Learning Lyapunov (Potential) Functions from Counterexamples and Demonstrations

Hadi Ravanbakhsh and Sriram Sankaranarayanan
Department of Computer Science, University of Colorado Boulder, Boulder, Colorado 80302
Email: firstname.lastname@colorado.edu

Abstract—We present a technique for learning control Lyapunov (potential) functions, which are used in turn to synthesize controllers for nonlinear dynamical systems. The learning framework uses a demonstrator that implements a black-box, untrusted strategy to synthesize control Lyapunov functions, which are used in turn to synthesize control inputs to stabilize the system starting from that state. However, synthesizing CLFs for nonlinear systems remains a challenge [41, 42]. Standard approaches to finding CLFs include the use of dynamic programming, wherein the value function satisfies the conditions of a CLF [4], or using non-convex bilinear matrix inequalities (BMI) [13]. BMIs can be solved using alternating minimization methods [8, 51, 31]. However, these approaches often get stuck in local minima and exhibit poor convergence guarantees [12].

In this article, we investigate the problem of learning a CLF using a black-box demonstrator that can be queried with a given system state, and responds by demonstrating control inputs to stabilize the system starting from that state. However, our framework uses just the control input at the query state. Such a demonstrator can be realized using an expensive nonlinear model predictive controller (MPC) that uses a local optimization scheme, or even a human operator under certain assumptions [4]. The framework has a learner which selects a candidate CLF and a verifier that tests whether this CLF is valid. If the CLF is invalid, the verifier returns a state at which the current candidate fails. The learner queries the demonstrator to obtain a control input corresponding to this state. It subsequently eliminates the current candidate along with a set of related functions from further consideration. The framework continues to exhaust the space of candidate CLFs until no CLFs remain or a valid CLF is found in this process.

We prove the process can converge in finitely many steps provided the learner chooses the candidate function appropriately at each step. We also provide efficient SDP-based approximations to the verification problem that can be used to drive the framework. Finally, we test this approach on a variety of examples, by solving stabilization problems for nonlinear dynamical systems. We show that our approach can successfully find simple CLFs using finite horizon nonlinear MPC schemes with appropriately chosen cost functions to serve as demonstrators. In these instances, the CLFs yield control laws that are computationally inexpensive, and guaranteed against the original dynamical model.

I. INTRODUCTION

We propose a novel learning from demonstration scheme for inferring control Lyapunov functions (potential functions) for stabilizing nonlinear dynamical systems to reference states/trajectories. Control Lyapunov functions (CLFs) have wide applications to motion planning problems in robotics. They extend the classic notion of Lyapunov functions to systems involving control inputs [3]. Finding a CLF also leads us to an associated feedback control law that can be used to solve the stabilization problem. Additionally, they can be extended for feedback motion planning using extensions to time-varying or sequential CLFs [5][53]. Likewise, they have been investigated in the robotics community in many forms including artificial potential functions to solve path planning problems involving obstacles [30].

However, synthesizing CLFs for nonlinear systems remains a challenge [41, 42]. Standard approaches to finding CLFs include the use of dynamic programming, wherein the value function satisfies the conditions of a CLF [4], or using non-convex bilinear matrix inequalities (BMI) [13]. BMIs can be solved using alternating minimization methods [8, 51, 31]. However, these approaches often get stuck in local minima and exhibit poor convergence guarantees [12].

In this article, we investigate the problem of learning a CLF using a black-box demonstrator that can be queried with a given system state, and responds by demonstrating control inputs to stabilize the system starting from that state. However, our framework uses just the control input at the query state. Such a demonstrator can be realized using an expensive nonlinear model predictive controller (MPC) that uses a local optimization scheme, or even a human operator under certain assumptions [4]. The framework has a learner which selects a candidate CLF and a verifier that tests whether this CLF is valid. If the CLF is invalid, the verifier returns a state at which the current candidate fails. The learner queries the demonstrator to obtain a control input corresponding to this state. It subsequently eliminates the current candidate along with a set of related functions from further consideration. The framework continues to exhaust the space of candidate CLFs until no CLFs remain or a valid CLF is found in this process.

We prove the process can converge in finitely many steps provided the learner chooses the candidate function appropriately at each step. We also provide efficient SDP-based approximations to the verification problem that can be used to drive the framework. Finally, we test this approach on a variety of examples, by solving stabilization problems for nonlinear dynamical systems. We show that our approach can successfully find simple CLFs using finite horizon nonlinear MPC schemes with appropriately chosen cost functions to serve as demonstrators. In these instances, the CLFs yield control laws that are computationally inexpensive, and guaranteed against the original dynamical model.

