta(p, T)$, then also $\delta(p \circ r, T')$. Note that $\mathbf{Ex}_W$, $\mathbf{Ex}_C$, $\mathbf{Bc}_W$, $\mathbf{Bc}_C$ and $\mathbf{CInd}$ are delayable restrictions.

Another useful notion are *locking sequences*. Intuitively, these contain enough information such that a learner, after seeing this information, converges correctly and does not change its mind anymore whatever additional information from the target language it is given. Formally, let L be a language and let $\sigma \in L^*_{\#}$. Given a $\mathbf{G}$-learner $h \in \mathcal{P}$, σ is a *locking sequence* for h on L if and only if for all sequences $\tau \in L^*_{\#}$ we have $h(\sigma) = h(\sigma\tau)$ and $h(\sigma)$ is a correct hypothesis for L [2]. This concept can immediately be transferred to other interaction operators. Exemplary, given a $\mathbf{Sd}$-learner h and a locking sequence σ of the starred learner h^*, we call the set $\mathrm{content}(\sigma)$ a *locking set*. Analogously, one transfers this definition to the other interaction operators. It shall not remain unmentioned that, when considering $\mathbf{Psd}$-learners, we speak of *locking information*. In the case of $\mathbf{Bc}_W$-learning we do not require the learner to syntactically converge. Therefore, we call a sequence $\sigma \in L^*_{\#}$ a $\mathbf{Bc}_W$*-locking sequence* for a $\mathbf{G}$-learner h on L if, for all sequences $\tau \in L^*_{\#}$, $h(\sigma\tau)$ is a correct hypothesis for L [11]. We omit the transfer to other interaction operators as it is immediate. It is an important observation that for any learner h and any language L it learns, there exists a ($\mathbf{Bc}_W$-) locking sequence [2]. These notions and results directly transfer to $\mathbf{Ex}_C$- and $\mathbf{Bc}_C$-learning. When it is clear from the context, we omit the index.

3 Requiring C-Indices as Output

This section is dedicated to proving Figure 1(a), giving all pairwise relations for the different settings of requiring C-indices for output in the various mentioned modes of data

presentation. In general, we observe that the later we require C-indices, the more learning power the learner has. This holds except for transductive learners which converge to C-indices. We show that they are as powerful as $\mathbf{CInd}$-transductive learners.

Although we learn classes of recursive languages, the requirement to converge to characteristic indices does heavily limit a learners capabilities. In the next theorem we show that even transductive learners which converge to W-indices can learn classes of languages which no Gold-style $\mathbf{Ex}_C$-learner can learn. We exploit the fact that C-indices, even if only conjectured eventually, must contain both positive and negative information about the guess.

Theorem 1. *We have that* $[\mathbf{TxtTdEx}_W]_{\mathbf{REC}} \setminus [\mathbf{TxtGEx}_C]_{\mathbf{REC}} \neq \emptyset$.

Proof. We show this by using the Operator Recursion Theorem (**ORT**) to provide a separating class of languages. To this end, let h be the $\mathbf{Td}$-learner with $h(\#) = ?$ and, for all $x, y \in \mathbb{N}$, let $h(\langle x, y\rangle) = x$. Let $\mathcal{L} = \mathbf{TxtTdEx}_W(h) \cap \mathbf{REC}$. Assume $\mathcal{L}$ can be learned by a $\mathbf{TxtGEx}_C$-learner h'. We may assume $h' \in \mathcal{R}$ [13]. Then, by **ORT** there exist indices $e, p, q \in \mathbb{N}$ such that

