On the Teaching Complexity of Linear Sets

Ziyuan Gao[1], Hans Ulrich Simon[2], and Sandra Zilles[1]

[1] Department of Computer Science
University of Regina, Regina, SK, Canada S4S 0A2
Email: {gao257,zilles}@cs.uregina.ca
[2] Horst Görtz Institute for IT Security and Faculty of Mathematics,
Ruhr-Universität Bochum, D-44780 Bochum, Germany
Email: hans.simon@rub.de

Abstract. Linear sets are the building blocks of semilinear sets, which are in turn closely connected to automata theory and formal languages. Prior work has investigated the learnability of linear sets and semilinear sets in three models – Valiant's *PAC-learning* model, Gold's *learning in the limit* model, and Angluin's *query learning* model. This paper considers a *teacher-learner* model of learning families of linear sets, whereby the learner is assumed to know all the smallest sets $T_1, T_2, \ldots$ of labelled examples that are consistent with exactly one language in the class $\mathcal{L}$ to be learnt, and is always presented with a sample S of labelled examples such that S is contained in at least one of $T_1, T_2, \ldots$; the learner then interprets S according to some fixed protocol. In particular, we will apply a generalisation of a recently introduced model – the *recursive teaching model* of teaching and learning – to several infinite classes of linear sets, and show that the maximum sample complexity of teaching these classes can be drastically reduced if each of them is taught according to a carefully chosen sequence. A major focus of the paper will be on determining two relevant teaching parameters, the *teaching dimension* and *recursive teaching dimension*, for various families of linear sets.

1 Introduction

A *linear set* L is defined by a nonnegative lattice point (called a *constant*) and a finite set of nonnegative lattice sets (called *periods*); the members of L are generated by adding to the constant an arbitrary finite sequence of the periods (allowing repetitions of the same period in the sequence). A *semilinear set* is a finite union of linear sets. Semilinear sets are not only objects of mathematical interest, but have also been linked to finite-state machines and formal languages. One of the earliest and most important results on the connection between semilinear sets and context-free languages is *Parikh's theorem* [9], which states that any context-free language is mapped to a semilinear set via a function known as the Parikh vector of a string. Another interesting result, due to Ibarra [6], characterises semilinear sets in terms of reversal-bounded multicounter machines. Moving beyond abstract theory, semilinear sets have also recently been applied in the fields of DNA self-assembly [3] and membrane computing [7].

The learnabilities of linear sets and semilinear sets have been investigated in Valiant's PAC-learning model [1], Gold's learning in the limit model [12], and Angluin's query learning model [12]. Abe [1] showed that when the integers are encoded in unary, the class of semilinear sets of dimension 1 or 2 is polynomially PAC-learnable; on the other hand, the question as to whether classes of semilinear sets of higher dimensions are PAC-learnable is open. Takada [12] established that for any fixed dimension, the family of linear sets is learnable from positive examples but the family of semilinear sets is not learnable from only positive examples. Takada also showed the existence of a learning procedure via restricted subset and restricted superset queries that identifies any semilinear set and halts; however, he proved at the same time that any such algorithm must necessarily be time consuming.

This paper is primarily concerned with the *sample complexity of teaching* classes of linear sets with a fixed dimension, which we determine mainly with two combinatorial parameters (and some variants), the *teaching dimension* (TD) and the *recursive teaching dimension* (RTD). These teaching complexity measures are based on a variant of the online learning model in which a cooperative teacher selects the instances presented to the learner [5, 13, 11]. In the teacher-learner model, the teacher must present a finite set of instances so that the learner achieves *exact identification* of the target concept via some consistent teaching-learning protocol. To preclude any unnatural collusion between the teacher and learner that could arise from, say, encoding concepts in examples, the teaching-learning protocol must, in some definite sense, be "collusion-free." To this end, Zilles, Lange, Holte and Zinkevich [13] proposed a rigorous definition of a "collusion-free" teaching-learning protocol. They designed a protocol – the *recursive teaching protocol* – that only exploits an inherent hierarchical structure of any concept class, and showed that this protocol is collusion-free. The RTD of a concept class is the maximum sample complexity derived by applying the recursive teaching protocol to the class. The RTD possesses several regularity properties and has been fruitfully applied to the analysis of pattern languages [4, 8]. A somewhat simpler protocol, the teaching set protocol [5, 11], only requires that the teacher present, for each target concept C, a sample S from C of smallest possible size so that C is the only concept in the class consistent with S. The teaching set protocol is also collusion-free, although the maximum sample complexity in this case – the TD – is generally larger than the RTD.

Our results may be of interest from a formal language perspective as well as from a computational learning theory perspective. First, they uncover a number of structural properties of linear sets, especially in the one-dimensional case, which could be applied to study formal languages via the Parikh vector function. Consider, for example, the set $L(\pi)$ of all words obtained by substituting nonempty strings over $\{a\}$ for variables in some nonempty string π of symbols chosen from $\{a\} \cup X$, where X is an infinite set of variables.[3] As will be seen later, the Parikh vector maps $L(\pi)$ to a linear subset L of the natural numbers such that the sum of L's periods does not exceed the constant associated

[3] $L(\pi)$ is known as a *non-erasing pattern language*.

to L. Thus one could determine various teaching complexity measures of any non-erasing pattern language from the teaching complexity measures of a certain linear subset of the natural numbers. Second, the class of linear sets affords quite a natural setting to study models of teaching and learning over infinite concept classes. Besides showing that the RTD can be significantly lower than the TD for many infinite classes of linear sets, we will consider a more stringent variant of the RTD, the RTD$^+$, which considers sequential teaching of classes using only positive examples. It will be shown that there are natural classes of linear sets that cannot even be taught sequentially using only positive examples while the RTD is finite; these examples illustrate how supplying negative information may sometimes be indispensable to successful teaching and learning.

2 Preliminaries

$\mathbb{N}_0$ denotes the set of all nonnegative integers and $\mathbb{N}$ denotes the set of all positive integers. For each integer $m \geq 1$, let $\mathbb{N}_0^m = \mathbb{N}_0 \times \ldots \times \mathbb{N}_0$ (m times). $\mathbb{N}_0^m$ is regarded as a subset of the vector space of all m-tuples of rational numbers over the rational numbers. For each $r \in \mathbb{N}$, $[r]$ denotes $\{1, \ldots, r\}$. For any $v = (a_1, \ldots, a_m) \in \mathbb{N}_0^m$, define $\|v\|_1 = \sum_{i=1}^m a_i$. 0 will denote the zero vector in $\mathbb{N}_0^m$ when there is no possibility of confusion.

2.1 Linear Sets

A subset L of $\mathbb{N}_0^m$ is said to be *linear* iff there exist an element c and a finite subset P of $\mathbb{N}_0^m$ such that $L = c + \langle P \rangle := \{q : q = c + n_1 p_1 + \ldots + n_k p_k, n_i \in \mathbb{N}_0, p_i \in P\}$. c is called the *constant* and each p_i is called a *period* of $c + \langle P \rangle$. Denote $0 + \langle P \rangle$ by $\langle P \rangle$. For any linear set L, if $L = c + \langle P \rangle$, then (c, P) is called a *representation of* L. Any finite $P \subset \mathbb{N}_0^m$ is *independent* iff for all $P' \subsetneq P$, it holds that $\langle P' \rangle \neq \langle P \rangle$. A representation (c, P) of a linear set L is *canonical* iff P is independent. A linear subset of $\mathbb{N}_0^m$ will also be called a *linear set of dimension* m. $\langle \{p_1, \ldots, p_k\} \rangle$ will often be written as $\langle p_1, \ldots, p_k \rangle$ and $d\langle p_1, \ldots, p_k \rangle$ will denote $\langle dp_1, \ldots, dp_k \rangle$.

Our paper will focus on the linear subsets of $\mathbb{N}_0$. The main classes of linear sets investigated are denoted as follows. In these definitions, $k \in \mathbb{N}$.

