Barrier Certificates for Assured Machine Teaching

Mohamadreza Ahmadi[1], Bo Wu[2], Yuxin Chen[1], Yisong Yue[1], and Ufuk Topcu[2]

Abstract— **Machine teaching can be viewed as optimal control for learning. Given a learner's model, machine teaching aims to determine the optimal training data to steer the learner towards a target hypothesis. In this paper, we are interested in providing assurances for machine teaching algorithms using control theory. In particular, we study a well-established learner's model in the machine teaching literature that is captured by the local preference over a version space. We interpret the problem of teaching a preference-based learner as solving a partially observable Markov decision process (POMDP). We then show that the POMDP formulation can be cast as a special hybrid system, i.e., a discrete-time switched system. Subsequently, we use barrier certificates to verify set-theoretic properties of this special hybrid system. We show how the computation of the barrier certificate can be decomposed and numerically implemented as the solution to a sum-of-squares (SOS) program. For illustration, we show how the proposed framework based on control theory can be used to verify the teaching performance of two well-known machine teaching methods.**

I. Introduction

From an optimal control perspective, a learning system (e.g., a machine learning algorithm, or a human learner) defines a dynamical system where the state (i.e., learner's hypothesis) is driven by training data [1]. In this respect, machine teaching, i.e., the algorithmic framework of designing an optimal training set for learning a target hypothesis, can be viewed as optimal control for learning [2]. In a typical setting of machine teaching, the target hypothesis is *given* to the algorithm, and the goal of the teacher (machine) is to generate a minimal sequence of training examples such that the target hypothesis can be learned by a learner (human or another machine) from a finite set of hypotheses.

One popular learner's model studied in the machine teaching literature is the version space learner. In such settings, the learner maintains a subset of hypotheses that are consistent with the examples received from a teacher, and outputs a hypothesis from this subset. Based on different assumptions on the learner's behavior, multiple variants of the version space learner model has been studied in algorithmic machine teaching, leading to different notions of teaching complexity: For instance, (i) the "worst-case" model [3] assumes that the learner's behavior is completely unpredictable, and (ii) the "preference-based" model [4] assumes that she has a global preference over the hypotheses. These models are typically studied under the batch setting, where the teacher constructs a set of examples and provides them to the learner at once. Recently, [5] considered the *state-dependent* preference-based model, which generalizes the preference-based model of [4] to the adaptive setting. The state-dependent preference-based model assumes that the learner's choice of next hypothesis depends on some local preferences defined by the learner's state (i.e., the current hypothesis). In the sequential machine teaching setting, the teacher, after showing each example, obtains feedback about the hypothesis that the learner is currently entertaining; such feedback is further utilized to guide the selection of future teaching examples.

In this paper, we use notions from hybrid systems analysis framework to study the state-dependent preference-based machine teaching model with the aim of verifying whether a given machine teaching method has assured teaching performance. We first show that state-dependent preference-based machine teaching model can be represented by a POMDP. Once this POMDP is formulated, we show that the evolution of the *beliefs* over the states of this POMDP can be described by a discrete-time switched system (also see [6], [7]). We use barrier certificates to verify whether the beliefs of this POMDP belong to some subset of the reachable belief space, which, in turn, corresponds to the probability of teaching of a hypothesis. From a computational standpoint, we show that these barrier certificates can be decomposed and constructed using SOS programming. We demonstrate the efficacy of our proposed methodology by comparing and analyzing two machine teaching methods.

The rest of this paper is organized as follows. We describe the state-dependent teaching model in the next section. In Section III, we propose a POMDP representation for machine teaching. In Section IV, we briefly discuss a hybrid system that describe the evolution of this POMDP. In Section V, we formulate a set of conditions based on barrier certificates for verifying the teaching performance and show how the calculations can be decomposed. In Section VI, we propose a computational approach using SOS programming to find the barrier certificates. We elucidate the proposed method with an example in Section VII and conclude the paper in Section VIII.

Notation: $\mathbb{R}$ and $\mathbb{N}$ denote the sets of real numbers and non-negative integers $\{0,1,2,\ldots\}$, respectively. $\mathbb{N}_{\geq l}$, with $l \in \mathbb{N}$, denotes $\{l, l+1, l+2, \ldots\}$. $\mathcal{R}[x]$ accounts for the set of polynomial functions with real coefficients in $z \in \mathbb{R}^n$, $p : \mathbb{R}^n \to \mathbb{R}$ and $\Sigma \subset \mathcal{R}$ is the subset of polynomials with an SOS decomposition; i.e., $p \in \Sigma[x]$ if and only if there are $p_i \in \mathcal{R}[x],\ i \in \{1, \ldots, k\}$ such that $p = p_i^2 + \cdots + p_k^2$.

This work was supported by AFOSR FA9550-19-1-0005, DARPA D19AP00004, NSF 1646522 and NSF 1652113.

[2]Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Peter O'Donnel Jr. Building, 201 E 24th St. Austin, TX 78712, USA. {bwu3,utopcu}@utexas.edu

[1] California Institute of Technology 1200 E. California Blvd., MC 104-44, Pasadena, CA 91125, USA. {mrahmadi,chenyux,yyue}@caltech.edu

II. The State-Dependent Teaching Model

We now state the adaptive machine teaching protocol, and describe the state-dependent learner's model of [5].