$$
\begin{aligned}
L \coloneqq W_e &= \operatorname{range}(\varphi_p);\\
\forall x\colon \tilde{T}(x) \coloneqq \varphi_p(x) &= \langle e, \varphi_q(\tilde{T}[x])\rangle;\\
\varphi_q(\varepsilon) &= 0;\\
\forall \sigma \neq \varepsilon\colon \bar{\sigma} &= \min\{\sigma' \subseteq \sigma \mid \varphi_q(\sigma') = \varphi_q(\sigma)\};\\
\forall \sigma \neq \varepsilon\colon \varphi_q(\sigma) &= \begin{cases}
\varphi_q(\bar{\sigma}), & \text{if } \forall \sigma', \bar{\sigma} \subseteq \sigma' \subseteq \sigma\colon \Phi_{h'(\sigma')}(\langle e, \varphi_q(\bar{\sigma}) + 1\rangle) > |\sigma|;\\
\varphi_q(\bar{\sigma}) + 1, & \text{else, for min. } \sigma' \text{ contradicting the previous case, if}\\
& \qquad \varphi_{h'(\sigma')}(\langle e, \varphi_q(\bar{\sigma}) + 1\rangle) = 0;\\
\varphi_q(\bar{\sigma}) + 2, & \text{otherwise.}
\end{cases}
\end{aligned}
$$

Here, Φ is a Blum complexity measure [3]. Intuitively, to define the next $\varphi_p(x)$, we add the same element to $\operatorname{content}(\tilde{T})$ until we know whether $\langle e, \tilde{T}[x] + 1\rangle \in C_{h'(\bar{\sigma})}$ holds or not. Then, we add the element contradicting this outcome.

We first show that $L \in \mathcal{L}$ and afterwards that L cannot be learned by h'. To show the former, note that either L is finite or $\tilde{T}$ is a non-decreasing unbounded computable enumeration of L. Therefore, we have $L \in \mathbf{REC}$. We now prove that h learns L. Let $T \in \mathbf{Txt}(L)$. For all $n \in \mathbb{N}$ where $T(n)$ is not the pause symbol, we have $h(T(n)) = e$. With $n_0 \in \mathbb{N}$ being minimal such that $T(n_0) \neq \#$, we get for all $n \geq n_0$ that $\mathbf{Td}(h, T)(n) = e$. As e is a correct hypothesis, h learns L from T and thus we have that $L \in \mathbf{TxtTdEx}_W(h)$. Altogether, we get that $L \in \mathcal{L}$.

By assumption, h' learns L from the text $\tilde{T} \in \mathbf{Txt}(L)$. Therefore, there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,

$$h'(\tilde{T}[n]) = h'(\tilde{T}[n_0]) \text{ and } \chi_L = \varphi_{h'(\tilde{T}[n])},$$

that is, $h'(\tilde{T}[n])$ is a C-index for L. Now, as h' outputs C-indices when converging, there are $t, t' \geq n_0$ such that

$$\Phi_{h'(\tilde{T}[t'])}(\langle e, \varphi_q(\tilde{T}[n_0]) + 1\rangle) \leq t.$$

Let t'_0 and t_0 be the first such found. We show that $h'(\tilde{T}[t'_0])$ is no correct hypothesis of L by distinguishing the following cases.

1. Case: $\varphi_{h'(\tilde{T}[t'_0])}(\langle e, \varphi_q(\tilde{T}[n_0])+1\rangle) = 0$. By definition of φ_q and by minimality of t'_0, we have that $\langle e, \varphi_q(\tilde{T}[n_0])+1\rangle \in L$, however, the hypothesis of $h'(\tilde{T}[t'_0])$ says differently, a contradiction.
2. Case: $\varphi_{h'(\tilde{T}[t'_0])}(\langle e, \varphi_q(\tilde{T}[n_0])+1\rangle) = 1$. By definition of φ_q and by minimality of t'_0, we have that $\langle e, \varphi_q(\tilde{T}[n_0])+1\rangle \in L$, but $\langle e, \varphi_q(\tilde{T}[n_0])+1\rangle \notin L$. However, the hypothesis of $h'(\tilde{T}[t'_0])$ conjectures the latter to be in L, a contradiction. □

Furthermore, the following known equalities from learning W-indices directly apply in the studied setting as well.

Theorem 2 ([12], [19,9]). *We have that*

$$[\mathbf{TxtItEx}_W]_{\mathbf{REC}} \subseteq [\mathbf{TxtSdEx}_W]_{\mathbf{REC}},$$
$$[\mathbf{TxtPsdEx}_W]_{\mathbf{REC}} = [\mathbf{TxtGEx}_W]_{\mathbf{REC}}.$$

The remaining separations we will show in a more general way, see Theorems 11 and 12. We generalize the latter result [19,9], namely that Gold-style learners may be assumed partially set-driven, to all considered cases. The idea here is to, just as in the $\mathbf{Ex}_W$-case, mimic the given learner and to search for minimal locking sequences. Incorporating the result that unrestricted Gold-style learners may be assumed total [13], we even get a stronger result.

