

# Elaborating on Learned Demonstrations with Temporal Logic Specifications

Craig Innes  
School of Informatics  
University of Edinburgh, UK  
Email: craig.innes@ed.ac.uk

Subramanian Ramamoorthy  
School of Informatics  
University of Edinburgh, UK  
Email: s.ramamoorthy@ed.ac.uk

**Abstract**—Most current methods for learning from demonstrations assume that those demonstrations alone are sufficient to learn the underlying task. This is often untrue, especially if extra safety specifications exist which were not present in the original demonstrations. In this paper, we allow an expert to elaborate on their original demonstration with additional specification information using linear temporal logic (LTL). Our system converts LTL specifications into a differentiable loss. This loss is then used to learn a dynamic movement primitive that satisfies the underlying specification, while remaining close to the original demonstration. Further, by leveraging adversarial training, the system learns to robustly satisfy the given LTL specification on unseen inputs, not just those seen in training. We show that our method is expressive enough to work across a variety of common movement specification patterns such as obstacle avoidance, patrolling, keeping steady, and speed limitation. In addition, we show how to modify a base demonstration with complex specifications by incrementally composing multiple simpler specifications. We also implement our system on a PR-2 robot to show how a demonstrator can start with an initial (sub-optimal) demonstration, then interactively improve task success by including additional specifications enforced with a differentiable LTL loss.

## I. INTRODUCTION

Giving physical demonstrations is an established method of teaching a robot complex movements [1]. However, such Learning from Demonstration (LFD) techniques often assume the expert’s demonstrations alone contain enough information to completely learn the underlying task. This is often untrue for at least two reasons:

First, the task may be difficult to demonstrate. This may be due to noisy sensors, control latency, fatigue, or lack of teleoperation skill. Simply copying the expert’s movements may therefore be insufficient to perform the task adequately. For example, in a pouring task, the demonstrator might have been shaky and inadvertently spilled some of the liquid; on a peg insertion task, the demonstrator may overshoot the hole initially then re-adjust.

Second, and more importantly, many tasks have *underlying* specifications which are simple to express verbally, but impossible to express through examples without an enormous amount of data [23]. For example, if a reaching task has an easily formulated safety constraint such as “don’t move your hand above a certain speed when there is a fragile object nearby”, it may take a large number of demonstrations to tease out this underlying behaviour.

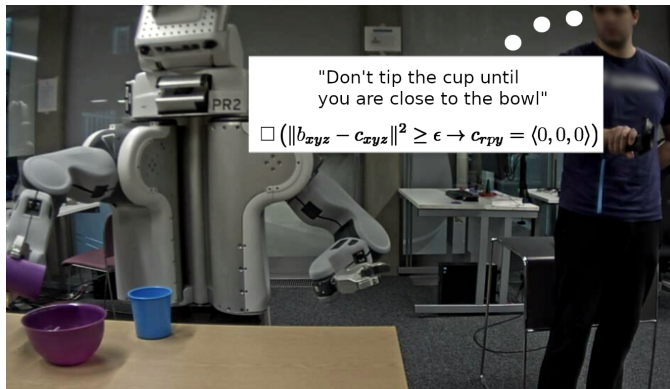


Fig. 1: A user pouring from a cup via teleoperation. The demonstration is later elaborated upon with the LTL specification: “Don’t tip the cup until you are close to the bowl”.

For these reasons, we would ideally like to follow a model closer to human teaching — one where we first give initial demonstrations, but can choose to elaborate on the precise detail of the task with additional instructions. In robotics, the language of Linear Temporal Logic (LTL) [24] provides us with an intuitive semantics with which to convey such instructions. LTL is also expressive enough to capture most of the common specification patterns seen in typical robotic tasks [20]. Further, using LTL opens up the potential for a user to add specifications directly with natural language, then later convert them into formal constraints using semantic parsing (we do not attempt this in this paper, but see e.g., Brunello et al. [2] for details).

In this paper, we present a neural network architecture which learns to imitate expert demonstrations by learning the weights of a *Dynamic Movement Primitive* (DMP) [27] — A generalizable representation for learning complex continuous movements. However, rather than merely optimizing the difference between the learned and target demonstrations, our system trades-off staying close to the original demonstration against satisfying additional LTL specifications provided by the user. To achieve this, we provide a measure which transforms LTL statements into a quantitative, differentiable metric. This means we can use this metric within the loss term of our optimization process. Further, by building on existing adversarial

techniques for optimizing networks with simple propositional constraints (e.g., Fischer et al. [8]), our system can propose unseen counterexamples which are maximally likely to violate the temporal constraint given the current DMP.

We incrementally build up task-learning by first learning a demonstration, then providing additional specifications later. In this sense, our work is close to the methodology of *Interactive Task Learning* (ITL). In such work, the goal is not just for the system to learn a given task, but to incrementally build an understanding of the task itself with help from the user.

In summary, our main contributions are:

- 1) A loss function which transforms LTL statements into a continuous, differentiable measure of satisfiability (Section II-B). Achieving a loss of zero guarantees that the constraint is satisfied for the given input.
- 2) A model which takes in object locations and robotic poses, and outputs the weights of a DMP (Section II-A). The model balances imitating the given demonstrations against satisfying the given LTL constraints. By using adversarial learning, the model learns to satisfy constraints on not just the training data, but on unseen examples designed to violate the constraints (Section II-C).
- 3) Experiments on 2-D curves across exemplary movement specifications such as obstacle avoidance, patrolling, staying within bounds, and staying within speed limits. These experiments combine demonstrations with task-relevant specifications to show two related uses of our system: First, we can elaborate on a single demonstration with LTL to satisfy specifications not originally present (Section III-A). Second, we can take a model trained on a batch of demonstrations, and use our LTL loss to satisfy a specification on an unseen test set (Section III-B).
- 4) Application of our model to reaching and pouring tasks on a PR-2 robot. The inputs are the gripper pose and object locations (extracted using the PR-2’s RGB-D camera). The output is a sequence of poses, which is synthesized into robotic motion using an inverse kinematics library. We show how a user can start with an initial demonstration captured via HTC-Vive VR controllers, then elaborate on that demonstration using LTL statements to either provide hints about how to better satisfy the task, or to modify the original task specification (Section III-C).