1) The Teaching Domain: Let $\mathcal{X}$ denote a ground set of unlabeled examples, and the set $\mathcal{Y}$ denotes the possible labels that could be assigned to elements of $\mathcal{X}$. We denote by $\mathcal{H}$ a finite class of hypotheses, each element $h \in \mathcal{H}$ is a function $h : \mathcal{X} \to \mathcal{Y}$. In our model, $\mathcal{X}$, $\mathcal{H}$, and $\mathcal{Y}$ are known to both the teacher and the learner. There is a target hypothesis $h^* \in \mathcal{H}$ that is known to the teacher, but not the learner. Let $\mathcal{Z} \subseteq \mathcal{X} \times \mathcal{Y}$ be the ground set of labeled examples. Each element $z = (x_z, y_z^*) \in \mathcal{Z}$ represents a labeled example, where the label is given by the target hypothesis h^*, i.e., $y_z^* = h^*(x_z)$. Here, we define the notion of *version space* needed to formalize our model of the learner. Given a set of labeled examples $Z \subseteq \mathcal{Z}$, the version space induced by Z is the subset of hypotheses $\mathcal{H}(Z) \in \mathcal{H}$ that are consistent with labels of all the examples, i.e., $\mathcal{H}(Z) := \{h : h \in \mathcal{H} \texttt{ and } \forall (x, y) \in Z, h(x) = y\}$.

2) State-dependent Preference-based Model: The preference function encodes the learner's preferences of transitioning to a hypothesis. Consider that the learner's current hypothesis is h, and there are two hypotheses h', h'' that they could possibly pick as the next hypothesis. We define the preference function as $\sigma : \mathcal{H} \times \mathcal{H} \to \mathbb{R}_+$. Given current hypothesis h and any two hypothesis h', h'', we say that h' is preferred to h'' from h, iff $\sigma(h'; h) < \sigma(h''; h)$. If $\sigma(h'; h) = \sigma(h''; h)$, then the learner could pick either one of these two.

The learner starts with an initial hypothesis $h^0 \in \mathcal{H}$ before receiving any labeled examples from the teacher. Then, the interaction between the teacher and the learner proceeds in discrete time steps (trials). At any trial t, let us denote the labeled examples received by the learner up to (but not including) time step t via a set Z^t, the learner's version space as $\mathcal{H}^t = \mathcal{H}(Z^t)$, and the current hypothesis as h^t. At trial t, we model the learning dynamics as follows:

1) the learner receives a new labeled example, and
2) the learner updates the version space $\mathcal{H}^{t+1}$, and picks the next hypothesis based on the current hypothesis h^t, version space $\mathcal{H}^{t+1}$, and the preference function σ:

$$h^{t+1} \in \{h \in \mathcal{H}^{t+1} : \sigma(h; h^t) = \min_{h' \in \mathcal{H}^{t+1}} \sigma(h'; h^t)\}. \quad (1)$$

3) The Teaching Protocol and Objective: The teacher's goal is to steer the learner towards the target hypothesis h^* by providing a sequence of labeled examples. At trial t, we consider the following teaching protocol:

1) the teacher selects an unlabeled example $x^t \in \mathcal{X}$ and presents it to the learner;
2) the learner makes a guess of the label, i.e. $y^t := h^t(x^t)$.
3) the teacher receives feedback from the learner[1] and provides the true label $h^*(x^t)$;
4) the learner transitions from the current h^t to the next hypothesis h^{t+1} as per the model described in the previous subsection.
5) Teaching finishes if the learner's updated hypothesis $h^{t+1} = h^*$.

The goal of teaching algorithms is to achieve this goal in the minimal number of time steps.

The state-dependent teaching model is also found to be consistent with simple human learning models in cognitive science, including the "win-stay lose-shift" model [8], [9] (e.g., when $\sigma(h'; h) = 0$ if $h = h'$ and 1 otherwise, the learner prefers to stay at the same hypothesis if it is consistent with the observed data).

III. POMDP Model for Machine Teaching

Given the state-dependent teaching model as described in Section II, we can represent machine teaching as a sequential decision making under uncertainty scenario. To this end, we propose a POMDP representation for the learner based on the state-dependent teaching model. The POMDP model can be described as follows.

Definition 1 (Learning POMDP): The learning POMDP $\mathcal{P}_L$ is a tuple $(\mathcal{H}, p_0, \mathcal{Z}, T, \mathcal{Y}, O)$

- the hypotheses set $\mathcal{H}$ is a finite set of hidden states;
- p_0 is the probability of having an initial hypothesis $h_0 \in \mathcal{H}$;
- the set of labelled examples $\mathcal{Z}$ constitute the finite set of actions;
- T describes the transitions from one hypothesis (state) to another characterized by the preference functions as given by (1);
- $\mathcal{Y}$ denotes the set of observations made by the teacher.
- $O(y_t \mid h_t, z_t)$ is determined by the current hypothesis function.