(I) $\mathrm{LINSET}_k := \{\langle P \rangle : P \subset \mathbb{N}_0 \wedge \exists p \in P[p \neq 0] \wedge |P| \leq k\}$.
(II) $\mathrm{LINSET} := \bigcup_{k \in \mathbb{N}} \mathrm{LINSET}_k$.
(III) CF–LINSET$_k$[4] $:= \{\langle P \rangle : \emptyset \neq P \subset \mathbb{N} \wedge gcd(P) = 1 \wedge |P| \leq k\}$.
(IV) CF–LINSET $:= \bigcup_{k \in \mathbb{N}}$ CF–LINSET$_k$.
(V) NE–LINSET$_k$[5] $:= \{c + \langle P \rangle : c \in \mathbb{N}_0 \wedge P \subset \mathbb{N}_0 \wedge |P| \leq k \wedge \sum_{p \in P} p \leq c\}$.
(VI) NE–LINSET $:= \bigcup_{k \in \mathbb{N}}$ NE–LINSET$_k$.

Note that the classes in items (I) to (IV) exclude singleton linear sets; the reason for this omission will be explained later. The motivation for studying each subfamily in items (III) to (VI) will be explained as it is introduced in the forthcoming sections.

[4] CF stands for “cofinite.”
[5] NE stands for “non-erasing.”

2.2 Teaching Dimension and Recursive Teaching Dimension

The two main teaching parameters studied in this paper are the *teaching dimension* and the *recursive teaching dimension*.

Let $\mathcal{L}$ be a family of subsets of $\mathbb{N}_0^m$. Let $L \in \mathcal{L}$ and T be a subset of $\mathbb{N}_0^m \times \{+,-\}$. Furthermore, let T^+ (resp. T^-) be the set of vectors in T that are labelled "+" (resp. "−"). $X(T)$ is defined to be $T^+ \cup T^-$. A subset L of $\mathbb{N}_0^m$ is said to be *consistent* with T iff $T^+ \subseteq L$ and $T^- \cap L = \emptyset$. T is a *teaching set* for L w.r.t. $\mathcal{L}$ iff T is consistent with L and for all $L' \in \mathcal{L} \setminus \{L\}$, T is not consistent with L'. Every element of $\mathbb{N}_0^m \times \{+,-\}$ is known as a *labelled example*.

Definition 1. [5, 11] Let $\mathcal{L}$ be any family of subsets of $\mathbb{N}_0^m$. Let $L \in \mathcal{L}$. The size of a smallest teaching set for L w.r.t. $\mathcal{L}$ is called the *teaching dimension of L w.r.t. $\mathcal{L}$*, denoted by $\mathrm{TD}(L,\mathcal{L})$. The *teaching dimension of* $\mathcal{L}$ is defined as $\sup\{\mathrm{TD}(L,\mathcal{L}) : L \in \mathcal{L}\}$ and is denoted by $\mathrm{TD}(\mathcal{L})$.

Another complexity parameter recently studied in computational learning theory is the recursive teaching dimension. It refers to the maximum size of teaching sets in a series of nested subfamilies of the family.

Definition 2. (Based on [13, 8]) Let $\mathcal{L}$ be any family of subsets of $\mathbb{N}_0^m$. A *teaching sequence for* $\mathcal{L}$ is any sequence $R = ((\mathcal{F}_0, d_0), (\mathcal{F}_1, d_1), \ldots)$ where (i) the families $\mathcal{F}_i$ form a partition of $\mathcal{L}$ and each $\mathcal{F}_i$ is nonempty, and (ii) $d_i = \mathrm{TD}(L, \mathcal{L} \setminus \bigcup_{0 \le j < i} \mathcal{F}_j)$ for all i and all $L \in \mathcal{F}_i$. $\sup\{d_i : i \in \mathbb{N}_0\}$ is called the *order of* R, and is denoted by $ord(R)$. The *recursive teaching dimension of* $\mathcal{L}$ is defined as $\inf\{ord(R) : R$ is a teaching sequence for $\mathcal{L}\}$ and is denoted by $\mathrm{RTD}(\mathcal{L})$.

One can also restrict the instances of the teaching sets in a teaching sequence to positive examples; the best possible order of such a teaching sequence will be denoted by RTD^+.

Definition 3. Let $\mathcal{L}$ be any family of subsets of $\mathbb{N}_0^m$. A *teaching sequence with positive examples for* $\mathcal{L}$ (or *positive teaching sequence for* $\mathcal{L}$) is any sequence $P = ((\mathcal{F}_0, d_0), (\mathcal{F}_1, d_1), \ldots)$ such that (i) the families $\mathcal{F}_i$ form a partition of $\mathcal{L}$ and each $\mathcal{F}_i$ is nonempty, and (ii) for all i and all $L \in \mathcal{F}_i$, there is a subset $S_L \subseteq L$ with $|S_L| = d_i < \infty$ such that for all $L' \in \bigcup_{j \ge i} \mathcal{F}_j$, it holds that $S_L \subseteq L' \Rightarrow L = L'$. $\sup\{d_i : i \in \mathbb{N}_0\}$ is called the *order of* P, and is denoted by $ord(P)$. If $\mathcal{L}$ has at least one teaching sequence with positive examples, then the *positive recursive teaching dimension of* $\mathcal{L}$ is defined as $\inf\{ord(P) : P$ is a teaching sequence with positive examples for $\mathcal{L}\}$ and is denoted by $\mathrm{RTD}^+(\mathcal{L})$. If $\mathcal{L}$ does not have any teaching sequence with positive examples, define $\mathrm{RTD}^+(\mathcal{L}) = \infty$.

A *teaching plan for* $\mathcal{L}$ is a teaching sequence $((\mathcal{F}_0, d_0), (\mathcal{F}_1, d_1), \ldots)$ for $\mathcal{L}$ such that $|\mathcal{F}_i| = 1$ for each i. A teaching plan $((\{L_0\}, d_0), (\{L_1\}, d_1), (\{L_2\}, d_2), \ldots)$ for $\mathcal{L}$ will often be written as $((L_0, S_0), (L_1, S_1), (L_2, S_2), \ldots)$, where S_i is a teaching set for L_i w.r.t. $\mathcal{L} \setminus \{L_j : 0 \le j < i\}$. A *teaching plan with positive examples* *for* $\mathcal{L}$ is defined analogously. Note that for any family $\mathcal{L}$, if $\mathrm{TD}(L, \mathcal{L}) = \infty$ for some $L \in \mathcal{L}$, then any teaching plan for $\mathcal{L}$ must have an infinite order. Moreover, TD, RTD and RTD^+ are monotonic, that is, for all $\mathcal{L}' \subseteq \mathcal{L}$ and $K \in \{\mathrm{TD}, \mathrm{RTD}, \mathrm{RTD}^+\}$, $K(\mathcal{L}') \leq K(\mathcal{L})$. Another useful fact is that for any family $\mathcal{L}$, $\inf\{\mathrm{TD}(L, \mathcal{L}) : L \in \mathcal{L}\} \leq \mathrm{RTD}(\mathcal{L})$.

For any $\mathcal{L}' \subseteq \mathcal{L}$, call $R = ((\mathcal{F}_0, d_0), (\mathcal{F}_1, d_1), \ldots)$ a *teaching subsequence for* $\mathcal{L}$ *covering* $\mathcal{L}'$ iff $\mathcal{F}_0, \mathcal{F}_1, \ldots$ are nonempty, pairwise disjoint subsets of $\mathcal{L}$ such that $\mathcal{L}' \subseteq \bigcup_{i \in \mathbb{N}_0} \mathcal{F}_i$ and $d_i = \mathrm{TD}(L, \mathcal{L} \setminus \bigcup_{0 \leq j < i} \mathcal{F}_j)$ for all i and all $L \in \mathcal{F}_i$. Define $ord(R) = \sup\{d_i : i \in \mathbb{N}_0\}$ and $\mathrm{RTD}(\mathcal{L}', \mathcal{L}) = \inf\{ord(R) :$ R is a teaching subsequence for $\mathcal{L}$ covering $\mathcal{L}'\}$.

A family $\mathcal{L}$ of subsets of $\mathbb{N}_0^m$ is said to have *finite thickness* [2] iff for every $v \in \mathbb{N}_0^m$, the class of linear sets in $\mathcal{L}$ that contain v is finite. Note that finite thickness is a sufficient condition for families that do not contain the empty set to have a teaching plan with positive examples. The proof is omitted.