## II. COMBINING LFD AND LTL

In this section, we describe a method for teaching a robot to learn movements for *finite horizon* tasks from a combination of expert demonstrations and specifications given in LTL. We first briefly introduce learning to imitate demonstrations with DMPs (II-A), as well as the basics of LTL (II-B). Then, we outline our contributions, which are to formulate a differentiable loss function for LTL statements, and then optimize this loss using adversarial training techniques (II-C).

### A. Learning from Demonstrations with DMPs

Lets assume we have data for several demonstrations  $D = \{D_0 \dots D_M\}$  in the form  $D_i = \langle \mathcal{I}_i, y_i \rangle$ . Here,  $\mathcal{I}_i$  denotes the starting input for the demonstration, which may include relevant objects positions, the initial joint-angles/pose of the robot, or many other sensory inputs. The term  $y_i = [y_{i,0} \dots y_{i,T}]$  represents the *trajectory* of the robot, where  $y_{i,j}$  is the recorded pose at time  $j$  of demonstration  $i$ .

The goal is to learn a function that, given input  $\mathcal{I}_i$ , outputs trajectories that are close to the demonstrations. One common way to represent such trajectories is as a Dynamic Movement Primitive (DMP). DMPs allow us to compactly represent a trajectory as a series of differential equations:

$$\ddot{y} = \alpha_y(\beta_y(y_{goal} - \dot{y}) - \dot{y}) + f(x) \quad (1)$$

$$\dot{x} = -\alpha_x x \quad (2)$$

The acceleration of the end effector,  $\ddot{y}$ , has two parts: The first is a point attractor, which linearly moves towards  $y_{goal}$  at a rate determined by the distance  $(y_{goal} - y)$ , a drag term  $(-\dot{y})$ , and scalars  $\alpha_y$  and  $\beta_y$ . The second is the function  $f(x)$ , which expresses complex non-linear motion as a combination of radial basis functions  $\psi_{0\dots N}$ , weighted by parameters  $w_{0\dots N}$ :

$$f(x) = \frac{\sum_{i=0}^N w_i \psi_i(x)}{\sum_{i=0}^N \psi_i(x)} x (y_{goal} - y_{start}) \quad (3)$$

Equations (1-3) introduce a structural bias which favours smooth progress towards a goal over arbitrary movement.

Rather than using time directly, these equations are governed by the *canonical system*  $x$ , which starts at 1 and decays logarithmically towards 0.

With the above equations, we can now formally define the LFD problem: Given a demonstration  $D_i$ , we wish to learn a parameterized function  $g_\theta$  which takes  $\mathcal{I}_i$  as input and produces weights  $w$  as outputs such that the following loss function is minimized:

$$\mathcal{L}_d(\theta, \mathcal{I}_i, y_i) = \frac{1}{T} \sum_{t=0}^T \|\text{DMP}_{\theta, \mathcal{I}_i}(t) - y_{i,t}\|^2 \quad (4)$$

Here,  $\text{DMP}_{\theta, \mathcal{I}_i}(t)$  denotes the value at time  $t$  of a DMP defined by  $y_{start} = y_{i,0}$ ,  $y_{goal} = y_{i,T}$ , and weights learned via a neural network function parameterized by  $\theta$ :

$$w = g_\theta(\mathcal{I}_i) \quad (5)$$

In words, (4) gives the average distance between the learned and demonstrated trajectories. We call (4) the *imitation loss*.

Figure 2 shows an architecture overview. Like other works [22, 21], we represent  $g_\theta$  as a feed-forward network, allowing us to optimize (4) using stochastic gradient-descent<sup>1</sup>.

<sup>1</sup>For simple DMPs, we can approximate  $w$  in closed-form [13]. However, such methods do not accommodate additional sensory inputs, nor the modifications we make to the loss function in the next section.

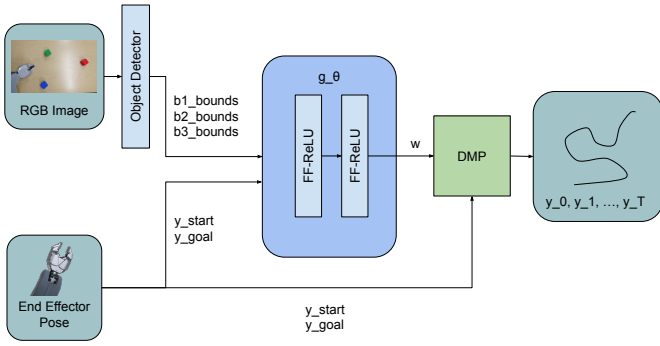


Fig. 2: Architecture overview. Inputs are the gripper pose and object descriptions (either given, or extracted from RGBD with an off-the-shelf object detector). We run these through two feed-forward layers with ReLU activations, and output weights  $w$  for equation (3). To calculate the trajectory, we roll out equations (1-2).

We now have a system which imitates demonstrated trajectories as closely as possible. The next sections outline how to extend (4) beyond mere imitation with additional specifications.

### B. A Loss Function from Temporal Logic

Linear Temporal Logic (LTL) [24] is a language which extends traditional propositional logic with modal operators. With these additional operators, we can specify robotic constraints that must hold *through time*. The syntax for constructing an LTL constraint  $\varphi$  is given by (6):

$$\begin{aligned} \varphi &:= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \mathcal{N}\varphi \mid \square\varphi \mid \diamond\varphi \mid \varphi_1\mathcal{U}\varphi_2 \\ p &:= T_1 < T_2 \mid T_1 \leq T_2 \mid T_1 = T_2 \mid T_1 \geq T_2 \mid T_1 > T_2 \end{aligned} \quad (6)$$

Here,  $\mathcal{N}\varphi$  means that  $\varphi$  must hold at the *next* time-step,  $\square\varphi$  means that  $\varphi$  must hold at *every* time-step,  $\diamond\varphi$  means that  $\varphi$  must *eventually* hold at some future time step, and  $\varphi_1\mathcal{U}\varphi_2$  means that  $\varphi_1$  must remain true *until*  $\varphi_2$  holds. Atomic comparisons  $p$  may be made up of *static* terms (which do not depend on time), or *dynamic* terms (which do depend on time, such as the value of a DMP). Terms are denoted by  $T_i$ .