Here, the observation model $O(y_t \mid h_t, z_t)$ defines how the version space gets updated. When referring to the "version space" learners, we are implicitly considering the "noise-free" setting, i.e., all consistent hypotheses are uniformly distributed, or equivalently, $O(y_t \mid h_t, z_t)$ is binary. Moreover, according to (1), the transition function $T(h, z_{t-1}, h')$ defines a *uniform* distribution: the learner only goes to the hypotheses h' that are the most preferred; hence, T induces a uniform distribution over the most preferred hypothesis according to the preference function σ.

The learner starts with an initial hypothesis h_0 and over a sequence of trials, in which an example $z_t \in \mathcal{Z}$ is shown and the learner receives a corresponding observation $y_t \in \mathcal{Y}$, develops a belief in the new hypothesis h. Then, the hypothesis belief evolves according to

$$b_t(h') = \frac{O(y_t \mid h', z_{t-1}) \sum_{h \in \mathcal{H}} T(h, z_{t-1}, h') b_{t-1}(h)}{\sum_{h' \in \mathcal{H}} O(y_t \mid h', z_{t-1}) \sum_{h \in \mathcal{H}} T(h, z_{t-1}, h') b_{t-1}(h)}, \quad (2)$$

The objective of a teaching policy is then to assure that the learner learns the target hypothesis $h^* \in \mathcal{H}$ in t^* number of trials. That is,

$$b_{t^*}(h^*) \geq \lambda, \quad (3)$$

[1] We consider two variants of the learner feedback: (a) the teacher indirectly observes the learner's hypothesis h^t via label y^t; (b) the teacher directly observes the learner's current hypothesis h^t. Our analysis in the subsequent sections applies to both scenarios. For discussion simplicity we focus on the more general setting (a) in Section III-VI.

where we refer to $0 < \lambda \leq 1$ as the *teaching performance*. In addition, given a teaching policy, we are often interested in finding the minimum number of trials such that the learner learns a target hypothesis, i.e.,

$$\min\ t^* \text{ subject to } b_{t^*}(h^*) \geq \lambda. \tag{4}$$

Ideally, given a pre-specified number of trials t^*, a teaching algorithm is *perfect*, if $\lambda = 1$, i.e., the probability of learning the target hypothesis after t^* number of examples is one. However, achieving a perfect teaching algorithm in t^* number of trials may not be realistic. In practice, it is desirable that we teach the target hypothesis with teaching performance $\lambda \geq 0.75$.

IV. Belief Evolution as a Hybrid System

Checking whether (3) holds by solving the learning POMDP directly is a PSPACE-hard problem [10]. In this section, we show that the learning POMDP can be represented as a special hybrid system [11], specifically, a discrete-time switched system [12], [13], [14].

The belief update equation (2) can be characterized as a discrete-time switched system, where the actions $a \in A$ define the switching modes. Formally, the hypothesis belief *dynamics* (2) can be described as

$$b_t = f_z\left(b_{t-1}, y_t\right), \tag{5}$$

where b denote the belief vector belonging to the belief unit simplex $\mathcal{B}$ and $b_0 = p_0$. In (5), $z \in \mathcal{Z}$ denote the examples that can be interpreted as the indices for the switching modes, $y \in \mathcal{Y}$ are the observations representing inputs, and $t \in \mathbb{N}_{\geq 1}$ denote the discrete time instances. The (rational) vector fields $\{f_z\}_{z \in \mathcal{Z}}$ with $f_z : [0,1]^{|\mathcal{Z}|} \times \mathcal{Y} \to [0,1]^{|\mathcal{Z}|}$ are described as the vectors with rows

$$f_z^{h'}(b, y) = \frac{O(y \mid h', z) \sum_{h \in \mathcal{H}} T(h, z, h') b_{t-1}(h)}{\sum_{h' \in \mathcal{H}} O(y \mid h', z) \sum_{h \in \mathcal{H}} T(h, z, h') b_{t-1}(h)},$$

where $f_z^{h'}$ denotes the h'th row of f_z.

We consider two classes of problems in learning POMDP verification:

1. *Arbitrary-Policy Verification*: This case corresponds to analyzing (5) under *arbitrary switching* with switching modes determined by the examples $z \in \mathcal{Z}$.
2. *Fixed-Policy Verification*: This corresponds to analyzing (5) under *state-dependent switching*. In fact, a teaching policy $\pi : \mathcal{B} \to \mathcal{Z}$ (a mapping from the hypothesis beliefs into examples) determines regions in the belief space where each mode (example) is active.

Both cases of switched systems with *arbitrary switching* and *state-dependent switching* are well-known in the systems and controls literature (see [15], [16] and references therein).

V. Verifying Teaching Performance Using Barrier Certificates

In the following, we describe a method based on barrier certificates to verify the teaching performance as given by (3). We then focus on the two cases of arbitrary policy verification and fixed-policy verification. We further show that in both cases, the calculation of the barrier certificates can be decomposed.

In order to check the teaching performance, we consider following teaching-failure set

$$\mathcal{B}_f = \{b \in \mathcal{B} \mid b_{t^*}(h^*) < \lambda\}, \tag{6}$$

which is the complement of (3).