Theorem 3. *For $\delta, \delta' \in \{\mathbf{CInd}, \mathbf{T}\}$, we have that*

$$[\tau(\delta)\mathbf{TxtG}\delta'\mathbf{Ex}_C]_{\mathbf{REC}} = [\tau(\delta)\mathcal{R}\mathbf{TxtPsd}\delta'\mathbf{Ex}_C]_{\mathbf{REC}}.$$

We also generalize the former result of Theorem 2 to hold in all considered cases. The same simulating argument (where one mimics the iterative learner on ascending text with a pause symbol between two elements) suffices regardless the exact setting.

Theorem 4. *Let $\delta, \delta' \in \{\mathbf{CInd}, \mathbf{T}\}$ and $\mathcal{C} \in \{\mathcal{R}, \mathcal{P}\}$. Then, we have that*

$$[\tau(\delta')\mathcal{C}\mathbf{TxtIt}\delta\mathbf{Ex}_C]_{\mathbf{REC}} \subseteq [\tau(\delta')\mathcal{C}\mathbf{TxtSd}\delta\mathbf{Ex}_C]_{\mathbf{REC}}.$$

Interestingly, totality is not restrictive solely for Gold-style (and due to the equality also partially set-driven) learners. For the other considered learners with restricted memory, being total lessens the learning capabilities. This weakness results from the need to output some guess. A partial learner can await this guess and outperform it. This way, we obtain self-learning languages [5] to show the three following separations.

Theorem 5. *We have that* $[\mathcal{R}\mathbf{TxtSdCIndEx}_C]_{\mathbf{REC}} \subsetneq [\mathbf{TxtSdCIndEx}_C]_{\mathbf{REC}}$.

Theorem 6. *We have that* $[\mathcal{R}\mathbf{TxtItCIndEx}_C]_{\mathbf{REC}} \subsetneq [\mathbf{TxtItCIndEx}_C]_{\mathbf{REC}}$.

Theorem 7. *We have that* $[\mathcal{R}\mathbf{TxtTdCIndEx}_C]_{\mathbf{REC}} \subsetneq [\mathbf{TxtTdCIndEx}_C]_{\mathbf{REC}}$.

Next, we show the gradual decrease of learning power the more we require the learners to output characteristic indices. We have already seen in Theorem 1 that converging to C-indices lessens learning power. However, this allows for more learning power than outputting these indices during the whole learning process as shows the next theorem. The idea is that such learners have to be certain about their guesses as these are indices of characteristic functions. When constructing a separating class using self-learning languages [5], one forces the $\mathbf{CInd}$-learner to output C-indices on certain languages to, then, contradict its choice there. This way, the $\mathbf{Ex}_C$-learner learns languages the $\mathbf{CInd}$-learner cannot.

Theorem 8. *We have that* $[\mathbf{TxtItEx}_C]_{\mathbf{REC}} \setminus [\mathbf{TxtGCIndBc}_C]_{\mathbf{REC}} \neq \emptyset$.

Since languages which can be learned by iterative learners can also be learned by set-driven ones (see Theorem 4), this result suffices. Note that the idea above requires some knowledge on previous elements. Thus, it is no coincidence that this separation does not include transductive learners. Since these learners base their guesses on single elements, they cannot see how far in the learning process they are. Thus, they are forced to always output C-indices.

Theorem 9. *We have that* $[\mathbf{TxtTdCIndEx}_C]_{\mathbf{REC}} = [\mathbf{TxtTdEx}_C]_{\mathbf{REC}}$.

For the remainder of this section, we focus on learners which output characteristic indices on *arbitrary* input, that is, we focus on $\tau(\mathbf{CInd})$-learners. First, we show that the requirement of always outputting C-indices lessens a learners learning power, even when compared to total $\mathbf{CInd}$-learners. To provide the separating class of self-learning languages, one again awaits the $\tau(\mathbf{CInd})$-learner's decision and then, based on these, learns languages this learner cannot.