Let’s say we wanted to express the constraint: “The robot must always stay away from obstacle  $o_{bad}$ , and must at some point visit  $o_{good}$ ”. We could express this in LTL as:

$$(\square\|\text{DMP}_{\mathcal{I}_i, \theta} - o_{bad}\|^2 \geq c) \wedge (\diamond\text{DMP}_{\mathcal{I}_i, \theta} = o_{good}) \quad (7)$$

We assume here that the positions of  $o_{bad}$  and  $o_{good}$  are static, while the current output of  $\text{DMP}_{\mathcal{I}_i, \theta}$  is dynamic, as it depends on the current time-step  $t$ .

The LTL definition given above is useful for discrete verification and high-level synthesis [16]. However, we cannot currently use it in our optimization process outlined in section II-A as it does not provide a continuous gradient to optimize with respect to. To use it for this purpose, we introduce a *constraint loss*  $\mathcal{L}_c$  which takes as input an LTL statement  $\varphi$  and

time-step  $t$ , and outputs a real-valued, differentiable loss. This loss measures quantitatively how close  $\varphi$  is to being satisfied.

Equation (8) gives the quantitative semantics for the atomic comparison terms. These build on the work by Fischer et al. [8] who define a differentiable quantitative semantics for propositional logic:

$$\begin{aligned} \mathcal{L}_c(t_1 \leq t_2, i) &= \max(t_1(i) - t_2(i), 0) \\ \mathcal{L}_c(t_1 \neq t_2, i) &= \zeta[|t_1(i) - t_2(i)|] \\ \mathcal{L}_c(t_1 < t_2, i) &= \mathcal{L}_c(t_1 \leq t_2 \wedge t_1 \neq t_2, i) \\ \mathcal{L}_c(t_1 = t_2, i) &= \mathcal{L}_c(t_1 \leq t_2 \wedge t_1 \geq t_2, i) \end{aligned} \quad (8)$$

Here,  $[a = b]$  is a function which evaluates to 1 when  $a = b$  and 0 otherwise. The constant  $\zeta$  is a small positive scalar.

Equation (9) gives the quantitative semantics for the modal and composition operators. It builds on quantitative measures of robustness from signal temporal logic [6]:

$$\begin{aligned} \mathcal{L}_c(\varphi_1 \wedge \varphi_2, i) &= \max^\gamma(\mathcal{L}_c(\varphi_1, i), \mathcal{L}_c(\varphi_2, i)) \\ \mathcal{L}_c(\varphi_1 \vee \varphi_2, i) &= \min^\gamma(\mathcal{L}_c(\varphi_1, i), \mathcal{L}_c(\varphi_2, i)) \\ \mathcal{L}_c(\mathcal{N}\varphi, i) &= \mathcal{L}_c(\varphi, i + 1) \\ \mathcal{L}_c(\square\varphi, i) &= \max^\gamma(\mathcal{L}_c(\varphi, i), \dots, \mathcal{L}_c(\varphi, T)) \\ \mathcal{L}_c(\diamond\varphi, i) &= \min^\gamma(\mathcal{L}_c(\varphi, i), \dots, \mathcal{L}_c(\varphi, T)) \\ \mathcal{L}_c(\neg\mathcal{N}\varphi, i) &= \mathcal{L}_c(\mathcal{N}(\neg\varphi), i) \\ \mathcal{L}_c(\neg\square\varphi, i) &= \mathcal{L}_c(\diamond(\neg\varphi), i) \\ \mathcal{L}_c(\neg\diamond\varphi, i) &= \mathcal{L}_c(\square(\neg\varphi), i) \\ \mathcal{L}_c(\varphi_1\mathcal{U}\varphi_2, i) &= \max^\gamma( \\ &\quad \min^\gamma(\mathcal{L}_c(\varphi_1, i), \mathcal{L}_c(\varphi_2, i)), \\ &\quad \min^\gamma(\mathcal{L}_c(\varphi_1, i + 1), \mathcal{L}_c(\varphi_2, i), \mathcal{L}_c(\varphi_2, i + 1)), \\ &\quad \dots, \\ &\quad \min^\gamma(\mathcal{L}_c(\varphi_1, T), \mathcal{L}_c(\varphi_2, i), \mathcal{L}_c(\varphi_2, i + 1), \dots, \mathcal{L}_c(\varphi_2, T))) \end{aligned} \quad (9)$$

Because maximization and minimization functions are non-differentiable, we use the soft approximations  $\max^\gamma$  and  $\min^\gamma$  (inspired by the work of Cuturi and Blondel [4] on soft approximations for dynamic time warping). These are defined in equations (10) and (11) below:

$$\max^\gamma(x_1, \dots, x_n) = \gamma \ln \sum_{i=0}^n e^{x_i/\gamma} \quad (10)$$

$$\min^\gamma(x_1, \dots, x_n) = -\gamma \ln \sum_{i=0}^n e^{-x_i/\gamma} \quad (11)$$

The above loss function has two useful properties. First, it is differentiable<sup>2</sup>. Second, it has the following soundness property (by construction):

**Property II.1.** *Soundness of  $\mathcal{L}_c$ : Given  $\gamma \rightarrow 0$ , for any  $\varphi$  and  $i$ , if  $\mathcal{L}_c(\varphi, i) = 0$ , then the constraint  $\varphi$  is satisfied at time-step  $i$  according to the original qualitative semantics of LTL.*

<sup>2</sup>To be precise, it is technically *almost-everywhere* differentiable.

In words, this means that minimizing  $\mathcal{L}_c$  is equivalent to learning to satisfy the constraint.

With the above properties, we can now augment our LFD loss from (4) with LTL constraints  $\varphi$  to give equation (12):

$$\mathcal{L}_{full}(\theta, D, \varphi) = \frac{1}{M} \sum_{i=0}^M \mathcal{L}_d(\theta, \mathcal{I}_i, y_i) + \eta \mathcal{L}_c(\varphi(\theta, \mathcal{I}_i), 0) \quad (12)$$

Here, we somewhat abuse notation and write  $\varphi(\theta, \mathcal{I})$  to mean evaluating  $\varphi$  with the values  $\theta$ , and  $\mathcal{I}$  substituted in as the values of the corresponding variables within the constraint. The term  $\eta$  is a constant which decides how much to trade off between perfectly imitating the expert demonstrations, and ensuring that  $\varphi$  is satisfied across all training inputs.