We have the following result.

Theorem 1: Given the learning POMDP $(\mathcal{H}, p_0, \mathcal{Z}, T, \mathcal{Y}, O)$, a target hypothesis $h^* \in \mathcal{H}$, and a teaching performance λ, and a pre-set number of trials t^*, if there exists a function $B : \mathbb{N} \times \mathcal{B} \to \mathbb{R}$ called the barrier certificate such that

$$B(t^*, b_{t^*}) > 0, \quad \forall b_{t^*} \in \mathcal{B}_f, \tag{7}$$

with $\mathcal{B}_f$ as described in (6),

$$B(0, b_0) < 0, \quad \text{for} \quad b_0 = p_0, \tag{8}$$

and

$$B\left(t, f_z(b_{t-1}, y)\right) - B(t-1, b_{t-1}) \leq 0, \\ \forall t \in \{1, 2, \ldots, t^*\},\ \forall z \in \mathcal{Z},\ \forall y \in \mathcal{Y},\ \forall b \in \mathcal{B}, \tag{9}$$

then there the teaching performance λ is satisfied, i.e., inequality (3) holds.

Proof: The proof is carried out by contradiction. Assume at trial t^*, the teaching performance is not satisfied. Thus, there is a solution to the hypothesis belief update equation (5) with $b_0 = p_0$ such that $b_{t^*}(h^*) < \lambda$. From inequality (9), we have

$$B(t, b_t) \leq B(t-1, b_{t-1})$$

for all $t \in \{1, 2, \ldots, t^*\}$ and all examples $z \in \mathcal{Z}$. Hence, $B(t, b_t) \leq B(0, b_0)$ for all $t \in \{1, 2, \ldots, t^*\}$. Furthermore, inequality (8) implies that

$$B(0, b_0) < 0$$

for $b_0 = p_0$. Since the choice of t^* can be arbitrary, this is a contradiction because it implies that $B(t^*, b_{t^*}) \leq B(0, b_0) < 0$. Therefore, there exist no solution of (5) such that $b_0 = p_0$ and $b_{t^*} \in \mathcal{B}_f$ for any sequence of examples $z \in \mathcal{Z}$. Hence, the teaching performance is satisfied. ■

In practice, we may have a large number of examples. Then, finding a barrier certificate that satisfies the conditions of Theorem 1 becomes prohibitive to compute. In the next result, we show how the calculation of the barrier certificate can be decomposed into finding a set of barrier certificates for each example and then taking the convex hull of them.

Theorem 2: Given the learning POMDP $(\mathcal{H}, p_0, \mathcal{Z}, T, \mathcal{Y}, O)$, a target hypothesis $h^* \in \mathcal{H}$, a teaching performance λ, and a pre-set number of trials t^*, if there exists a set of function $B_z : \mathbb{N} \times \mathcal{B} \to \mathbb{R}$, $z \in \mathcal{Z}$, such that

$$B_z(t^*, b_{t^*}) > 0, \quad \forall b_{t^*} \in \mathcal{B}_f, \quad \forall z \in \mathcal{Z}, \tag{10}$$

with $\mathcal{B}_f$ as described in (6),

$$B_z(0, b_0) < 0, \quad \text{for} \quad b_0 = p_0, \quad \forall z \in \mathcal{Z}, \tag{11}$$

and

$$B_z\left(t, f_z(b_{t-1}, y)\right) - B_z(t-1, b_{t-1}) \leq 0, \\ \forall t \in \{1, 2, \ldots, t^*\},\ \forall z \in \mathcal{Z},\ \forall y \in \mathcal{Y},\ \forall b \in \mathcal{B}, \quad (12)$$

then there the teaching performance λ is satisfied, i.e., inequality (3) holds. Furthermore, the overall barrier certificate is given by $B = \mathrm{co}\{B_z\}_{x \in \mathcal{Z}}$.

Proof: The proof was omitted due to lack of space here. Please refer to the extended version [17]. ■

The efficacy of the above result is that we can search for each example-based barrier certificate B_z, $z \in \mathcal{Z}$, independently or in parallel and then verify whether the overall teaching algorithm (described by the learning POMDP) satisfies a pre-specified teaching performance (see Fig. 1 for an illustration).

Next, we demonstrate that, if a teaching policy is given, the search for the barrier certificate can be decomposed into the search for a set of local barrier certificates. As discussed earlier, a teaching policy $\pi : \mathcal{B} \to \mathcal{Z}$ assigns an example to different regions of the belief space (refer to Section IV). Without loss of generality, we consider policies of the form

$$\pi(b) = \begin{cases} z_1, & b \in \mathcal{B}_1, \\ \vdots & \vdots \\ z_{|\mathcal{Z}|}, & b \in \mathcal{B}_N, \end{cases} \quad (13)$$

where N denotes the number of partitions of $\mathcal{B}$ and $\cup_{i=1}^{N} \mathcal{B}_i = \mathcal{B}$. Note that the number of partitions and the number of examples are not necessarily equal. We denote by z_i the example active in the partition $\mathcal{B}_i$.