A. Illustrative Example: TORA System

Figure 1(a) shows a mechanical system consisting of a cart attached to a wall using a spring. The position of the cart is controlled by an arm with a weight that can be moved back and forth by applying a force u, as shown. The goal is to stabilize the cart to $x = 0$, with its velocity, angle, and angular velocity now corresponds to $\dot{x} = \dot{\theta} = 0$. We refer the reader to the Jankovic et al. [17] for a derivation of the dynamics shown below in terms of state variables $(x_1, \ldots, x_4)$ and control input $u_1$, after a suitable change of basis transformation:

$$\begin{align*}
\dot{x}_1 &= x_2, \\
\dot{x}_2 &= -x_1 + \epsilon \sin(x_3), \\
\dot{x}_3 &= x_4, \\
\dot{x}_4 &= u_1.
\end{align*}$$

(1)

We approximate $\sin(x_3)$ using a degree 3 approximation which is quite accurate over the range $x_3 \in [-2, 2]$. The equilibrium $x = 0, \dot{x} = \dot{\theta} = 0$ now corresponds to $x_1 = x_2 = x_3 = x_4 = 0$. The state space is taken to be $S : [-1, 1] \times [-1, 1] \times [-2, 2] \times [-1, 1]$, the control input $u_1 \in [-1.5, 1.5]$.

---

1We do not handle noisy or erroneous demonstrations in this paper.
Furthermore, the closed loop lacks formal guarantees despite a first approach to doing so uses a nonlinear model-predictive control (MPC) scheme using the time horizon $H = 30$, time step $\tau = 1$ and a simple cost function

$$\sum_{t=0,\tau,\ldots,H} (\|x(t)\|^2 + \|u(t)\|^2) + H \|x(H)\|^2.$$ 

Given the model in (1), such a control is implemented using a first order numerical gradient descent method to minimize the cost function over a finite time horizon. The convergence was informally confirmed by observing hundreds of such simulations from different initial conditions for the system. However, the MPC scheme is expensive, requiring repeated solutions to (constrained) nonlinear optimization problems in real-time. Furthermore, the closed loop lacks formal guarantees despite the high confidence gained from numerous simulations.

Learning a Control Lyapunov Function: The approach in this paper uses the MPC scheme as a “DEMONSTRATOR”, and attempts to learn a simpler control law through a control Lyapunov function. The key idea depicted in Fig. 2 is to pose queries to the MPC at finitely many witness states $W = \{x_1, \ldots, x_j\}$ and use the corresponding instantaneous control inputs $u_1, \ldots, u_j$, respectively. The LEARNER attempts to find a candidate function $V(x)$ that is positive definite and which decreases at each witness state $x_j$ through the control input $u_j$. This function is fed to the VERIFIER, which checks whether $V(x)$ is indeed a CLF, or discover a state $x_{j+1}$ at which the condition fails. This new state is added to the witness set and the process is iterated.

The procedure described in this paper synthesizes the control Lyapunov function after 60 iterations of the learning loop and synthesizes the CLF $V(x)$ below:

$$V(x) = 1.22 x_2^2 + 0.31 x_2 x_3 + 0.44 x_3^2 - 0.28 x_4 x_2 + 0.8 x_4 x_3 + 1.69 x_2^2 + 0.069 x_1 x_2 - 0.68 x_1 x_3 - 1.85 x_4 x_1 + 1.60 x_1^2$$

This function yields a simple associated control law that can be implemented and guarantees the stabilization of the model (1). Figure (b) shows a closed loop trajectory of this control law vs control law extracted by MPC. The advantage of this law is that its calculation is much simpler.

II. BACKGROUND

In this section, we briefly describe control Lyapunov (potential) functions. A state feedback control system $\Psi (X, U, \mathcal{P}, \mathcal{C})$ consists of a plant $\mathcal{P}$ and a controller $\mathcal{C}$ over state space $X \subseteq \mathbb{R}^n$ and input space $U \subseteq \mathbb{R}^m$. The plant $\mathcal{P}$ has a state $x \in X$ and input $u \in U$. The vector field for the plant is defined by a smooth function $f : X \times U \to \mathbb{R}^n$. Throughout the paper, we consider control affine systems that are possibly nonlinear in $x$, but affine in $u$, of the form

$$\dot{x} = f(x, u) = f_0(x) + \sum_{i=1}^m f_i(x)u_i,$$ 

where $f_i : X \to \mathbb{R}^n$. The controller $\mathcal{C}$ measures the state of the plant ($x \in X$) and provides feedback $u \in U$. The controller is defined by a continuous feedback function $K : X \to U$. 