### C. Adversarial Training for Robust Constraint Satisfaction

By optimizing equation (12), we can (in the best case) learn parameters  $\theta$  which output movements guaranteed to satisfy the user-specified constraints. However, such guarantees only hold for the training inputs  $\mathcal{I}_{0\dots M}$ . For scenarios in which we wish to augment the *existing* behaviour of a *single* demonstration, this is fine. However in scenarios where we are training a *general* behaviour with a *batch* of varying demonstrations, we cannot be sure of whether our model’s ability to satisfy the given constraints will generalize to unseen examples.

This lack of generalization is a particular issue for LTL constraints which represent important safety specifications. For such constraints, we are often concerned with not only the average case behaviour of our model, but how the model will behave in the *worst possible situation* — we want to ensure our model satisfies the constraint *robustly*.

Let us clarify this notion of robustness with reference to the example from equation (7): Ideally, our robot should avoid  $o_{bad} = \mathcal{I}_{i,j}$  and reach  $o_{good} = \mathcal{I}_{i,k}$  even if we slightly change all relevant inputs by some small amount  $\epsilon$ . We want to be *robust* to perturbations, which we can write formally as equation (13):

$$\forall z_i \in B_\epsilon(\mathcal{I}_i), \quad (\square \|\text{DMP}_{z_i, \theta} - z_{i,j}\|^2 \geq c) \wedge (\diamond \text{DMP}_{z_i, \theta} = z_{i,k}) \quad (13)$$

Here,  $B_\epsilon(\mathcal{I}) = \{z \mid \|\mathcal{I} - z\|^\infty < \epsilon\}$  is the set of all points which lie within  $\epsilon$  of  $\mathcal{I}$  according to the  $L_\infty$  norm.

It is impractical to exhaustively check (13) holds at every valid instantiation of  $z_i$ . However, we can still robustly enforce this constraint by leveraging the following equivalence noted in Fischer et al. [8]: Saying  $\varphi$  holds everywhere is the same as saying there is no counter-example  $z'$  where  $\varphi$  *does not hold*.

This insight suggests the following two-step adversarial training loop: First, have an adversary find a  $z'$  which minimizes the *negated* loss  $\mathcal{L}_c(\neg\varphi, 0)$ . Second, give this  $z'$  as input to our network, and attempt to find the parameters  $\theta$  which minimize the full loss  $\mathcal{L}_{full}(\theta, D, \varphi)$ .

Algorithm 1 outlines our training process. Lines (6-9) show the adversarial inner loop. First, we sample a  $z$  from the region

around the current training example. We then repeatedly run  $z$  through our network  $g_\theta$ , and minimize the negated loss of our constraint  $\varphi$  using projected gradient descent. In lines (10-11), we then take the  $z$  chosen by our adversary, and use its gradient with respect to  $\mathcal{L}_{full}$  to update our network parameters  $\theta$ .

---

#### Algorithm 1 Adversarial training with demos and LTL

---

```

1: function TRAIN-DMP( $D, \epsilon, \gamma, \eta, \varphi, epochs, iterations$ )
2:   Initialize  $\theta$ 
3:   for  $1 \dots epochs$  do
4:     for  $\langle \mathcal{I}_i, y_i \rangle \in D$  do
5:        $z \leftarrow$  Sample from  $B_\epsilon(\mathcal{I}_i)$ 
6:       for  $1 \dots iterations$  do
7:          $y' \leftarrow$  Rollout DMP with weights  $w = g_\theta(z)$ 
8:          $z \leftarrow$  Update using  $\nabla_z \mathcal{L}_c(\neg\varphi(\theta, z), 0)$ 
9:          $z \leftarrow$  Project  $z$  onto  $B_\epsilon(\mathcal{I}_i)$ 
10:       $y \leftarrow$  Rollout DMP with weights  $w = g_\theta(z)$ 
11:       $\theta \leftarrow$  Update with:
12:         $\nabla_\theta (\mathcal{L}_d(\theta, \mathcal{I}_i, y_i) + \eta \mathcal{L}_c(\varphi(\theta, z), 0))$ 
return  $\theta$ 

```

---

In summary, we have a model which takes expert demonstrations combined with additional LTL specifications as input, and outputs the weights of a DMP. We can train this model to alter a single initial demonstration to obey additional constraints. Alternatively, we can leverage adversarial training and train our model to produce DMPs which satisfy the given constraints not just on the training data, but on unseen examples in a robust region.

## III. EXPERIMENTS

In this section, we show that the model presented in section II is practically useful via experiments across a range of tabletop movement tasks and across a variety of specifications. Our model is useful in two senses: First, we show that we can incrementally elaborate on the behaviour of a *single* demonstration using additional LTL specifications. Second, we show that given a *batch* of demonstrations augmented with an LTL specification, our adversarial learning approach outlined in algorithm 1 allows us to generalize the desired specification to unseen inputs.

We start by applying our model to 2-D planar motion scenarios. In subsequent sections we show example applications of our model on a PR-2 robot.

### A. One-Shot 2-D Tasks

We first consider single-shot demonstrations in a two-dimensional environment. Our goal is to observe whether our model can successfully adapt a base demonstration to one of four LTL constraints, which represent a selection of common robotic movement specifications. These are: Avoiding an obstacle, ensuring we reach a collection of patrol points, keeping movement within fixed bounds, and ensuring we do not move too quickly.

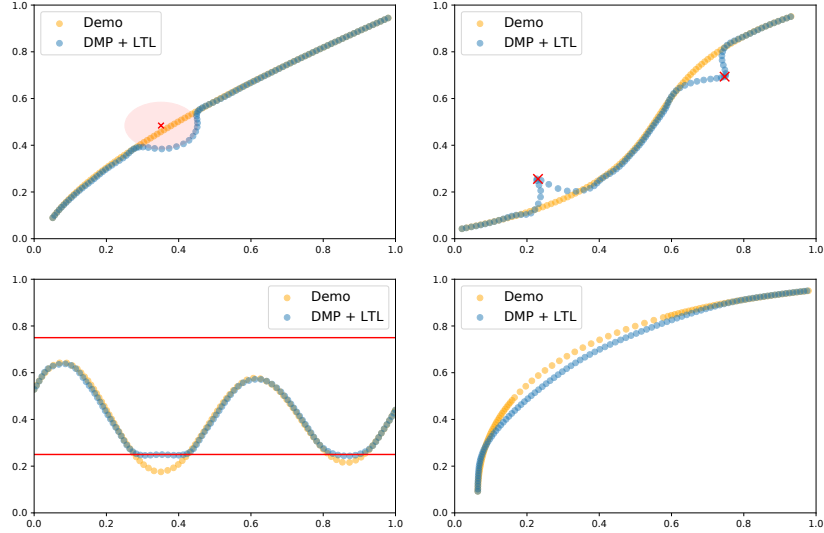


Fig. 3: The 2-D trajectories for (from left-to-right) avoid, patrol, steady, and slow. The original demonstration is in orange, while the learned trajectory is in blue. Points marked in red provide a visual representation of the associated constraint.