Theorem 3: Given the learning POMDP $(\mathcal{H}, p_0, \mathcal{Z}, T, \mathcal{Y}, O)$, a target hypothesis $h^* \in \mathcal{H}$, a teaching performance λ, a teaching policy $\pi : \mathcal{B} \to \mathcal{Z}$ as described in (13), and a pre-set number of trials t^*, if there exists a set of function $B_i : \mathbb{N} \times \mathcal{B}_i \to \mathbb{R}$, $i \in \{1, 2, \ldots, N\}$, such that

$$B_i(t^*, b_{t^*}) > 0, \quad \forall b_{t^*} \in \mathcal{B}_f \cap \mathcal{B}_i,\ \ i \in \{1, 2, \ldots, N\}, \quad (14)$$

with $\mathcal{B}_f$ as described in (6),

$$B_i(0, b_0) < 0, \quad \text{for} \quad b_0 = p_0, \quad i \in \{1, 2, \ldots, N\}, \quad (15)$$

and

$$B_i\left(t, f_{z_i}(b_{t-1}, y)\right) - B_i(t-1, b_{t-1}) \leq 0, \\ \forall t \in \{1, 2, \ldots, t^*\},\ \forall y \in \mathcal{Y},\ \forall b \in \mathcal{B}_i, \\ i \in \{1, 2, \ldots, N\}, \quad (16)$$

then there the teaching performance λ is satisfied, i.e., inequality (3) holds. Furthermore, the overall barrier certificate is given by $B = \mathrm{co}\{B_i\}_{i=1}^{N}$.

Proof: The proof was omitted due to lack of space here. Please refer to the extended version [17]. ■

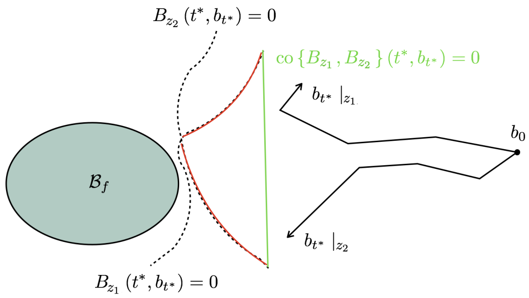

$B_{z_2}(t^*, b_{t^*}) = 0$
$\mathrm{co}\{B_{z_1}, B_{z_2}\}(t^*, b_{t^*}) = 0$
$b_{t^*}|_{z_1}$
b_0
$\mathcal{B}_f$
$b_{t^*}|_{z_2}$
$B_{z_1}(t^*, b_{t^*}) = 0$

Fig. 1: Decomposing the barrier certificate computation for a learning POMDP with two examples z_1 and z_2: the zero-level sets of B_{z_1} and B_{z_2} at trial t^* separate the evolutions of the hypothesis beliefs starting at b_0 from $\mathcal{B}_f$. The green line illustrate the zero-level set of the barrier certificate formed by taking the convex hull of B_{z_1} and B_{z_2}.

We proposed two techniques for decomposing the construction of the barrier certificates and checking a pre-set teaching performance. Our method relied on barrier certificates that take the form of the convex hull of a set of local barrier certificates (see similar results in [18], [19]). Though the convex hull barrier certificate may introduce a level of conservatism, it is computationally easier to find (as will be discussed in more detail in the next section). We remark that another technique that can be used for decomposition may use non-smooth barrier certificates [20], i.e., max or min of a set of local barrier certificates.

VI. Computational Method via SOS Programming

In this section, we propose techniques for finding the barrier certificates and checking whether a teaching performance is satisfied using SOS programming [21], [22].

In order to cast the conditions of Theorem 1-3 into SOS programs, we need polynomial/rational variables and require the associated sets to be semi-algebraic. Fortunately, these requirements naturally fit our problem. The hypothesis belief space is a semi-algebraic set. Moreover, the right-hand side of the belief update equation (5) is composed of rational functions in the belief states $b_t(h)$, $h \in \mathcal{H}$. That is,

$$b_t(h') = \frac{S_z\left(b_{t-1}(h'), y_{t-1}\right)}{R_z\left(b_{t-1}(h'), y_{t-1}\right)} \\ = \frac{O(h', z_{t-1}, y_t) \sum_{h \in \mathcal{H}} T(h, z_{t-1}, h') b_{t-1}(h)}{\sum_{h' \in \mathcal{H}} O(h', z_{t-1}, y_t) \sum_{h \in \mathcal{H}} T(h, z_{t-1}, h') b_{t-1}(h)}. \quad (17)$$

Furthermore, the teaching-failure set (6) is a semi-algebraic set.

At this point, we present conditions based on SOS programs to verify a given teaching performance of a teaching algorithm.