Proposition 4. *Let $\mathcal{L}$ be a family of subsets of $\mathbb{N}_0^m$ such that $\mathcal{L}$ has finite thickness and $\emptyset \notin \mathcal{L}$. Then there exists a teaching plan with positive examples for $\mathcal{L}$, Q, such that $ord(Q) = RTD^+(\mathcal{L})$.*

The next proposition provides a necessary condition for any family to have a teaching sequence with positive examples. This condition will be used later to establish the non-existence of positive teaching sequences for some families of linear sets.

Proposition 5. *Let $\mathcal{L}$ be a family of subsets of $\mathbb{N}_0^m$ that has at least one positive teaching sequence. Then for every $L \in \mathcal{L}$, there does not exist any infinite descending chain $H_0 \supsetneq H_1 \supsetneq H_2 \supsetneq \ldots$ such that $\{H_0, H_1, H_2, \ldots\} \subseteq \mathcal{L}$ and $L \subsetneq H_i$ for each i.*

Proof. Suppose there is some $L \in \mathcal{L}$ for which there exists an infinite descending chain $H_0 \supsetneq H_1 \supsetneq H_2 \supsetneq \ldots$ with $\{H_0, H_1, H_2, \ldots\} \subseteq \mathcal{L}$ and $L \subsetneq H_i$ for each i. Assume by way of a contradiction that $((\mathcal{L}_0, d_0), (\mathcal{L}_1, d_1), \ldots)$ were a positive teaching sequence for $\mathcal{L}$. Suppose $L \in \mathcal{L}_i$ for some i. Note that for all $j \in \{0, \ldots, i\}$, $L \subsetneq H_j$ implies that $H_j \in \mathcal{L}_{k_j}$ for some $k_j < i$. Further, for all $j \in \{0, \ldots, i-1\}$, since $H_{j+1} \subsetneq H_j$, it must hold that $k_j < k_{j+1}$. This contradicts the fact that $0 \leq k_j < i$ for all $j \in \{0, \ldots, i\}$. ∎

3 Linear Subsets of $\mathbb{N}_0$ With Constant 0

This section will analyse the class LINSET of linear subsets of $\mathbb{N}_0$ with constant 0. Even in the apparently simple one-dimensional case, the teaching complexity measures can vary quite widely across families of linear sets. Many proofs will exploit the fact that linear sets of dimension 1 are ultimately periodic, a property that has no exact analogue for linear sets of higher dimensions.

Proposition 6. *[10] Let $P \subset \mathbb{N}$ be a finite set such that $gcd(P) = 1$. Then $\mathbb{N} \setminus \langle P \rangle$ is finite. The largest number in $\mathbb{N} \setminus \langle P \rangle$ is known as the* Frobenius number *of $\langle P \rangle$.*

For any $P = \{p_1, \dots, p_k\}$ with $gcd(P) = 1$, $F(P)$ and $F(p_1, \dots, p_k)$ will denote the Frobenius number of $\langle P \rangle$. We will characterise the teaching sets of all linear sets $\langle P \rangle$ such that $gcd(P) = 1$ with respect to LINSET in terms of P and a certain (finite) subset of $\mathbb{N} \setminus \langle P \rangle$. The following variation of a notion from the theory of numerical semigroups will help to formulate the characterisation.

The *partial ordering induced by* P (modified from [10]) is defined as follows: $x \le_P y \iff \exists a \in \mathbb{N} : y - ax \in \langle P \rangle$. We write $x <_P y$ as an abbreviation of $x \le_P y \wedge x \neq y$. One has $x \in \langle P \rangle \wedge x \le_P y \Rightarrow y \in \langle P \rangle$, or equivalently, $y \notin \langle P \rangle \wedge x \le_P y \Rightarrow x \notin \langle P \rangle$.

For the rest of this section, "maximal" (resp. "minimal") always means "maximal w.r.t. $\le_P$" (resp. "minimal w.r.t. $\le_P$") unless specified otherwise. Let MAX_P be the set of maximal elements in $\mathbb{N} \setminus \langle P \rangle$ and let MIN_P denote the set of minimal elements in $\langle P \rangle \setminus \{0\}$. The following lemma collects some useful known facts. Many of these facts are proven in [10], or may be directly deduced from related results proven in [10].

Lemma 7. (I) $\mathbb{N} \setminus \langle P \rangle$ *contains an infinite ascending chain* $x_0 <_P x_1 <_P x_2 <_P \dots$ *(e.g.* $x_i = 1 + ip$ *with an arbitrary choice of* $p \in P$*) iff* $\mathrm{gcd}(P) > 1$.

(II) *If* $H \subseteq \langle P \rangle$*, then the following hold:*
  (a) $\langle H \rangle \subseteq \langle P \rangle$.
  (b) The partial ordering $\le_P$ *is a refinement of the partial ordering* $\le_H$*; that is, for any* x, y*,* $x \le_H y$ *implies* $x \le_P y$.
  (c) $MIN_P \cap \langle H \rangle \subseteq MIN_H$.

(III) *Let* $s = a_1 p_1 + \dots + a_r p_r \in \langle P \rangle$ *and let* $I = \{i \in [r] |\ a_i \neq 0\}$*. Then* $p_i \le_P s$ *for each* $i \in I$*. This implies that* $MIN_P \subseteq P$.

(IV) *If* $p, p' \in P$ *and* $p <_P p'$*, then* p' *is* superfluous*, i.e.,* $\langle P \rangle = \langle P \setminus \{p'\} \rangle$.

(V) *If* P *is independent, then* $P = MIN_P$.

For the rest of this section, it will always be assumed that P is independent. We now study teaching sets.

Lemma 8. *1. Let* T *be a teaching set for* $\langle P \rangle$ *w.r.t. LINSET. Then* $P \subseteq T^+$.
2. Let T *be a teaching set for* $\langle P \rangle$ *w.r.t.* $LINSET_k$ *for* $k = |P| + 1$*. Then, for each* $x \in \mathbb{N} \setminus \langle P \rangle$*, there exists* $y \in T^-$ *such that* $x \le_P y$.

Proof.

1. Since the labelling in T is consistent with $\langle P \rangle$, it follows that $T^+ \subseteq \langle P \rangle$ and $T^- \cap \langle P \rangle = \emptyset$. Therefore, $\langle T^+ \rangle \subseteq \langle P \rangle$ so that $\langle T^+ \rangle$ is consistent with T. Since T is a teaching set for $\langle P \rangle$, we may conclude that $\langle T^+ \rangle = \langle P \rangle$, which implies that $\le_{T^+}$ is the same partial ordering as $\le_P$. Since P (by the general convention made above) is independent, it follows that $P = \mathrm{MIN}_P = \mathrm{MIN}_{T^+} \subseteq T^+$. Thus $P \subseteq T^+$, as desired.
2. Pick an arbitrary but fixed $x \in \mathbb{N} \setminus \langle P \rangle$. We have to show that T^- contains some element y such that $x \le_P y$. Let $P' = P \cup \{x\}$. Clearly, $\langle P \rangle$ is a proper subset of $\langle P' \rangle$ and $\langle P' \rangle$ is consistent with T^+. Since T is a teaching

set for $\langle P\rangle$ w.r.t. LINSET_k and $\langle P\rangle, \langle P'\rangle \in \mathrm{LINSET}_k$, it follows that T^- contains an element $y \in \langle P'\rangle \setminus \langle P\rangle$. Thus y can be written in the form $y = ax + \sum_{p\in P} a(p)p$ for some properly chosen $a \in \mathbb{N}$ and $a(p) \in \mathbb{N}_0$. It follows that $x \leq_P y$, as desired. ▮

Corollary 9. *If* $\gcd(P) > 1$ *and* $k = 1 + |P|$, *then* $TD(\langle P\rangle, LINSET_k) = \infty$.