To create our 2-D demonstrations, we generate cubic-spline curves with start/end points in the region  $[0, 1]^2$ . Our demonstration trajectory  $y_i$  then consists of 100 equally-spaced points sampled from the generated curve. We also randomly generate the positions of three “objects” within the scene:  $o_1$ ,  $o_2$ , and  $o_3$ . These objects may (or may not) be relevant to the underlying task.

The LTL formulas for the four specifications mentioned above are given below. Here, we write  $p = \text{DMP}_{\mathcal{L}_i, \theta}$  as a shorthand for our learned DMP,  $p_y$  to mean the y dimension of a two-dimensional output, and  $\dot{p}$  as the current velocity of the DMP:

a) *Avoid*:

$$\square \|p - o_2\|^2 \geq 0.1$$

b) *Patrol*:

$$(\diamond p = o_2) \wedge (\diamond p = o_3)$$

c) *Steady*:

$$\square (p_y \geq 0.25 \wedge p_y \leq 0.75)$$

d) *Slow*:

$$\square (\|\dot{p}\| \leq 0.015)$$

For each task, we randomly generate 100 curves, varying the start, end and object positions. To train our DMP weights, we use the architecture outlined in figure 2. Our inputs are the start position of the demonstration, the goal position, and the positions of the three generated objects. We skip the object detection step here, as their positions are given to use directly in this section). Each hidden layer contains 256 units. We use the loss term given by equation (12) with  $\eta = 1.0$ ,  $\gamma = 0.005$ , and train using the Adam optimizer [15] for 200 epochs with

	Unconstrained		Constrained	
	$\mathcal{L}_d$	$\mathcal{L}_c$	$\mathcal{L}_d$	$\mathcal{L}_c$
Avoid	0.0046	0.0695	0.0127	<b>0.0003</b>
Patrol	0.0035	0.1257	0.0228	<b>0.0228</b>
Steady	0.0048	0.0272	0.0079	<b>0.0002</b>
Slow	0.0040	0.0098	0.0045	<b>0.0090</b>

TABLE I: The two separate parts of our loss function—the imitation loss ( $\mathcal{L}_d$ ) and constraint loss ( $\mathcal{L}_c$ )—on the single-demonstration experiments for the model trained with/without the LTL constraint. Results averaged over 20 iterations.

a learning rate of  $10^{-3}$ . For our DMP, we use the default parameters described by Hoffmann et al. [11].

Table I shows the results of our single-demonstration experiments each task. In all cases, the system learns parameters which almost completely minimize the constraint loss  $\mathcal{L}_c$ , at the cost of slightly increasing in the imitation loss  $\mathcal{L}_d$ .

Figure 3 gives a representative example of the qualitative behavior of our model on each of the task — The learned trajectory sticks as closely as possible to the original demonstration, except on those areas in which it must diverge to satisfy the additional constraint. Perhaps the most interesting instance of this behaviour (and the one that the model had the most difficulty optimizing) is the “Slow” constraint. Here, the learned trajectory appears to “cut-the-corner”, and turn right early. This pushes adjacent points closer together (minimizing the constraint loss), while keeping as close as possible to each corresponding point in the original demonstration (minimizing the imitation loss).

We note here that part of the strength of this approach is that it allows the user to build up increasingly complex specifications declaratively by incrementally composing LTL

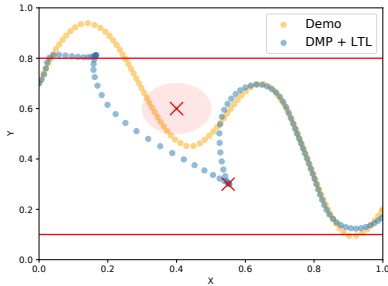


Fig. 4: 2-D trajectory for a demonstration augmented with avoid, steady, and patrol constraints. Constraints shown in red. This shows the ability of our model to adapt to complex specifications built up from combinations of simpler ones.

	Unconstrained	Train-Only	Adversarial
Avoid	0.0605	0.0272	<b>0.0192</b>
Patrol	0.1390	0.0413	<b>0.0369</b>
Steady	0.0286	0.0022	<b>0.0002</b>
Slow	0.0089	0.0070	<b>0.0069</b>

TABLE II: Constraint loss on test set (size 20) for generalized experiments. Models trained without the constraint (Unconstrained), with the constraint but only on the training data (Train-Only), and with adversarial counter-examples (Adversarial).

formulas. For example, figure 4 shows an example of adding *avoid*, *patrol*, and *steady* specifications incrementally on top of the one base demonstration. The model then computes the best way to minimally augment the original demonstration to satisfy this combination of constraints.

### B. Generalized Synthetic Tasks

We next look at whether, given a *batch* of demonstrations, our model can generalize the specified behaviour to unseen inputs. Further, we aim to show that training using the adversarial learning technique outlined in algorithm 1 results in more robust satisfaction of the given specification than training using the training data alone.

Our setup is identical to that of section III-A, but instead of optimizing for a single demonstration, we train on a batch of 100 demonstrations per task. We then evaluate the loss on a test set of 20 unseen inputs.

During training, we use an input batch-size of 32, and the inner loop of our adversarial search (lines 6-9) runs for 10 iterations with a robustness domain of  $\epsilon = 0.01$ . To evaluate the particular contribution of the adversarial steps, we also train a model exclusively on the training data inputs (“Train-Only”). This can be seen as equivalent to setting  $\epsilon = 0.0$ .