Corollary 1: Given the learning POMDP $(\mathcal{H}, p_0, \mathcal{Z}, T, \mathcal{Y}, O)$, a target hypothesis $h^* \in \mathcal{H}$, a teaching performance λ, and a pre-set number of trials t^*, if there exist polynomial functions $B \in \mathcal{R}[t, b]$ of degree d and $p^f \in \Sigma[b]$, and constants $s_1, s_2 > 0$ such that

$$B\left(t^*, b_{t^*}\right) + p^f\left(b_{t^*}\right)\left(b_{t^*}(h^*) - \lambda\right) - s_1 \in \Sigma\left[b_{t^*}\right], \quad (18)$$

$$-B\left(0, p_0\right) - s_2 > 0, \quad (19)$$

and

$$-R_z\left(b_{t-1}\right)^d\Bigg(B\left(t,\frac{S_z\left(b_{t-1},y\right)}{R_z\left(b_{t-1},y\right)}\right)-B(t-1,b_{t-1})$$
$$\in\Sigma[t,b_{t-1}],\forall t\in\{1,2,\ldots,t^*\},\ y\in\mathcal{Y},\ z\in\mathcal{Z},\quad(20)$$

then there exists no solution of (5) such that $b_0=p_0$ and $b_{t^*}\in\mathcal{B}_f$ and, hence, the teaching performance is satisfied.

Proof: The proof was omitted due to lack of space here. Please refer to the extended version [17]. ■

Similarly, we can formulate SOS feasibility conditions for checking the inequalities in Theorem 2.

Corollary 2: Given the learning POMDP $(\mathcal{H},p_0,\mathcal{Z},T,\mathcal{Y},O)$, a target hypothesis $h^*\in\mathcal{H}$, a teaching performance λ, and a pre-set number of trials t^*, if there exist polynomial functions $B_z\in\mathcal{R}[t,b]$, $z\in\mathcal{Z}$, of degree d and $p_z^f\in\Sigma[b]$, $z\in\mathcal{Z}$, and constants $s_z^1,s_z^2>0$, $z\in\mathcal{Z}$, such that

$$B_z\left(t^*,b_{t^*}\right)+p_z^f(b_{t^*})\left(b_{t^*}(h^*)-\lambda\right)$$
$$-s_z^1\in\Sigma\left[b_{t^*}\right],\quad z\in\mathcal{Z},\quad(21)$$

$$-B_z\left(0,p_0\right)-s_z^2>0,\quad z\in\mathcal{Z},\quad(22)$$

and

$$-R_z\left(b_{t-1}\right)^d\Bigg(B_x\left(t,\frac{S_z\left(b_{t-1},y\right)}{R_z\left(b_{t-1},y\right)}\right)-B_x(t-1,b_{t-1})\Bigg)$$
$$\in\Sigma[t,b_{t-1}],\forall t\in\{1,2,\ldots,t^*\},$$
$$y\in\mathcal{Y},\ z\in\mathcal{Z},\quad(23)$$

then there exists no solution of (5) such that $b_0=p_0$ and $b_{t^*}\in\mathcal{B}_f$ and, hence, the teaching performance is satisfied.

We assume that a teaching policy in the form of (13) assigns examples to semi-algebraic partitions of the hypothesis belief space $\mathcal{B}$ described as

$$\mathcal{B}_i=\left\{b\in\mathcal{B}\mid g_i(b)\leq 0\right\},\quad i\in\{1,2,\ldots,N\}.\quad(24)$$

We then have the following SOS formulation for Theorem 3 using Positivstellensatz.

Corollary 3: Given the learning POMDP $(\mathcal{H},p_0,\mathcal{Z},T,\mathcal{Y},O)$, a target hypothesis $h^*\in\mathcal{H}$, a teaching performance λ, a teaching policy $\pi:\mathcal{B}\to\mathcal{Z}$ as described in (13), a teaching performance λ, and a pre-set number of trials t^*, if there exist polynomial functions $B_i\in\mathcal{R}[t,b]$, $i\in\{1,2,\ldots,N\}$, of degree d, $p_i^{l_1}\in\Sigma[b]$, $i\in\{1,2,\ldots,N\}$, $p_i^{l_2}\in\Sigma[b]$, $i\in\{1,2,\ldots,N\}$, $p_i^{l_3}\in\Sigma[b]$, $i\in\{1,2,\ldots,N\}$, and $p_i^f\in\Sigma[b]$, $i\in\{1,2,\ldots,N\}$, and constants $s_i^1,s_i^2>0$, $i\in\{1,2,\ldots,N\}$, such that

$$B_i\left(t^*,b_{t^*}\right)+p_i^f(b_{t^*})\left(b_{t^*}(h^*)-\lambda\right)+p_i^{l_1}(b_{t^*})g_i(b_{t^*})$$
$$-s_i^1\in\Sigma\left[b_{t^*}\right],\quad i\in\{1,2,\ldots,N\},\quad(25)$$

$$-B_i\left(0,p_0\right)+p_i^{l_2}(p_0)g_i(p_0)-s_i^2>0,\quad i\in\{1,2,\ldots,N\},\quad(26)$$

and

$$-R_z\left(b_{t-1}\right)^d\Bigg(B_i\left(t,\frac{S_z\left(b_{t-1},y\right)}{R_z\left(b_{t-1},y\right)}\right)-B_i(t-1,b_{t-1})\Bigg)$$
$$+p_i^{l_3}(b_{t-1})g_i(b_{t-1})\in\Sigma[t,b_{t-1}],\forall t\in\{1,2,\ldots,t^*\},$$
$$y\in\mathcal{Y},\ z\in\mathcal{Z},\quad i\in\{1,2,\ldots,N\},\quad(27)$$

then there exists no solution of (5) such that $b_0=p_0$ and $b_{t^*}\in\mathcal{B}_f$ and, hence, the teaching performance is satisfied.