Proof. According to Lemma 7, $\mathbb{N} \setminus \langle P\rangle$ contains an infinite ascending chain $x_1 <_P x_2 <_P x_3 <_P \ldots$. According to the second assertion in Lemma 8, a teaching set for $\langle P\rangle$ must contain infinitely many elements of this chain. ▮

Corollary 10. *If* $\gcd(P) = 1$, *then the set* $T(P)$ *given by* $T(P)^+ = P$ *and* $T(P)^- = MAX_P$ *is the unique smallest teaching set for* $\langle P\rangle$ *w.r.t. LINSET.*

Proof. Suppose that $\langle H\rangle$ is consistent with $T(P)$. We show that $\langle H\rangle = \langle P\rangle$ (implying that $T(P)$ is a teaching set for $\langle P\rangle$). Since $P \subseteq \langle H\rangle$ (by consistency), it follows that $\langle P\rangle \subseteq \langle H\rangle$. Pick an arbitrary but fixed element x from $\mathbb{N} \setminus \langle P\rangle$. Recall that $\mathbb{N} \setminus \langle P\rangle$ is finite. Thus, by the definition of MAX_P, there must exist an element $y \in \mathrm{MAX}_P$ such that $x \leq_P y$. From $x \leq_P y$ and $\langle P\rangle \subseteq \langle H\rangle$, we may conclude that $x \leq_H y$. The number $y \in \mathrm{MAX}_P$ cannot belong to $\langle H\rangle$ (because $\langle H\rangle$ is consistent with $T(P)$). Now, $x \leq_H y$ implies that $x \notin \langle H\rangle$. Thus, $\langle H\rangle$ does not contain any element outside $\langle P\rangle$. It follows that $\langle P\rangle = \langle H\rangle$. Thus, $T(P)$ is a teaching set for $\langle P\rangle$ w.r.t. LINSET, indeed. According to Lemma 8, any other teaching set must contain $T(P)$ as a subset. ▮

Remark 11. If T is a teaching set for $L' \subseteq \mathbb{N}_0^m$ w.r.t. $\mathcal{L}$, then for any $c \in \mathbb{N}_0^m$, $c+T = \{(c+x,+) : x \in T^+\} \cup \{(c+y,-) : y \in T^-\}$ is a teaching set for $c+L'$ w.r.t. $\mathcal{L}[n] = \{c+L : L \in \mathcal{L}\}$. Thus Lemma 8 and Corollaries 9 and 10 may be readily generalised, *mutatis mutandis*, to the class $\mathrm{LINSET}[c] = \{c + L : L \in \mathrm{LINSET}\}$ for any $c \in \mathbb{N}_0$.

The proof of [8, Theorem 6] provides a construction that may be slightly modified to show that $\mathrm{TD}(\mathrm{LINSET}_1) = \infty$ even though $\mathrm{TD}(\langle q\rangle, \mathrm{LINSET}_1) < \infty$ for any $q > 0$. By the monotonicity of TD, $\mathrm{TD}(\mathrm{LINSET}) = \mathrm{TD}(\mathrm{LINSET}_k) = \infty$ for all $k > 0$.

By Corollary 9, LINSET contains infinitely many members that have an infinite TD w.r.t. LINSET. Thus for any $\mathcal{L} \subseteq \mathrm{LINSET}$, it may be difficult to interpret a value of ∞ for $\mathrm{TD}(\mathcal{L})$: (1) on the one hand, all cofinite subclasses of $\mathcal{L}$ may have an infinite TD w.r.t. $\mathcal{L}$; (2) on the other hand, there may be a cofinite subclass of $\mathcal{L}$ that has a finite TD w.r.t. $\mathcal{L}$. Intuitively, it seems that $\mathcal{L}$ in Case (2) is unteachable in a weaker sense than in Case (1), but the TD makes no such distinction. It shall be shown, however, that the RTD is a bit more well-behaved when applied to LINSET. In particular, for all $\mathcal{L} \subset \mathrm{LINSET}$, $\mathrm{RTD}(\mathcal{L}, \mathrm{LINSET})$ grows only linearly with $\sup\{\min(P) : \langle P\rangle \in \mathcal{L} \wedge \min(P) > 0\}$. We will also give a finer analysis of LINSET_k for $k \in \{1,2,3\}$, showing that while LINSET_2 does not have any positive teaching sequence, $\mathrm{RTD}(\mathrm{LINSET}_k) < \infty$ for $k \in \{1,2,3\}$. In addition, $\mathrm{RTD}(\mathrm{LINSET}_k)$ grows at least linearly in k, implying that

RTD(LINSET) $= \infty$. The question of whether RTD(LINSET_k) $< \infty$ for any $k > 3$ remains open.

First, the following proposition explains why the singleton $\{0\}$ was excluded from the definition of LINSET. The proof is quite similar to that of Proposition 21, which will be proven later.

Proposition 12. *$RTD(\{\{0\}\}, \{\{0\}\} \cup LINSET_1) = \infty$.*

Proposition 12 should be contrasted with the observation that RTD(LINSE-T$_1$) $= 1$: $((\langle 1\rangle, 1), (\langle 2\rangle, 1), \ldots)$ is a teaching plan with positive examples for LINSET$_1$, where the ith linear set in the plan is $\langle i\rangle$ and $\{(i, +)\}$ is a teaching set for $\langle i\rangle$ w.r.t. $\{\langle j\rangle : j \geq i\}$. The next theorem shows on the other hand that for any finite $\mathcal{L} \subset$ LINSET, RTD($\mathcal{L}$, LINSET) $< \infty$; in fact, for *any* $\mathcal{L} \subset$ LINSET, RTD($\mathcal{L}$, LINSET) is at most linear in $\sup\{\min(P) : \langle P\rangle \in \mathcal{L} \wedge \min(P) > 0\}$.

Theorem 13. *Let $\mathcal{F}_n = \{\langle P\rangle : P \text{ is independent } \wedge \min(P) \leq n\}$. Then $RTD(\mathcal{F}_n, LINSET) \leq 2n - 1$.*

Proof. (Sketch.) To streamline the proof, we will adopt some graph terminology. Let $\mathcal{L}$ = LINSET. Let $L \mapsto T(L)$ be a mapping that assigns a set of labelled examples to every $L \in \mathcal{L}$. Define the *digraph induced by* T as the graph $G = (V_G, A_G)$, where the nodes of G are identified with the members L of $\mathcal{L}$, i.e., $V_G = \mathcal{L}$, and a pair $(L', L) \in \mathcal{L} \times \mathcal{L}$ is included in A_G iff L' is consistent with $T(L)$. Define the *depth* of a node v in a digraph as the length of the longest path ending in v (or as ∞ if the paths ending in v can become arbitrarily long). Say that the mapping $L \mapsto T(L)$ with L ranging over all members of $\mathcal{L}$ is *RTD-admissible for* $\mathcal{L}$ if the digraph G induced by T is acyclic and every node in G has a finite depth.

We shall use the following two facts (proofs omitted due to space constraints): (1) there exists a partition of $\mathcal{L}$ into $\mathcal{L}_0, \mathcal{L}_1, \ldots$ such that, for all i and all $L \in \mathcal{L}_i$, it holds that $T(L)$ is a teaching set for L w.r.t. $\bigcup_{j \geq i} \mathcal{L}_j$ iff T is RTD-admissible; (2) if $P = \{p_1, \ldots, p_k\} \subseteq \mathbb{N}_0$ is independent, $p_1 = \min(P)$ and $d = \gcd(P)$, then $k = |P| \leq p_1$ and $|\text{MAX}_P| \leq p_1 - 1$.

Let P range over finite independent subsets of $\mathbb{N}_0$. We shall show that the mapping $\langle P\rangle \mapsto T(\langle P\rangle)$ given by $T(\langle P\rangle)^+ = P$ and $T(\langle P\rangle)^- = \text{MAX}_P$ is RTD-admissible for $\mathcal{L}$. It suffices to show that the digraph $G = (\mathcal{L}, A)$ induced by the mapping $\langle P\rangle \mapsto T(\langle P\rangle)$ is acyclic and every node $\langle P\rangle \in \mathcal{L}$ has a finite depth. Suppose that $(\langle P'\rangle, \langle P\rangle) \in A$. It follows from the construction of G that $\langle P'\rangle$ is consistent with $T(\langle P\rangle)$. The consistency with $T(\langle P\rangle)^+ = P$ implies that $d' = \gcd(P')$ is a divisor of $d = \gcd(P)$. It suffices to show that d' is a proper divisor of d since this implies that G is acyclic and that the depth of $\langle P\rangle$ is bounded by the number of prime power divisors of $d = \gcd(P)$. Suppose for sake of contradiction that $d' = d$ so that $\langle P'\rangle, \langle P\rangle \subseteq d \cdot \mathbb{N}_0$. Now we may argue as follows. Since $(\langle P'\rangle, \langle P\rangle) \in A$, the two linear sets do not coincide so that we may pick a point u from their symmetric difference. Since both linear sets are subsets of $d \cdot \mathbb{N}_0$, there exists $u' \in \mathbb{N}$ such that $u = du'$. But then u' is in the symmetric difference of $(1/d)\langle P'\rangle$ and $(1/d)\langle P\rangle$. On the other hand, since $\langle P'\rangle$ is consistent with $T(\langle P\rangle)$, it follows that $(1/d)\langle P'\rangle$ is consistent with $(1/d)P$ labelled "+" and $(1/d)\mathrm{MAX}_P = \mathrm{MAX}_{(1/d)P}$ labelled "−". According to Corollary 10, $(1/d)\cdot P$ labelled "+" and $\mathrm{MAX}_{(1/d)P}$ labelled "−" is a teaching set for $(1/d)\langle P\rangle$ w.r.t. $\mathcal{L}$, a contradiction.