Table II shows the average constraint loss on the test data for each of our tasks. The model trained without the LTL specifications fails in general to satisfy those constraints on the test set (as the given demonstrations themselves did not satisfy the constraints). When training the system with the constraint

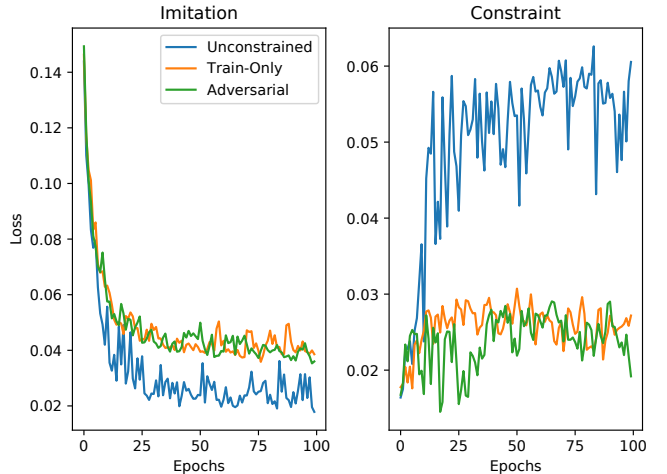


Fig. 5: Losses on test set for the “avoid” specification.

loss on the training data alone, the model successfully learns to satisfy the additional specifications to within 2-3 decimal places.

When we train using our adversarial approach, we see minor improvements across each task. Figure 5 shows the typical trend during the learning process: In earlier epochs, the adversarial approach is initially able to more aggressively lower the constraint loss, but as the model learns to better imitate the demonstrations, the train-only and adversarial approaches tend to level out. In the later epochs, the train-only and adversarial approaches maintain similar constraint losses, with the adversarial approach typically producing a marginally smaller loss by the final epoch.

An outlier in these results is the *steady* constraint, where the *adversarial* model outperforms the *train-only* approach by an order of magnitude. We hypothesize that this is because the points at which the demonstration trajectories actually violate the boundaries within the training data are sparse, making it difficult for the *train-only* approach to generalize to the unseen test data. In contrast, the adversarial approach proposes new random inputs at every epoch, then pushes these inputs towards places that are maximally likely to violate the boundary conditions. This suggests that the *adversarial* approach might be particularly useful in scenarios where the training data does not fully represent the given specification across the entire space.

### C. Experiments on the PR-2 Robot

In this section, we show two applications of our system on real-world manipulation tasks on a PR-2 robot. In the first, we show an example of a pouring task where the user observes the robot imitating their demonstration, then decides to elaborate on their demonstration with an additional LTL specification which captures the underlying constraint they were trying to observe. In the second application, we give an example of a reaching task, where the user later augments the tasks with two completely new specifications that were not present at all



Fig. 6: Setup for the pour (left) and reach (right) tasks

in the original demonstration.

We provide demonstrations by teleoperating the PR-2 using a pair of HTC-Vive virtual reality controllers. During teleoperation, we record the six-dimensional end-effector pose of the PR-2’s left-hand gripper over time at a rate of 20Hz. This constitutes the demonstrated trajectory. As in previous sections, we use the architecture described by figure 2 for learning. The initial scene-object locations are extracted by running the an image from the PR-2’s RGB-D camera through off-the-shelf object detection software. To synthesize movement, we then take the end-effector trajectory output by our learned DMP and use an inverse kinematics library<sup>3</sup> to calculate the appropriate joint movements<sup>4</sup>.

Figure 6 shows the setup for the two tasks we consider. We describe them in detail below.

*a) Pouring Task:* In this task, the PR-2 is holding a purple cup filled with rubber ducks in one hand, and a blue container in the other. The task is to pour all the ducks from the cup into the container without spilling any. After demonstrating this task several times, we noticed that it was difficult to keep the cup consistently steady via teleoperation, and that we would often tip the cup early before reaching the container (as it is difficult to observe from a distance). We therefore elaborated on our demonstration with the following advice designed to capture an underlying specification: “Don’t tip the cup until you are close to the container”.

The LTL formula for this specification is given below. As before, we write  $p = \text{DMP}_{\mathcal{I}, \theta}$  as a shorthand for the pose extracted from our learned DMP, and  $p_y$  to denote the  $y$  dimension of that pose:

$$\begin{aligned} & \square(\|p_{xyz} - o_1\|^2 \geq 0.1 \wedge p_z \geq o_{1,z}) \\ \implies & (\langle 0, 0, -1.0 \rangle \leq p_{rpy} \leq \langle 0.2, 0.2, 0.0 \rangle) \end{aligned}$$

*b) Reaching Task:* In this task, we started by demonstrating a simple specification — Reach from the edge of the table to the red cube. After the demonstration was recorded, we then incrementally add two additional specifications: “Avoid the purple bowl”, and “Visit the green cube”. The LTL for this additional specification is given below.

$$(\square\|p_{xyz} - o_2\|^2 \geq 0.2) \wedge (\diamond p_{xyz} = o_3)$$

<sup>3</sup><https://moveit.ros.org/>

<sup>4</sup>The demonstration data used, as well as videos of the final learned system in action, are available at <https://sites.google.com/view/ltl-dmp-rss-2020/>

	Unconstrained		Constrained	
	$\mathcal{L}_d$	$\mathcal{L}_c$	$\mathcal{L}_d$	$\mathcal{L}_c$
Robot-Pour	0.0033	0.0833	0.0293	0.0038
Robot-Reach	0.0059	0.1143	0.0710	0.0104

TABLE III: Imitation loss ( $\mathcal{L}_d$ ), constraint loss ( $\mathcal{L}_c$ ) for the robotic demonstration tasks.

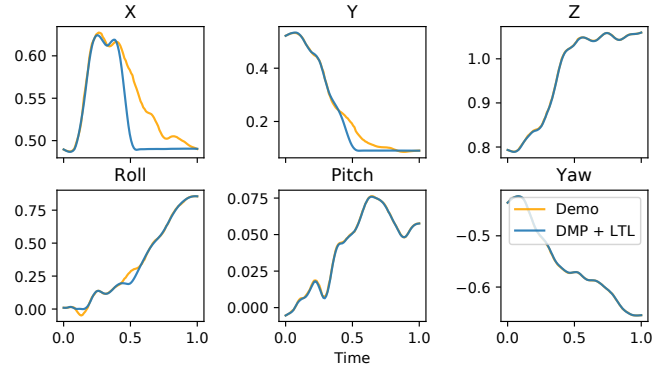


Fig. 7: Values of the xyz-rpy dimensions of the robot end-effector over time on the “cup-pour” experiment. Rotational axes are normalized between -1 and 1.