VII. Example

In order to illustrate the proposed framework, we consider a toy scenario, where the teacher aims to teach/steer a human learner to reach a goal state in a physical environment. Each hypothesis/node corresponds to some unexplored territory, and there exists an example which flags the territory as explored. The learner prefers local moves, and if all neighboring territories are explored, the learner jumps to the next closest one.

The physical environment is characterized by a 4×4 lattice corresponding to 16 hypotheses. The target hypothesis is located at $h^*=(4,4)$. The teacher has 16 choices of locations on the lattice to show to the student as examples. The student then receives two labels based on its answer $y\in\{-1,1\}$. The preference function $\sigma(h';h)$ is given by the minimum distance between hypotheses described by $\ell_1(h';h)$.

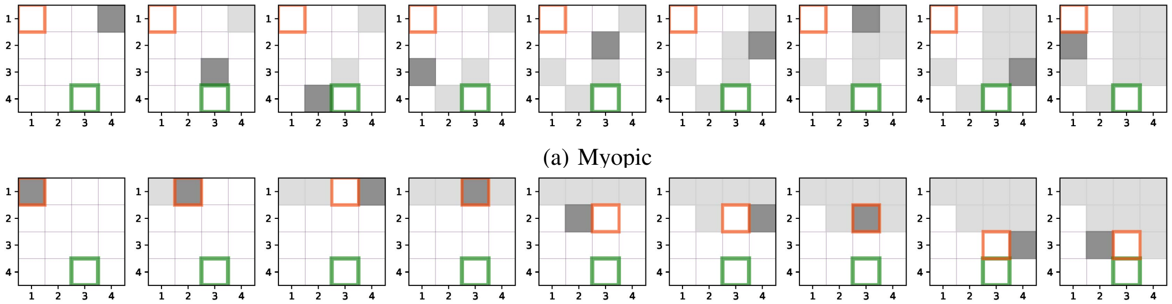
In this example, we compare two teaching algorithms in the adaptive setting, where the teacher observes the learner's hypothesis at each iteration. The Myopic algorithm is a greedy approach which, at each iteration, picks the teaching example such that after observing the label, the worst-case rank of the target hypothesis in the learner's resulting version space is the smallest. The Ada-L algorithm aims to teach the learner some intermediate hypothesis at each iteration, i.e., it aims to direct the learner to transit to a hypothesis that is "closer" to the target hypothesis. For more details of the algorithms please refer to [5].

Each algorithm provides a set of policies for which we seek to find the minimum number of trials such that the following teaching performance is assured

$$b_{t^*}(h^*)\geq\lambda.$$

To this end, we minimize the number of trials t^* such that (27)-(29) are satisfied. We start by a large number of trials (16 in this case) and decrease it until no barrier certificate can be found to verify the teaching performance. We fix the degree of variables B_i, $p_i^{l_1}$, $p_i^{l_2}$, $p_i^{l_3}$, and $p_i^f\in\Sigma[b]$, $i\in\{1,2,\ldots,N\}$ in Corollary 3 to 2 and search for the certificates. In order to check the SOS conditions formulated in Section VI, we use diagonally-dominant-SOS (DSOS) relaxations of the SOS programs implemented through the SPOTless tool [23] (for more details see [24], [25]).

The results on finding the minimum number of trials t^* for which the teaching performance is satisfied were as follows.



(a) Myopic

(b) Ada-L

Fig. 2: Teaching sequences generated by Myopic and Ada-L algorithms on a 4×4 lattice, with $h_0 = (1, 1), h^* = (3, 4)$. The learner's initial hypothesis is marked by orange, and the target is marked by green. The dark gray square represents the teaching example at the current time step, while light gray squares represent the previous teaching examples.

1) $h_0 = (1, 1)$ and $h^* = (3, 4)$: For the Myopic algorithm, we could not find any certificate for $\lambda = 0.8$. Changing the the teaching performance to $\lambda = 0.55$ yielded certificates for only $t^* = 15$. On the other hand, for the Ada-L algorithm, we obtained $t^* = 9$ assuring teaching performance $\lambda = 0.8$ and $t^* = 10$ assuring teaching performance $\lambda = 0.9$.

The results can also be corroborated from simulations. As can be see in Figure 2, the Myopic algorithm perform poorly on simple teaching tasks as compared to the Ada-L algorithm.

VIII. Conclusions

We presented a method based on barrier certificates to assure the performance of machine teaching algorithms. Our computational method was in terms of SOS programs, where we used DSOS relaxations. It was shown in [26] that using sparse SOS (SSOS) programs leads to more efficient and less conservative results. Future work can explore the use of more scalable SOS relaxations such as SSOS.