Theorem 10. *We have* $[\mathcal{R}\mathbf{TxtTdCIndEx}_C]_{\mathbf{REC}} \setminus [\tau(\mathbf{CInd})\mathbf{TxtGBc}_C]_{\mathbf{REC}} \neq \emptyset$.

Proof. We prove the result by providing a separating class of languages. Let h be the $\mathbf{Td}$-learner with $h(\#) = ?$ and, for all $x, y \in \mathbb{N}$, let $h(\langle x, y\rangle) = x$. By construction, h is total and computable. Let $\mathcal{L} = \mathcal{R}\mathbf{TxtTdCIndEx}_C(h) \cap \mathbf{REC}$. We show that there is no $\tau(\mathbf{CInd})\mathbf{TxtGBc}_C$-learner learning $\mathcal{L}$ by way of contradiction. Assume there is a $\tau(\mathbf{CInd})\mathbf{TxtGBc}_C$-learner h' which learns $\mathcal{L}$. With the Operator Recursion Theorem (**ORT**), there are $e, p \in \mathbb{N}$ such that for all $x \in \mathbb{N}$

$$
\begin{aligned}
L &\coloneqq \operatorname{range}(\varphi_p);\\
\varphi_e &= \chi_L;\\
\tilde{T}(x) \coloneqq \varphi_p(x) &= \begin{cases} \langle e, 2x\rangle, & \text{if } \varphi_{h'(\varphi_p[x])}(\langle e, 2x\rangle) = 0;\\ \langle e, 2x+1\rangle, & \text{otherwise.}\end{cases}
\end{aligned}
$$

Intuitively, for all x either $\varphi_p(x)$ is an element of L if it is not in the hypothesis of h' after seeing $\varphi_p[x]$, or there is an element in this hypothesis that is not in $\operatorname{content}(\tilde{T})$. As any hypothesis of h' is a C-index, we have that $\varphi_p \in \mathcal{R}$ and, as φ_p is strictly monotonically increasing, that L is decidable.

We now prove that $L \in \mathcal{L}$ and afterwards that L cannot be learned by h'. First, we need to prove that h learns L. Let $T \in \mathbf{Txt}(L)$. For all $n \in \mathbb{N}$ where $T(n)$ is not the pause symbol, we have $h(T(n)) = e$. Let $n_0 \in \mathbb{N}$ with $T(n_0) \neq \#$. Then, we have, for all $n \geq n_0$, that $\mathbf{Td}(h, T)(n) = e$ and, since e is a hypothesis of L, h learns L from T. Thus, we have that $L \in \mathcal{R}\mathbf{TxtTdCIndEx}_C(h) \cap \mathbf{REC}$.

By assumption, h' learns $\mathcal{L}$ and thus it also needs to learn L on text $\tilde{T}$. Hence, there is x_0 such that for all $x \geq x_0$ the hypothesis $h'(\tilde{T}[x]) = h'(\varphi_p[x])$ is a C-index for L. We now consider the following cases.

1. Case: $\varphi_{h'(\varphi_p[x])}(\langle e, 2x\rangle) = 0$. By construction, we have that $\tilde{T}(x) = \langle e, 2x\rangle$. Therefore, $\langle e, 2x\rangle \in L$, which contradicts $h'(\varphi_p[x])$ being a correct hypothesis.
2. Case: $\varphi_{h'(\varphi_p[x])}(\langle e, 2x\rangle) = 1$. By construction, we have that $\tilde{T}(x) \neq \langle e, 2x\rangle$ and thus, because $\tilde{T}$ is strictly monotonically increasing, $\langle e, 2x\rangle \notin L = \mathrm{content}(\tilde{T})$. This, again, contradicts $h'(\varphi_p[x])$ being a correct hypothesis.

As in all cases $h'(\varphi_p[x])$ is a wrong hypothesis, h' cannot learn $\mathcal{L}$. □

It remains to be shown that memory restrictions are severe for such learners as well. First, we show that partially set-driven learners are more powerful than set-driven ones. Just as originally witnessed by for W-indices [19,9], this is solely due to the lack of learning time. In the following theorem, we already separate from behaviorally correct learners, as we will need this stronger version later on.