As Table III shows, our model was able to learn to satisfy the added constraints of both tasks, while keeping close to the initial demonstration. To get a sense of how the model altered the original demonstration, figure 7 shows the value of the six dimensions of the robot’s end-effector pose over time on the pouring task. In the roll dimension, we see that the model corrects for an early wobble at around  $t = 0.1$ , and also delays fully tipping the cup at around  $t = 0.5$ . However, it is in the X and Y dimensions that we observe the declarative nature of our specification: Rather than enforce a steady orientation throughout as one might expect, the model instead learns that it can stick closer to the original trajectory by instead negating the left-hand side of the specification implication by bringing the cup closer to the container earlier in the demonstration.

The combined experiments across these three sections show the potential of our method for incrementally learning a task by building additional user specifications on top of an initial physical demonstration.

#### IV. RELATED LITERATURE

The work closest to ours is Wen et al. [31], which learns from expert demonstrations with side-information provided by an LTL specification. However, their work frames the problem as inverse-reinforcement learning, where the LTL specification defines a reward function based on the likelihood that the robot will enter a terminating state which satisfies the specification. Other works [9, 17] also learn a reward function for which an optimal policy will satisfy a given temporal constraint. In contrast, our method uses the LTL specification to directly optimize the weights of a DMP, making it possible to inspect

how close the current motion comes to satisfying the given constraint.

Similarly, Hoffmann et al. [11] augment DMPs with a coupling term that avoids obstacles. However, this involves explicitly engineering specialized potential field terms into the underlying differential equations of the DMP. If the specification changes from obstacle avoidance to something else, those equations must be re-engineered. Our method instead allows the user to flexibly alter the behaviour of the DMP by giving an intuitive declarative specification in LTL.

Kasenberg and Scheutz [14] consider an inverse problem — learning an LTL specification *from* a demonstration. As we argued in the introduction however, this approach assumes that the underlying specification for a task is completely learnable from the given demonstrations alone.

Several recent works combine symbolic logic with neural networks. Xu et al. [32] derive a “semantic loss” for logical constraints, which can be used when a network’s output layer represents a distribution over a set of boolean variables. Similarly, Fischer et al. [8] use their Deep Learning with Differentiable Logic framework to transform propositional logic into differentiable costs for use in classification tasks. They also use an adversarial algorithm to search for constraint-violating counterexamples. We build on such techniques in section II-C, extending their expressiveness to temporal logic for robotic control.

There are many notions of robustness within robotic verification and control. Tabuada and Neider [30] refer to qualitative gradations of failure, such as violating a constraint once versus violating it infinitely. Fainekos and Pappas [6] convert temporal logic statements into a quantitative metric which measures distance from satisfiability when the robot fails to meet the constraint, and distance from unsatisfiability (robustness) when the robot does meet the constraint. Donzé and Maler [5] go further and use metric interval temporal logic to define *time-robustness*. However, the goal of all such work is typically either verification, or high-level planning using a known model of the problem [25, 16, 7, 26]. Our method also encodes a notion of robustness via algorithm 1, but our goals are different. We want to *learn* the low-level control parameters that satisfy the given specification. Consequently, our metric has the added properties of differentiability and having a zero-cost when satisfied. (Arguably, several works which devise metrics of *average robustness* could be also carried over into a network cost function in a similar manner without much effort [18, 19].)

Multiple works combine images, demonstrations and *natural language* [29, 3]. These works encode both images and language as a low-level embedding, then use this embedding to learn a control policy. Our system does not currently use natural language directly as input. Instead, we use a formal language to specify task constraints. This has the advantage that, unlike the natural language embeddings, the quantitative semantics from equations (8-9) have an intuitive interpretation, which allows the user to inspect the current model to see how close it is to satisfying the their specification.

Pervez et al. [22] and Pahic et al. [21] learn DMPs from images using neural networks. Unlike our paper however, these methods are pure imitation methods — they assume the given expert demonstration is all that is needed to achieve the underlying specification. Our method has a similar architecture, but elaborates on the initial demonstration using additional specifications encoded in the loss function.

Ho and Ermon [10] present an algorithm which alternates between imitating the given expert trajectories, and adversarially learning to discriminate between the current best policy and the expert trajectories. Both our paper and theirs take an adversarial approach to learning, but while their goal is to learn a robust imitation of a given demonstration, ours is ensure that additional LTL specifications are satisfied.

There are analogies between our method and the notion of residual policy learning [28]. Here, we start with a fixed control policy for a simpler task, then learn a residual policy to modify the fixed policy for a more complex situation. In this light, we can think of our initial expert demonstration as the fixed policy, and our additional elaboration with the LTL specification as the residual modification.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a quantitative, differentiable loss function which maps to the qualitative semantics of LTL, and an adversarial training regime for enforcing it. Through experiments on a variety of robotic motion tasks, we have shown that our method can successfully elaborate upon a single demonstration with additional specifications, and also ensure that a constraint can be robustly satisfied across unseen inputs when trained on a batch of demonstrations.

For future work, we aim to increase the range of expressible specifications in two ways. The first is to handle time-critical constraints by extending our framework to handle Metric Interval Temporal Logic (MITL) and investigate whether a notion of time-robustness [5] can be enforced adversarially. The second is to express constraints not just in terms of pre-defined symbols, but also in terms of relational predicates learned directly from data (as in e.g., [12]).

## ACKNOWLEDGMENTS

This work is supported by funding from the Alan Turing Institute, as part of the Safe AI for Surgical Assistance project. Additionally, we thank the members of the Edinburgh Robust Autonomy and Decisions Group for their feedback and support.

## REFERENCES

- [1] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 5 2009.
- [2] Andrea Brunello, Angelo Montanari, and Mark Reynolds. Synthesis of LTL formulas from natural language texts: State of the art and research directions.