References

[1] L. Lessard, X. Zhang, and X. Zhu, "An optimal control approach to sequential machine teaching," *arXiv preprint arXiv:1810.06175*, 2018.

[2] X. Zhu, "Machine teaching: An inverse problem to machine learning and an approach toward optimal education." in *AAAI*, 2015, pp. 4083–4087.

[3] S. A. Goldman and M. J. Kearns, "On the complexity of teaching," *Journal of Computer and System Sciences*, vol. 50, no. 1, pp. 20–31, 1995.

[4] Z. Gao, C. Ries, H. U. Simon, and S. Zilles, "Preference-based teaching," *JMLR*, vol. 18, no. 31, pp. 1–32, 2017.

[5] Y. Chen, A. Singla, O. M. Aodha, P. Perona, and Y. Yue, "Understanding the role of adaptivity in machine teaching: The case of version space learners," in *Proc. Conference on Neural Information Processing Systems (NeurIPS)*, December 2018.

[6] M. Ahmadi, B. Wu, H. Lin, and U. Topcu, "Privacy verification in POMDPs via barrier certificates," in *Decision and Control (CDC), 2018 IEEE 57th Annual Conference on,*, 2018.

[7] M. Ahmadi, M. Cubuktepe, N. Jansen, and U. Topcu, "Verification of uncertain POMDPs using barrier certificates," in *56th Annual Allerton Conference on Communication, Control, and Computing,*, 2018.

[8] E. Bonawitz, S. Denison, A. Gopnik, and T. L. Griffiths, "Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference," *Cognitive psychology*, vol. 74, pp. 35–65, 2014.

[9] A. N. Rafferty, E. Brunskill, T. L. Griffiths, and P. Shafto, "Faster teaching via pomdp planning," *Cognitive science*, vol. 40, no. 6, pp. 1290–1332, 2016.

[10] K. Chatterjee, M. Chmelík, and M. Tracol, "What is decidable about partially observable Markov decision processes with ω-regular objectives," *Journal of Computer and System Sciences*, vol. 82, no. 5, pp. 878–911, 2016.

[11] R. Goebel, R. G. Sanfelice, and A. R. Teel, "Hybrid dynamical systems," *IEEE Control Systems*, vol. 29, no. 2, pp. 28–93, 2009.

[12] A. A. Ahmadi and P. A. Parrilo, "Non-monotonic Lyapunov functions for stability of discrete time nonlinear and switched systems," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*. IEEE, 2008, pp. 614–621.

[13] A. Kundu and D. Chatterjee, "On stability of discrete-time switched systems," *Nonlinear Analysis: Hybrid Systems*, vol. 23, pp. 191 – 210, 2017.

[14] W. Zhang, A. Abate, J. Hu, and M. P. Vitus, "Exponential stabilization of discrete-time switched linear systems," *Automatica*, vol. 45, no. 11, pp. 2526–2536, 2009.

[15] D. Liberzon, *Switching in Systems and Control*, ser. Systems & Control: Foundations & Applications. Birkhäuser Boston, 2003.

[16] J. P. Hespanha, "Uniform stability of switched linear systems: Extensions of LaSalle's invariance principle," *IEEE Transactions on Automatic Control*, vol. 49, no. 4, pp. 470–482, 2004.

[17] M. Ahmadi, B. Wu, Y. Chen, Y. Yue, and U. Topcu, "Barrier certificates for assured machine teaching," *arXiv preprint arXiv:1810.00093*, 2018.

[18] M. Ahmadi, A. Israel, and U. Topcu, "Safety assessment based on physically-viable data-driven models," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 6409–6414.

[19] M. Ahmadi, A. Israel, and U. Topcu, "Controller Synthesis for Safety of Physically-Viable Data-Driven Models," *ArXiv e-prints*, Jan. 2018.

[20] P. Glotfelter, J. Cortés, and M. Egerstedt, "Nonsmooth barrier functions with applications to multi-robot systems," *IEEE Control Systems Letters*, vol. 1, no. 2, pp. 310–315, Oct 2017.

[21] P. Parrilo, "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization," Ph.D. dissertation, California Institute of Technology, 2000.

[22] S. Prajna, A. Papachristodoulou, P. Seiler, and P. Parrilo, "SOSTOOLS: Sum of squares optimization toolbox for MATLAB V3.00," 2013.

[23] A. Megretski, "Systems polynomial optimization tools (SPOT)," 2010. [Online]. Available: https://github.com/anirudhamajumdar/spotless/tree/spotless_isos

[24] A. A. Ahmadi and A. Majumdar, "DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization," *arXiv preprint arXiv:1706.02586*, 2017.

[25] A. A. Ahmadi, G. Hall, A. Papachristodoulou, J. Saunderson, and Y. Zheng, "Improving efficiency and scalability of sum of squares optimization: Recent advances and limitations," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 453–462.

[26] Y. Zheng, G. Fantuzzi, and A. Papachristodoulou, "Sparse sum-of-squares (SOS) optimization: A bridge between DSOS/SDSOS and SOS optimization for sparse polynomials," *arXiv preprint arXiv:1807.05463*, 2018.