Finally, observe that from Facts (1), (2), the RTD-admissibility of T for $\mathcal{L}$, and the condition that $\min(P) \leq n$ for all $\langle P\rangle \in \mathcal{F}_n$ with P independent, one can find a teaching subsequence for $\mathcal{L}$ covering $\mathcal{F}_n$ such that the order of this teaching subsequence is at most $n + (n-1) = 2n-1$. ∎

Theorem 15 (to be shown later) will imply that the order of any teaching sequence for LINSET must necessarily be infinite. Nonetheless, the preceding theorem shows roughly that the growth of $\mathrm{RTD}(\mathcal{L}, \mathrm{LINSET})$ with $\mathcal{L}$ is relatively modest if the minimum positive periods of all $L \in \mathcal{L}$ vary only linearly.

The next series of results will present a detailed study of CF–LINSET$_k$ for each k and CF–LINSET, which comprises all linear sets $\langle P\rangle$ such that P ($\neq \emptyset$) is a finite subset of $\mathbb{N}$ and and $gcd(P) = 1$. By Proposition 6, this is precisely the class of cofinite linear subsets of $\mathbb{N}_0$ with constant 0. Since LINSET$_k$ is a union of classes of linear sets, each of which is isomorphic to CF–LINSET$_k$, it is hoped that investigating the teaching complexity of CF–LINSET$_k$ may lead to some insights into the question of whether RTD(LINSET$_k$) is finite for each k. CF–LINSET is also perhaps interesting in its own right: on the one hand, the teaching dimension of CF–LINSET$_k$ is finite for $k \leq 3$ but infinite for $k \geq 5$; on the other hand, for all k, CF–LINSET$_k$ has a relatively simple teaching sequence that gives it an RTD$^+$ of k. The first result gives an almost complete analysis of TD(CF–LINSET$_k$) for all k; the case $k = 4$ is left open.

Theorem 14. (I) *TD(CF–LINSET$_1$) = 0;*
(II) *TD(CF–LINSET$_2$) = 3;*
(III) *TD(CF–LINSET$_3$) = 5;*
(IV) *for each $k \geq 5$, TD(CF–LINSET$_k$) $= \infty$.*

Proof. The proofs of Assertions (III) and (IV) are quite long and will be omitted. *Assertion (I).* Note that CF–LINSET$_1$ $= \{\langle 1\rangle\}$. The empty set is a teaching set for $\langle 1\rangle$ w.r.t. CF–LINSET$_1$.

Assertion (II). We first prove the upper bound. Note that $\mathbb{N}_0$ is the only member of CF–LINSET$_2$ that is generated by one number. $\{(1,+)\}$ is a teaching set for $\mathbb{N}_0$ w.r.t. CF–LINSET$_2$, and so $\mathrm{TD}(\mathbb{N}_0) = 1$. Now consider $L = \langle p_1, p_2\rangle$, where $gcd(p_1, p_2) = 1$. We claim that $T = \{(p_1,+), (p_2,+), (p_1p_2 - p_1 - p_2, -)\}$ is a teaching set for L w.r.t. CF–LINSET$_2$. Note that $F(p_1, p_2) = p_1p_2 - p_1 - p_2$ (see, for example, [10]), and so the labelling of T is consistent with L. For any $L' \in$ CF–LINSET$_2$ such that $\{p_1, p_2\} \subseteq L'$, it must hold that $L' \subseteq L$. Suppose further that $L' \neq L$, and take any $p_3 \in L' - L$. Then there exists some k with $0 \leq k \leq p_1 - 1$ such that $p_3 \equiv kp_2$ $(mod\ p_1)$. Since $p_3 \notin L$, this means that for some $m \geq 1$, $p_3 = kp_2 - mp_1$. As $p_1p_2 - p_1 - p_2 = kp_2 - mp_1 + (m-1)p_1 + (p_1 - k - 1)p_2 = p_3 + (m-1)p_1 + (p_1 - k - 1)p_2 \in \langle p_1, p_2, p_3\rangle \subseteq L'$, it follows that $p_1p_2 - p_1 - p_2 \in L'$, and so T cannot be consistent with L'. Hence TD(CF–LINSET$_2$) ≤ 3.

For the lower bound, choose primes p_1, p_2, p_3 such that $2 < p_1 < p_2 < p_3$. Note that $\langle 2, p_1p_2p_3\rangle \subsetneq \langle 2, p_1p_2\rangle \subsetneq \langle 2, p_1\rangle$ is a chain in CF–LINSET$_2$. Thus if T is any teaching set for $\langle 2, p_1p_2\rangle$ w.r.t. CF–LINSET$_2$ such that $|T| \leq 2$, then T must contain exactly one positive example $(x_1, +)$ and exactly one negative example $(x_2, -)$. Choose any prime $p > max(\{x_1, x_2, 2, p_1p_2\})$. Then $\langle x_1, p\rangle$ is consistent with T but $\langle 2, p_1p_2\rangle \neq \langle x_1, p\rangle$. Hence $|T| \geq 3$. ∎

Theorem 15. *For all $k \geq 1$, RTD(CF–LINSET$_k$) $\in \{k-1, k\}$ and RTD$^+$(CF–LINSET$_k$) $= k$. Moreover, RTD(CF–LINSET$_2$) $= 2$.*

Proof. (Sketch.) We prove RTD$^+$(CF–LINSET$_k$) $= k$. First, a teaching plan with positive examples for CF–LINSET$_k$ is constructed. Let $\langle P_0\rangle, \langle P_1\rangle, \ldots$ be a one-one enumeration of CF–LINSET$_k$ such that for all i, j with $i < j$, P_i is independent and $\langle P_i\rangle \not\subseteq \langle P_j\rangle$. Such an enumeration exists because for each $\langle P\rangle \in$ CF–LINSET$_k$, there are only finitely many $\langle P'\rangle \in$ CF–LINSET$_k$ such that $\langle P\rangle \subseteq \langle P'\rangle$, which implies that there are only finitely many chains $C_1, \ldots$ such that $\langle P\rangle$ is the least member (with respect to set inclusion) of each C_l, and each of these chains has finite length. Let Q be the teaching plan $((\langle P_0\rangle, S_0), (\langle P_1\rangle, S_1), (\langle P_2\rangle, S_2), \ldots)$ where, for each P_i, $S_i = \{(p, +) : p \in P_i\}$. Since $\langle P_i\rangle \not\subseteq \langle P_j\rangle$ for all $j > i$, S_i is a teaching set for $\langle P_i\rangle$ w.r.t. $\{\langle P_j\rangle : j \geq i\}$. Further, as $|S_i| \leq k$ for all i, Q is a teaching plan for CF–LINSET$_k$ of order at most k.