Theorem 11. *We have that* $[\tau(\mathbf{CInd})\mathbf{TxtPsdEx}_C]_{\mathbf{REC}} \setminus [\mathbf{TxtSdBc}_W]_{\mathbf{REC}} \neq \emptyset$.

In turn, this lack of time is not as severe as lack of memory. The standard class (of recursive languages) to separate set-driven learners from iterative ones [11] can be transferred to the setting studied in this paper.

Theorem 12. *We have that* $[\tau(\mathbf{CInd})\mathbf{TxtSdEx}_C]_{\mathbf{REC}} \setminus [\mathbf{TxtItEx}_W]_{\mathbf{REC}} \neq \emptyset$.

Lastly, we show that transductive learners, having basically no memory, do severely lack learning power. As they have to infer their conjectures from single elements they, in fact, cannot even learn basic classes such as $\{\{0\}, \{1\}, \{0, 1\}\}$. The following result concludes the map shown in Figure 1(a) and, therefore, also this section.

Theorem 13. *For* $\beta \in \{\mathbf{It}, \mathbf{Sd}\}$*, we have that*

$$[\tau(\mathbf{CInd})\mathbf{Txt}\beta\mathbf{Ex}_C]_{\mathbf{REC}} \setminus [\mathbf{TxtTdEx}_W]_{\mathbf{REC}} \neq \emptyset.$$

4 Syntactic versus Semantic Convergence to C-indices

In this section, we investigate the effects on learners when we require them to converge to characteristic indices. We study both syntactically converging learners as well as semantically converging ones. In particular, we compare learners imposed with different well-studied memory restrictions.

Surprisingly, we observe that, although C-indices incorporate and, thus, require the learner to obtain more information during the learning process than W-indices, the relative relations of the considered restrictions remain the same. We start by gathering results which directly follow from the previous section.

Corollary 1. *We have that*

$$\begin{aligned}
[\mathbf{TxtPsdEx}_C]_{\mathbf{REC}} &= [\mathbf{TxtGEx}_C]_{\mathbf{REC}}, (\textit{Theorem 3}),\\
[\mathbf{TxtItEx}_C]_{\mathbf{REC}} &\subseteq [\mathbf{TxtSdEx}_C]_{\mathbf{REC}}, (\textit{Theorem 4}),\\
[\mathbf{TxtGEx}_C]_{\mathbf{REC}} &\setminus [\mathbf{TxtSdBc}_C]_{\mathbf{REC}} \neq \emptyset, (\textit{Theorem 11}),\\
[\mathbf{TxtSdEx}_C]_{\mathbf{REC}} &\setminus [\mathbf{TxtItEx}_C]_{\mathbf{REC}} \neq \emptyset, (\textit{Theorem 12}),\\
[\mathbf{TxtItEx}_C]_{\mathbf{REC}} &\setminus [\mathbf{TxtTdEx}_C]_{\mathbf{REC}} \neq \emptyset, (\textit{Theorem 13}).
\end{aligned}$$

We show the remaining results. First, we show that, just as for W-indices, behaviorally correct learners are more powerful than explanatory ones. We provide a separating class exploiting that explanatory learners must converge to a single, correct hypothesis. We collect elements on which mind changes are witnessed, while maintaining decidability of the obtained language.

Theorem 14. *We have that* $[\mathbf{TxtSdBc}_C]_{\mathbf{REC}} \setminus [\mathbf{TxtGEx}_C]_{\mathbf{REC}} \neq \emptyset$.

Next, we show that, just as for W-indices, a padding argument makes iterative behaviorally correct learners as powerful as Gold-style ones.

Theorem 15. *We have that* $[\mathbf{TxtItBc}_C]_{\mathbf{REC}} = [\mathbf{TxtGBc}_C]_{\mathbf{REC}}$.

We show that the classes of languages learnable by some behaviorally correct Gold-style (or, equivalently, iterative) learner, can also be learned by partially set-driven ones. We follow the proof which is given in a private communication with Sanjay Jain [7]. The idea there is to search for minimal $\mathbf{Bc}$-locking sequences without directly mimicking the $\mathbf{G}$-learner. We transfer this idea to hold when converging to C-indices as well. We remark that, while doing the necessary enumerations, one needs to make sure these are characteristic. One obtains this as the original learner eventually outputs characteristic indices.