Fig. 1. Tora System. (a) A schematic diagram of the TORA system. (b) Execution traces of the system using MPC control (blue traces) and Lyapunov based control (red traces) starting from same initial point.

Fig. 2. Overview of the learning framework for learning a Lyapunov function.
For a given feedback function $\mathcal{K}$, the execution trace of the system $\Psi$ is defined as $x(.) : \mathbb{R}^+ \rightarrow X$, which maps time to state. Formally, given $x(0) = x_0$, $x(.)$ is defined according to $\dot{x}(t) = f_0(x(t)) + \sum_{i=1}^m f_i(x(t))(\mathcal{K}(x(t)))i$, where $\dot{x}(.)$ is the right derivative of $x(.)$.

In this article, we study stabilization of nonlinear systems using Lyapunov functions (or potential function) inside a compact set $S$. More precisely, we consider a compact and connected set $S \subset X$. Without loss of generality, the origin $0$ is the state we seek to stabilize to. Furthermore, $0 \in \text{int}(S)$. We restrict the set $S$ to be a basic semi-algebraic set defined by a conjunction of polynomial inequalities. Likewise, the control inputs $U$ are restricted to a polytope.

Our approach relies on control Lyapunov functions. A control Lyapunov function (CLF) $V$, is a continuous function that respects the following conditions:

\[
\begin{align*}
V(0) &= 0 \\
(\forall x \in S \setminus \{0\}) & \quad V(x) > 0 \\
(\forall x \in S \setminus \{0\}) & \quad (\exists u \in U) \quad \nabla V \cdot f(x, u) < 0.
\end{align*}
\]

The last condition ensures that the value of $V$ can be decreased everywhere by choosing a proper feedback $u$. Let $V^{\alpha, \beta} = \{x|V(x) \leq \beta\}$, where $\alpha, \beta \geq 0$.

Let $\beta^*$ be maximum $\beta$ s.t. $V^{\leq \beta} \subseteq S$. Once a CLF is obtained, it guarantees that the system initialized to any state belonging to $V^{\leq \beta^*}$, can be stabilized to the origin (Fig. 3). A control Lyapunov function guarantees that there is a control strategy, which stabilizes the system.

**Theorem 1:** Given a control affine system $\Psi$, where $U : \mathbb{R}^m$ and a polynomial control Lyapunov function $V$ satisfying Eq. (3), there is a feedback function $\mathcal{K}$ for which if $x_0 \in V^{< \beta}$, then:

1. $(\forall t \geq 0) \quad x(t) \in S$
2. $(\forall \epsilon > 0) \quad (\exists T \geq 0) \quad ||x(T) - 0|| < \epsilon.$

Given a control Lyapunov function, it is possible to then obtain a feedback law in a closed form that stabilizes the system. Sontag [49] provides a method for extracting a continuous feedback function $\mathcal{K}$ for control affine systems from a control Lyapunov function. This can be extended to systems with constraints on the control inputs [53]. Also, feedback synthesis for periodic time/event triggered switching is possible [43].

We have thus far considered the problem of stabilizing to a fixed equilibrium state. However, given this primitive, we can extend the CLF approach to related problems of (a) Reachwhile-stay: reaching a given target set of states $T$ starting from an initial set $I$, while staying inside a safe set $S$ using Lyapunov-barrier functions [39, 44]; (b) Trajectory stabilization: stabilizing to a trajectory $x(t)$ rather than to a fixed state using time-varying Lyapunov functions; or similarly, (c) Feedback motion planning which addresses the robustness of a plan using funnels [5, 53]; (d) Obstacles: problems that involve reaching while avoiding an obstacle region in the state-space using artificial potential functions [30]. We will focus our exposition on the basic formulation for stability (Eq. (3)) while demonstrating extensions to some of other applications mentioned above through numerical examples.