- Leibniz International Proceedings in Informatics*, 147 (17):171–1719, 2019.
- [3] John D. Co-Reyes, Abhishek Gupta, Suvansh Sanjeev, Nick Altieri, Jacob Andreas, John DeNero, Pieter Abbeel, and Sergey Levine. Guiding policies with language via meta-learning. *7th International Conference on Learning Representations*, 2019.
- [4] Marco Cuturi and Mathieu Blondel. Soft-DTW: A differentiable loss function for time-series. In *34th International Conference on Machine Learning*, volume 2, pages 1483–1505, 3 2017.
- [5] Alexandre Donzé and Oded Maler. Robust satisfaction of temporal logic over real-valued signals. In *Lecture Notes in Computer Science*, volume 6246 LNCS, pages 92–106, 2010.
- [6] Georgios E. Fainekos and George J. Pappas. Robustness of temporal logic specifications for continuous-time signals. *Theoretical Computer Science*, 410(42):4262–4291, 2009.
- [7] Samira S Farahani, Vasumathi Raman, and Richard M Murray. Robust model predictive control for signal temporal logic synthesis. *IFAC-PapersOnLine*, 48(27): 323–328, 2015.
- [8] Marc Fischer, Mislav Balunovic, Dana Drachler-Cohen, Timon Gehr, Ce Zhang, and Martin Vechev. DL2: Training and Querying Neural Networks with Logic. *International Joint Conference on Artificial Intelligence*, pages 1532–1538, 2019.
- [9] Shalini Ghosh, Susmit Jha, Ashish Tiwari, Patrick Lincoln, and Xiaojin Zhu. Model, Data and Reward Repair: Trusted Machine Learning for Markov Decision Processes. In *Proceedings - 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops*, pages 194–199. Institute of Electrical and Electronics Engineers Inc., 2018.
- [10] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. In *Advances in Neural Information Processing Systems*, pages 4572–4580, 2016.
- [11] Heiko Hoffmann, Peter Pastor, Dae-Hyung Park, and Stefan Schaal. Biologically-inspired dynamical systems for movement generation: automatic real-time goal adaptation and obstacle avoidance. In *2009 IEEE International Conference on Robotics and Automation*, pages 2587–2592. IEEE, 2009.
- [12] Yordan Hristov, Daniel Angelov, Michael Burke, Alex Lascarides, and Subramanian Ramamoorthy. Disentangled Relational Representations for Explaining and Learning from Demonstration. In *Conference on Robot Learning*, 2019.
- [13] Auke Jan Ijspeert, Jun Nakanishi, Heiko Hoffmann, Peter Pastor, and Stefan Schaal. Dynamical Movement Primitives: Learning Attractor Models for Motor Behaviors. *Neural Computation*, 25(2):328–373, 2 2013.
- [14] Daniel Kasenberg and Matthias Scheutz. Interpretable apprenticeship learning with temporal logic specifications. In *IEEE 56th Annual Conference on Decision and Control*, pages 4914–4921. Institute of Electrical and Electronics Engineers Inc., 2018.
- [15] Diederik P. Kingma and Jimmy Lei Ba. Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations*, 2015.
- [16] Hadas Kress-Gazit, Tichakorn Wongpiromsarn, and Ufuk Topcu. Correct, reactive, high-level robot control. In *IEEE Robotics and Automation Magazine*, volume 18, pages 65–74, 2011.
- [17] Xiao Li, Yao Ma, and Calin Belta. A policy search method for temporal logic specified reinforcement learning tasks. In *2018 Annual American Control Conference (ACC)*, pages 240–245. IEEE, 2018.
- [18] Lars Lindemann and Dimos V Dimarogonas. Robust control for signal temporal logic specifications using discrete average space robustness. *Automatica*, 101:377–387, 2019.
- [19] Noushin Mehdipour, Cristian-Ioan Vasile, and Calin Belta. Arithmetic-geometric mean robustness for control from signal temporal logic specifications. In *2019 American Control Conference (ACC)*, pages 1690–1695. IEEE, 2019.
- [20] Claudio Menghi, Christos Tsigkanos, Patrizio Pelliccione, Carlo Ghezzi, and Thorsten Berger. Specification Patterns for Robotic Missions. *IEEE Transactions on Software Engineering*, 2019.
- [21] Rok Pahic, Andrej Gams, Ales Ude, and Jun Morimoto. Deep Encoder-Decoder Networks for Mapping Raw Images to Dynamic Movement Primitives. In *Proceedings - IEEE International Conference on Robotics and Automation*. IEEE, 2018.
- [22] Affan Pervez, Yuecheng Mao, and Dongheui Lee. Learning deep movement primitives using convolutional neural networks. In *IEEE-RAS International Conference on Humanoid Robots*, pages 191–197. IEEE, 2017.
- [23] Lerrel Pinto and Abhinav Gupta. Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours. In *Proceedings - IEEE International Conference on Robotics and Automation*, pages 3406–3413. Institute of Electrical and Electronics Engineers Inc., 2016.
- [24] Amir Pnueli. The temporal logic of programs. In *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 46–57, 1977.
- [25] Vasumathi Raman, Cameron Finucane, and Hadas Kress-Gazit. Temporal logic robot mission planning for slow and fast actions. In *IEEE International Conference on Intelligent Robots and Systems*, pages 251–256, 2012.
- [26] Dorsa Sadigh and Ashish Kapoor. Safe Control under Uncertainty. 10 2015. URL <http://arxiv.org/abs/1510.07313>.
- [27] Stefan Schaal, Jan Peters, Jun Nakanishi, and Auke Ijspeert. Learning Movement Primitives. *Robotics Research*, pages 561–572, 2005.
- [28] Tom Silver, Kelsey Allen, Josh Tenenbaum, and Leslie Kaelbling. Residual policy learning. *arXiv preprint*

*arXiv:1812.06298*, 2018.

- [29] Simon Stepputtis, Joseph Campbell, Mariano Phielipp, Chitta Baral, and Heni Ben Amor. Imitation Learning of Robot Policies by Combining Language, Vision and Demonstration. *arXiv preprint arXiv:1911.11744*, 2019.
- [30] Paulo Tabuada and Daniel Neider. Robust linear temporal logic. In *Leibniz International Proceedings in Informatics, LIPIcs*, volume 62. Schloss Dagstuhl- Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, 2016.
- [31] Min Wen, Ivan Papusha, and Ufuk Topcu. Learning from demonstrations with high-level side information. In *International Joint Conference on Artificial Intelligence*, pages 3055–3061, 2017.
- [32] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Van Den Broeck. A semantic loss function for deep learning with symbolic knowledge. In *35th International Conference on Machine Learning*, volume 12, pages 8752–8760, 2018.