Theorem 16. *We have that* $[\mathbf{TxtPsdBc}_C]_{\mathbf{REC}} = [\mathbf{TxtGBc}_C]_{\mathbf{REC}}$.

Lastly, we investigate transductive learners. Such learners base their hypotheses on a single element. Thus, one would expect them to benefit from dropping the requirement to converge to a *single* hypothesis. Interestingly, this does not hold true. This surprising fact originates from C-indices encoding characteristic functions. Thus, one can simply search for the minimal element on which no "?" is conjectured. The next result finalizes the map shown in Figure 1(b) and, thus, this section.

Theorem 17. *We have that* $[\mathbf{TxtTdEx}_C]_{\mathbf{REC}} = [\mathbf{TxtTdBc}_C]_{\mathbf{REC}}$.

References

1. Berger, J., Böther, M., Doskoč, V., Gadea Harder, J., Klodt, N., Kötzing, T., Lötzsch, W., Peters, J., Schiller, L., Seifert, L., Wells, A., Wietheger, S.: Learning languages with decidable hypotheses. CoRR (2020)

2. Blum, L., Blum, M.: Toward a mathematical theory of inductive inference. Information and Control **28**, 125–155 (1975)
3. Blum, M.: A machine-independent theory of the complexity of recursive functions. Journal of the ACM **14**, 322–336 (1967)
4. Carlucci, L., Case, J., Jain, S., Stephan, F.: Results on memory-limited U-shaped learning. Information and Computation **205**, 1551–1573 (2007)
5. Case, J., Kötzing, T.: Strongly non-U-shaped language learning results by general techniques. Information and Computation **251**, 1–15 (2016)
6. Case, J., Lynes, C.: Machine inductive inference and language identification. In: Proc. of the International Colloquium on Automata, Languages and Programming (ICALP). pp. 107–115 (1982)
7. Doskoč, V., Kötzing, T.: Cautious limit learning. In: Proc. of the International Conference on Algorithmic Learning Theory (ALT) (2020)
8. Fulk, M.: A Study of Inductive Inference Machines. Ph.D. thesis (1985)
9. Fulk, M.A.: Prudence and other conditions on formal language learning. Information and Computation **85**, 1–11 (1990)
10. Gold, E.M.: Language identification in the limit. Information and Control **10**, 447–474 (1967)
11. Jain, S., Osherson, D., Royer, J.S., Sharma, A.: Systems that Learn: An Introduction to Learning Theory. MIT Press, Cambridge (MA), Second Edition (1999)
12. Kinber, E.B., Stephan, F.: Language learning from texts: Mindchanges, limited memory, and monotonicity. Information and Computation **123**, 224–241 (1995)
13. Kötzing, T., Palenta, R.: A map of update constraints in inductive inference. Theoretical Computer Science **650**, 4–24 (2016)
14. Kötzing, T., Schirneck, M., Seidel, K.: Normal forms in semantic language identification. In: Proc. of the International Conference on Algorithmic Learning Theory (ALT). pp. 76:493–76:516 (2017)
15. Kötzing, T.: Abstraction and Complexity in Computational Learning in the Limit. Ph.D. thesis, University of Delaware (2009)
16. Lange, S., Zeugmann, T., Zilles, S.: Learning indexed families of recursive languages from positive data: A survey. Theoretical Computer Science **397**, 194–232 (2008)
17. Osherson, D.N., Weinstein, S.: Criteria of language learning. Information and Control **52**, 123–138 (1982)
18. Rogers Jr., H.: Theory of recursive functions and effective computability. Reprinted by MIT Press, Cambridge (MA) (1987)
19. Schäfer-Richter, G.: Über Eingabeabhängigkeit und Komplexität von Inferenzstrategien. Ph.D. thesis, RWTH Aachen University, Germany (1984)
20. Wexler, K., Culicover, P.W.: Formal principles of language acquisition. MIT Press, Cambridge (MA) (1980)
21. Wiehagen, R.: Limes-Erkennung rekursiver Funktionen durch spezielle Strategien. Journal of Information Processing and Cybernetics **12**, 93–99 (1976)