Now it is shown that RTD$^+$(CF–LINSET$_k$) $\geq k$. Let $P = \{k, k+1, \ldots, 2k-1\}$, and consider the class $\mathcal{C}_k = \{H \in \text{CF–LINSET}_k : H \subseteq \langle P\rangle\}$. For any positive teaching sequence Q' of $\mathcal{C}_k$, $\langle P\rangle$ must be contained in the first nonempty subclass of $\mathcal{C}_k$ removed. If $\langle P\rangle$ has a teaching set with positive examples S (w.r.t. the subclass of linear sets in $\mathcal{C}_k$ that do not occur before $\langle P\rangle$ in Q') such that $|S| \leq k-1$, then $\langle X(S)\rangle$ is a proper subset of $\langle P\rangle$ that is consistent with S; further, there exists some prime $p \notin X(S)$ such that $\langle X(S) \cup \{p\}\rangle \in$ CF–LINSET$_k$ and $\langle X(S) \cup \{p\}\rangle \subsetneq \langle P\rangle$. Hence $|S| \geq k$. This proves that RTD$^+$(CF–LINSET$_k$) $\geq$ RTD$^+(\mathcal{C}_k) \geq k$. We skip the proof that RTD(CF–LINSET$_2$) $= 2$.

The construction that witnesses RTD(CF–LINSET$_k$) $\geq k-1$ is based on a hitherto unpublished proof [4]. For each k, let $\mathcal{L}_k = \{\langle k, p_1, \ldots, p_{k-1}\rangle : \forall i \in \{1, \ldots, k-1\}[p_i \in \{k+i, 2k+i\}]\}$. One can show that RTD(CF–LINSET$_k$) $\geq$ RTD$(\mathcal{L}_k) \geq \min(\{\text{TD}(L, \mathcal{L}_k) : L \in \mathcal{L}_k\}) \geq k-1$. ∎

Corollary 16. *TD(CF–LINSET) = RTD(CF–LINSET) = RTD$^+$(CF–LINSET) = RTD(LINSET) $= \infty$.*

For each $k \in \{1, 2, 3\}$, the result on TD(CF–LINSET$_k$) may be directly applied to construct a teaching sequence of finite order for LINSET$_k$.

Theorem 17. (I) *LINSET$_2$ does not have any positive teaching sequence.*
(II) *RTD$^+$(LINSET$_1$) = RTD(LINSET$_1$) = 1, RTD(LINSET$_2$) = 3 and $3 \leq$ RTD(LINSET$_3$) ≤ 5.*

Proof. *Assertion (I).* Let $p_1, p_2, p_3, \ldots$ be a strictly increasing infinite sequence of primes with $p_1 > 2$. Note that for all j, $\langle 2\rangle \subsetneq \langle 2, p_1 \ldots p_j\rangle$. Further, $\langle 2, p_1\rangle \supsetneq$

$\langle 2, p_1p_2\rangle \supsetneq \langle 2, p_1p_2p_3\rangle \supsetneq \ldots \supsetneq \langle 2, p_1 \ldots p_j\rangle \supsetneq \langle 2, p_1 \ldots p_jp_{j+1}\rangle \supsetneq \ldots$ is an infinite descending chain in LINSET_2. Thus by Proposition 5, LINSET_2 does not have a positive teaching sequence.

Assertion (II). (Sketch.) We had described earlier (after Proposition 12) a teaching plan with positive examples of order 1 for LINSET_1. Here, a teaching sequence for LINSET_2 of order no more than 3 is given; a teaching sequence for LINSET_3 can be constructed analogously using teaching sets of size at most 5 for linear sets in CF–LINSET_3. Define the sequence $((\mathcal{L}_0, d_0), (\mathcal{L}_1, d_1), \ldots)$ where, for all $i \in \mathbb{N}_0$, $\mathcal{L}_i = \{\langle p_1, p_2\rangle : p_1, p_2 \in \mathbb{N} \wedge gcd(p_1, p_2) = i + 1\}$. Consider any $\langle p_1, p_2\rangle \in \mathcal{L}_i$. The proof of Assertion (II) in Theorem 14 gives a teaching set T of size no more than 3 for $\left\langle \frac{p_1}{i+1}, \frac{p_2}{i+1} \right\rangle$ w.r.t. CF–LINSET_2. Now let $T' = \{((i+1)x, +) : x \in T^+\} \cup \{((i+1)y, -) : y \in T^-\}$. Note that since T' contains the two positive examples $(p_1, +), (p_2, +)$ and $gcd(p_1, p_2) = i + 1$, no linear set $\langle P\rangle \in \text{LINSET}_2$ with $gcd(P) > i + 1$ can be consistent with T'. Moreover, T' is a teaching set for $\langle p_1, p_2\rangle$ w.r.t. all $\langle P\rangle \in \text{LINSET}_2$ with $gcd(P) = i + 1$. Hence $d_i \leq 3$ for all $i \in \mathbb{N}_0$. We omit the proof that $\text{RTD}(\text{LINSET}_2) \geq 3$. ∎

4 Linear Subsets of $\mathbb{N}_0$ With Bounded Period Sums

The present section will examine a special family of linear subsets of $\mathbb{N}_0$ that arises from studying an invariant property of the class of non-erasing pattern languages over varying unary alphabets. Recall that the *commutative* image, or *Parikh* image, of $w \in \{a\}^*$ is the number of times that a appears in w, or the length of w. Thus the commutative image of the language generated by a non-erasing pattern $a^{k_0}x_1^{k_1} \ldots x_n^{k_n}$ is the linear subset $(k_0+k_1+\ldots+k_n)+\langle k_1, \ldots, k_n\rangle$ of $\mathbb{N}_0$, which is in NE–LINSET. Conversely, any $L \in$ NE–LINSET is the commutative image of a non-erasing pattern language. This gives a one-to-one correspondence between the class of non-erasing pattern languages and the class NE–LINSET, so that the two classes have equivalent teachability properties. The following theorem gives the exact value of $\text{RTD}^+(\text{NE–LINSET}_k)$.

Theorem 18. *$RTD^+(NE\text{–}LINSET_k) = k + 1$.*

Proof. (Sketch.) Note that as NE–LINSET_k has finite thickness and $\emptyset \notin$ NE–LINSET_k, Proposition 4 implies that NE–LINSET_k has a teaching plan with positive examples. $\text{RTD}^+(\text{NE–LINSET}_k) \leq k + 1$ is shown. A teaching plan Q for NE–LINSET_k is built in stages as follows. Let Q_c denote the segment of Q that has been defined up to stage c and let A_c denote the class of all $L \in$ NE–LINSET_k such that Q_c does not contain L. It is assumed inductively that A_c does not contain any linear subset of $\mathbb{N}_0$ with constant less than c. The idea of the construction is to design a teaching plan for the finite subclass $P_c^0 = \{c + \langle p_1, \ldots, p_k\rangle : p_1 + \ldots + p_k \leq c\}$ at stage c.

Inductively, assume that P_c^g has been defined for some $g \geq 0$. If $P_c^g = \emptyset$, then the teaching plan for P_c^0 is complete. Otherwise, suppose that P_c^g is nonempty. Choose $c+\langle p_1, \ldots, p_k\rangle \in P_c^g$, the next linear set to be taught in the teaching plan for P_c^0, so that $c+\langle p_1,\dots,p_k\rangle$ is maximal in P_c^g with respect to the subset inclusion relation. Set $P_c^{g+1} = P_c^g \setminus \{c+\langle p_1,\dots,p_k\rangle\}$. We claim that $S = \{(c,+)\} \cup \{(c+p_i,+) : 1 \le i \le k\}$ is a teaching set for $c+\langle p_1,\dots,p_k\rangle$ w.r.t. $P_c^g \cup (A_c \setminus P_c^0)$. Assume that $\{c\} \cup \{c+a_i : 1 \le i \le k\} \subseteq c' + \langle b_1,\dots,b_l\rangle \in P_c^g \cup (A_c \setminus P_c^0)$. Since A_c does not contain any linear subset of $\mathbb{N}_0$ with constant less than c, $c' = c$ and so $c' + \langle b_1,\dots,b_l\rangle \in P_c^g$. Then for all $i \in \{1,\dots,k\}$, $a_i = \sum_{i=1}^{l} q_i b_i$ for some non-negative integers $q_1,\dots,q_l$, and therefore $c+\langle a_1,\dots,a_k\rangle \subseteq c+\langle b_1,\dots,b_l\rangle$. By the maximality of $c+\langle a_1,\dots,a_k\rangle$, one has $c+\langle a_1,\dots,a_k\rangle = c+\langle b_1,\dots,b_l\rangle$. The construction continues until a stage g' is reached where $P_c^{g'} = \emptyset$. Q_{c+1} is then defined as the concatenation of Q_c and the teaching plan for P_c^0 (with Q_c as the prefix). We omit the somewhat long proof that $\mathrm{RTD}^+(\text{NE–LINSET}_k) \ge k+1$. ∎

The lower bound on $\mathrm{RTD}(\text{NE–LINSET}_k)$ in the following theorem may be obtained by adapting the proof of the corresponding result for CF–LINSET_k; the proof of [8, Theorem 6] immediately implies that $\mathrm{TD}(\text{NE–LINSET}_k) = \infty$.

Theorem 19. *For all $k \ge 1$, $k-1 \le RTD(NE\text{–}LINSET_k) \le k+1$ and $TD(NE\text{–}LINSET_k) = \infty$.*

Remark 20. Our results on NE–LINSET may be generalised to classes of linear subsets of $\mathbb{N}_0^m$ for any $m > 1$ in the following way. For each m, define

(i) $\text{NE–LINSET}_k^m := \{c+\langle P\rangle : c \in \mathbb{N}_0^m \wedge P \subset \mathbb{N}_0^m \wedge |P| \le k \wedge \|\sum_{p\in P} p\|_1 \le \|c\|_1\}$.
(ii) $\text{NE–LINSET}^m := \bigcup_{k\in\mathbb{N}} \text{NE–LINSET}_k^m$.

Note that $\text{NE–LINSET}_k^1 = \text{NE–LINSET}_k$ and $\text{NE–LINSET}^1 = \text{NE–LINSET}$. Then one has $\mathrm{RTD}^+(\text{NE–LINSET}_k^m) = \mathrm{RTD}^+(\text{NE–LINSET}_k) = k+1$, $k-1 \le \mathrm{RTD}(\text{NE–LINSET}_k^1) \le \mathrm{RTD}(\text{NE–LINSET}_k^m)$ and $\mathrm{RTD}(\text{NE–LINSET}^m) = \mathrm{RTD}(\text{NE–LINSET}) = \mathrm{RTD}^+(\text{NE–LINSET}^m) = \mathrm{RTD}^+(\text{NE–LINSET}) = \infty$.

5 Linear Subsets of $\mathbb{N}_0^2$ With Constant 0

Finally, we consider how our preceding results may be extended to general classes of linear subsets of higher dimensions. Finding teaching sequences for families of linear sets with dimension $m > 1$ seems to present a new set of challenges, as many of the proof methods for the case $m = 1$ do not carry over directly to the higher dimensional cases. The classes of linear subsets of $\mathbb{N}_0^2$ briefly studied in this section are denoted as follows. In the first definition, $k \in \mathbb{N}$.

(i) $\text{LINSET}_k^2 := \{\langle P\rangle : P \subset \mathbb{N}_0^2 \wedge \exists p \in P[p \ne 0] \wedge |P| \le k\}$.
(ii) $\text{LINSET}_{=2}^2 := \text{LINSET}_2^2 \setminus \text{LINSET}_1^2$.

The following result suggests that to identify interesting classes of linear subsets of $\mathbb{N}_0^m$ for $m > 1$ that have finite teaching complexity measures, it might be a good idea to first exclude certain linear sets.

Proposition 21. *$RTD(\{\langle(0,1)\rangle\}, LINSET_2^2) = \infty$.*

Proof. Assume that $R = ((\mathcal{L}_0, d_0), (\mathcal{L}_1, d_1), \ldots)$ were a teaching subsequence for LINSET_2^2 covering $\{\langle(0,1)\rangle\}$. Suppose that $\langle(0,1)\rangle \in \mathcal{L}_i$ and T were a teaching set for $\langle(0,1)\rangle$ w.r.t. $\mathrm{LINSET}_2^2 \setminus \bigcup_{j<i} \mathcal{L}_j$. Choose any $N > \max(\{d_j : j < i\})$ such that N is larger than every component of any instance $(a,b) \in \mathbb{N}_0^2$ in T. Further, let $p_0, \ldots, p_{N+i}$ be a strictly increasing sequence of primes. Observe that by the choice of N, $\langle(0,1), p_0 \ldots p_{N+i}(1,1)\rangle$ is consistent with T. Hence this linear set occurs in some $\mathcal{L}_{j_0}$ with $j_0 < i$. In addition, since, for any two distinct $(N+i)$-subsets S, S' of $\{p_0, \ldots, p_{N+i}\}$, $\langle(0,1), p_0 \ldots p_{N+i}(1,1)\rangle \subsetneq \langle(0,1), \prod_{x\in S} x(1,1)\rangle$ and $\langle(0,1), \prod_{x\in S} x(1,1)\rangle \cap \langle(0,1), \prod_{x\in S'} x(1,1)\rangle \subseteq \langle(0,1), p_0 \ldots p_{N+i}(1,1)\rangle$, the choice of N again gives that for some $(N+i)$-subset S_1 of $\{p_0, \ldots, p_{N+i}\}$, $\langle(0,1), \prod_{x\in S_1} x(1,1)\rangle \in \mathcal{L}_{j_1}$ for some $j_1 < j_0$. The preceding line of argument can be applied again to show that for some $(N+i-1)$-subset S_2 of S_1, $\langle(0,1), \prod_{x\in S_2} x(1,1)\rangle \in \mathcal{L}_{j_2}$ for some $j_2 < j_1$. Repeating the argument successively thus yields a chain $S_1 \supsetneq S_2 \supsetneq \ldots \supsetneq S_i$ of subsets of $\{p_0, \ldots, p_{N+i}\}$ such that $\langle(0,1), \prod_{x\in S_l} x(1,1)\rangle \in \mathcal{L}_{j_l}$ for all $l \in \{1, \ldots, i\}$, where $j_i < \ldots < j_1 < j_0 < i$, which is impossible as $j_i \geq 0$. Hence there is no teaching subsequence of LINSET_2^2 covering $\{\langle(0,1)\rangle\}$. ∎

One can define quite a meaningful subclass of LINSET_2^2 that does have a finite RTD. $\mathrm{LINSET}_{=2}^2$ consists of all linear sets in LINSET_2^2 that are *strictly 2-generated*. Examples of strictly 2-generated linear subsets include $\langle(1,0),(0,1)\rangle$ and $\langle(4,6),(6,9)\rangle$. $\langle(0,1)\rangle$ is not a strictly 2-generated linear subset.

Theorem 22. (I) *$TD(LINSET_{=2}^2) = \infty$.*
(II) *$LINSET_{=2}^2$ does not have any positive teaching sequence.*
(III) *$RTD(LINSET_{=2}^2) \in \{3,4\}$.*

Proof. *Assertion (I).* Observe from the proof of Proposition 21 that for any N distinct primes $p_0, p_1, \ldots, p_{N-1}$, $\mathrm{TD}(\langle(0,1), p_0p_1 \ldots p_{N-1}(1,0)\rangle, \mathrm{LINSET}_{=2}^2) \geq N$. Hence $\mathrm{TD}(\mathcal{L}, \mathrm{LINSET}_{=2}^2) = \infty$ for any cofinite subclass $\mathcal{L}$ of $\mathrm{LINSET}_{=2}^2$.

Assertion (II). Let $p_1, p_2, p_3, \ldots$ be a strictly increasing infinite sequence of primes. Note that for all j, $\langle(2,0),(3,0)\rangle \subsetneq \langle(1,0), p_1 \ldots p_j(0,1)\rangle$. Further, $\langle(1,0), p_1(0,1)\rangle \supsetneq \langle(1,0), p_1p_2(0,1)\rangle \supsetneq \langle(1,0), p_1p_2p_3(0,1)\rangle \supsetneq \ldots \supsetneq \langle(1,0), p_1 \ldots p_j(0,1)\rangle \supsetneq \langle(1,0), p_1 \ldots p_jp_{j+1}(0,1)\rangle \supsetneq \ldots$ is an infinite descending chain in $\mathrm{LINSET}_{=2}^2$. Thus by Proposition 5, $\mathrm{LINSET}_{=2}^2$ does not have a positive teaching sequence.

Assertion (III). (Sketch.) The main idea is that for each strictly 2-generated linear subset S of $\mathbb{N}_0^2$ with canonical representation $(0,P)$, if M denotes the class of all $S' \in \mathrm{LINSET}_{=2}^2$ for which each $S' \in M$ with canonical representation $(0,P')$ satisfies $\|\sum_{p'\in P'} p'\|_1 \geq \|\sum_{p\in P} p\|_1$, then $\mathrm{TD}(S,M) \leq 4$. The sequence $((\mathcal{L}_0, d_0), (\mathcal{L}_1, d_1), \ldots)$ defined by $\mathcal{L}_i = \{\langle u_1, u_2\rangle : \|u_1 + u_2\|_1 = i+2\}$ would then be a teaching sequence for $\mathrm{LINSET}_{=2}^2$ of order at most 4. To prove this assertion, it suffices to find a teaching set of size at most 4 for any $\langle u_1, u_2\rangle$ w.r.t. the class of all $S' \in \mathrm{LINSET}_{=2}^2$ such that if S' has the canonical representation $(0,P')$, then $\|\sum_{p'\in P'} p'\|_1 \geq \|u_1 + u_2\|_1$.

Owing to space constraints, we will only give a proof for the case when $\{u_1, u_2\}$ is linearly independent. For a given linear set L with canonical repre-

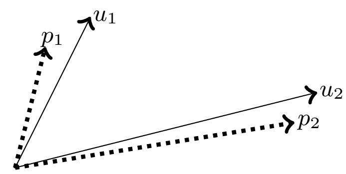

Fig. 1. p_1 and p_2 (not drawn to scale)

sentation (c, P), call each $p \in P$ a *minimal period of* L. Assume that u_1 lies to the left of u_2. Consider the set A of linear sets L in M such that $\langle u_1, u_2 \rangle \subsetneq L$. Since no single vector in $\mathbb{N}_0^2$ can generate two linearly independent vectors in $\mathbb{N}_0^2$, each $L \in A$ must have two linearly independent periods p_1 and p_2, neither of which lies strictly between u_1 and u_2; in addition, $\max(\{\|p_1\|_1, \|p_2\|_1\}) \leq \max(\{\|u_1\|_1, \|u_2\|_1\})$. Thus A is finite. Furthermore, for each $L \in A$ with canonical representation $(0, P')$, at least one of the periods in P' is not parallel to u_1 and also not parallel to u_2, for otherwise $\|\sum_{p' \in P'} p'\|_1 < \|u_1 + u_2\|_1$. If $A = \emptyset$, then $\{(u_1, +), (u_2, +)\}$ is a teaching set for $\langle u_1, u_2 \rangle$ w.r.t. M. Assume that $A \neq \emptyset$. Consider the set $Q = \bigcup_{L \in A} \{w : w$ is a minimal period of L not parallel to u_1 and not parallel to $u_2\}$. Choose some p_1 among the periods in Q that are closest to u_1 to the left of u_1, and choose p_2 so that p_2 is among the periods in Q that are closest to u_2 to the right of u_2 (see Figure 1); note that at least one of p_1, p_2 exists. For every $L \in A$ with canonical representation $(0, \{v_1, v_2\})$, at least one of p_1 and p_2 lies between (not necessarily strictly) v_1 and v_2, and $\{kp_1, k'p_2\} \cap \langle u_1, u_2 \rangle = \emptyset$ for all $k, k' \in \mathbb{N}$. Thus there is a sufficiently large $K \in \mathbb{N}$ such that for all $L \in A$, either $Kp_1 \in L \setminus \langle u_1, u_2 \rangle$ or $Kp_2 \in L \setminus \langle u_1, u_2 \rangle$. Therefore a teaching set for $\langle u_1, u_2 \rangle$ w.r.t. M is $\{(u_1, +), (u_2, +), (Kp_1, -), (Kp_2, -)\}$. If p_i does not exist for exactly one i, then remove $(Kp_i, -)$ from this teaching set. ∎

6 Conclusion

We have studied two main teaching parameters, the TD and RTD (and its variant RTD$^+$), of classes of linear sets with a fixed dimension. Notice that in Table 1, even though all the classes have an infinite TD, there are finer notions of teachability that occasionally yield different finite sample complexity measures. In particular, there are families of linear sets that have an infinite TD and RTD$^+$ and yet have a finite RTD. We broadly interpret a class that has an infinite RTD as being "unteachable" in a stronger sense than merely having an infinite TD. Quite interestingly, the fact that some classes in Table 1 have an infinite RTD contrasts with Takada's [12] theorem that the family of linear subsets of $\mathbb{N}_0^m$ is learnable in the limit from just positive examples. One possible interpretation of this contrast is that classes of linear sets may be generally harder to teach than to learn. Further, a number of quantitative problems remain open. For example, we did not solve the question of whether RTD(LINSET$_k$) is finite for each $k > 3$. A more precise analysis of the values of RTD for various families of linear sets studied in the present paper (see Table 1) would also be desirable.

Class	TD	RTD	RTD^+
$\mathrm{CF–LINSET}_k, k \geq 5$	∞ (Thm 14(IV))	$\mathrm{RTD} \in \{k-1, k\}$ (Thm 15)	k (Thm 15)
LINSET_1	∞ (Rem 11)	1 (Thm 17(II))	1 (Thm 17(II))
LINSET_2	∞ (Rem 11)	3 (Thm 17(II))	∞ (Thm 17(I))
LINSET_3	∞ (Rem 11)	$\mathrm{RTD} \in \{3, 4, 5\}$ (Thm 17(II))	∞ (Thm 17(I))
LINSET	∞ (Rem 11)	∞ (Cor 16)	∞ (Thm 17(I))
$\mathrm{NE–LINSET}_k^m, m, k \geq 1$	∞ (Rem 20)	$\mathrm{RTD} \in \{k-1, k, k+1\}$ (Rem 20)	$k+1$ (Rem 20)
$\mathrm{NE–LINSET}^m, m \geq 1$	∞ (Rem 20)	∞ (Rem 20)	∞ (Rem 20)
$\mathrm{LINSET}_{=2}^2$	∞ (Thm 22)	$\mathrm{RTD} \in \{3, 4\}$ (Thm 22)	∞ (Thm 22)

Table 1. Partial summary of results.

Acknowledgements. We thank the referees of ALT 2015 for their critical reading of the manuscript; special thanks go to one referee for pointing out an error in the original version of Proposition 12 and for helping to simplify our definition of a teaching set. Sandra Zilles was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Naoki Abe. Polynomial learnability of semilinear sets. In: Computational Learning Theory (COLT), 25–40, 1989.
2. Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2): 117–135, 1980.
3. David Doty, Matthew J. Patitz and Scott M. Summers. Limitations of self-assembly at temperature 1. *Theoretical Computer Science*, 412(1–2): 145–158, 2011.
4. Ziyuan Gao, Zeinab Mazadi, Regan Meloche, Hans Ulrich Simon and Sandra Zilles. Distinguishing pattern languages with membership examples. Manuscript, 2014.
5. Sally A. Goldman and Michael J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1): 20–31, 1995.
6. Oscar H. Ibarra. Reversal-bounded multicounter machines and their decision problems. *Journal of the Association for Computing Machinery*, 25(1): 116–133, 1978.
7. Oscar H. Ibarra, Zhe Dang and Omer Egecioglu. Catalytic P systems, semilinear sets, and vector addition systems. *Theoretical Computer Science*, 312(2–3): 379–399, 2004.
8. Zeinab Mazadi, Ziyuan Gao and Sandra Zilles. Distinguishing pattern languages with membership examples. LATA 2014, Proceedings. Springer 2014 Lecture Notes in Computer Science Proceedings: 528–540, 2014.
9. Rohit J. Parikh. On context-free languages. *Journal of the Association for Computing Machinery*, 13(4): 570–581, 1966.
10. J.C. Rosales, P. A. García-Sánchez. *Numerical semigroups.* New York: Springer, 2009.
11. Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Generation Computing*, 8(4): 337–347, 1991.
12. Yuji Takada. Learning semilinear sets from examples and via queries. *Theoretical Computer Science*, 104(2): 207–233, 1992.
13. Sandra Zilles, Steffen Lange, Robert Holte and Martin Zinkevich. Models of cooperative teaching and learning. *Journal of Machine Learning Research*, 12: 349–384, 2011.