### III. Algorithmic Learning Framework

Finding CLFs is known to be a hard problem, requiring the solution to BMIs [51] or hard polynomial constraints [43]. A standard approach to discovering such functions is to choose a set of basis functions $g_1, \ldots, g_r$ and search of a function of the form

\[
V_c(x) = \sum_{j=1}^r c_j g_j(x), \tag{4}
\]

where $c \in \mathbb{R}^r$ is vector of unknowns. One possible choice of basis functions involves monomials $g_j(x) : x^\alpha$, wherein $|\alpha_j|_1 \leq D_L$ for some degree bound $D_L$ for the learning concept (CLF). Then, the problem is reduced to finding $c$ s.t. $V_c$ satisfies Eq. (3).

We now present the algorithmic learning framework. Let us fix a control affine system $\mathcal{P}$ over a state-space $X$, control inputs $U$ given by (2). Let $x^* = 0$ be the equilibrium we wish to stabilize the system to, while remaining inside $S \subset X$.

Next, we assume a demonstrator as a function $\mathcal{D} : S \mapsto U$ that given a state $x \in S$, provides us an appropriate feedback $\mathcal{D}(x) \in U$ for the state $x$, such that $\mathcal{D}$ is presumed to be a valid function that stabilizes the system.

**Remark 1:** Our definition of a demonstrator is general enough to allow offline MPC, sample based methods [27, 23], human operator demonstrations [21], or even demonstrations that rely on opaque models such as neural networks.

Also, we assert that the demonstrator is presumed correct. However, the approach can work even if the demonstrator may fail on some input states. Finally, a faulty demonstrator may, at the worst, lead our technique to fail without finding a CLF. In particular, such a demonstrator will not cause our technique to synthesize an incorrect CLF.

**Definition 1 (Problem Statement):** The CLF learning problem has the following inputs:

1. A dynamical system $\mathcal{P}$ in the form (2).
2) A safe set $S$.
3) A “black-box” demonstrator function $D : S \mapsto U$ that presumably stabilizes the system, and
4) A candidate space for CLFs of the form $\sum_{j=1}^{r} c_{j}g_{j}(x)$ given by basis functions $g(x) : (g_{1}(x), \ldots, g_{r}(x))$ and a compact set $C \ni (c_{1}, \ldots, c_{r})$. We represent the coefficients $(c_{1}, \ldots, c_{r})$ collectively as $c$.

The output can be SUCCESS: a function $V_{c}(x) : c' \cdot g(x)$ that is a CLF; or FAILURE: no function could be discovered by our procedure.

A. Algorithmic Learning Framework

The algorithmic learning framework is shown in Fig. 4 and implements two modules (a) LEARNER and (b) VERIFIER that interact with each other and the demonstrator. The framework works iteratively until termination. At the $j^{th}$ iteration, the learner maintains a (witness) set

$$W_{j} : \{(x_{1}, u_{1}), \ldots, (x_{j}, u_{j})\} \subseteq S \times U.$$ 

$W_{j}$ is a finite set of pairs of states $x_{i}$ and corresponding demonstrated feedback $u_{i}$. Corresponding to $W_{j}$, $C_{j} \subseteq C$ is defined as set of candidate coefficients for functions $V_{c}(x) : c' \cdot g(x)$ with $c \in C_{j}$. Formally, $C_{j}$ is a set of all candidates $c$ s.t. $V_{c}$ satisfies the CLF condition (5) for every point in the finite set $W_{j}$:

$$C_{j} : \left\{ c \in C \left| \bigwedge_{(x_{i}, u_{i}) \in W_{j}} V_{c}(x_{i}) > 0 \land \nabla V_{c} \cdot f(x_{i}, u_{i}) < 0 \right\} \right. \right. \right.$$

The flowchart for the overall procedure is shown in Fig. 4. To begin with, $W_{0} : \emptyset$ and $C_{0} : C$. Each iteration works as follows:

1) The learner samples a value $c_{j} \in C_{j}$ and outputs the corresponding function $V_{j}(x) : c'_{j} \cdot g(x)$. If $C_{j} = \emptyset$ then no sample is found and the algorithm fails.
2) The verifier checks if $V_{j}$ is a CLF by checking the conditions in (5). If $V_{j}$ satisfies the conditions, then the algorithm stops to declare success. Otherwise the verifier selects a (counterexample) state $x_{j+1} \in S$ for which the CLF condition fails. Assume without loss of generality that $x_{j+1} \neq 0$.
3) Failing verification, the demonstrator is called to choose a suitable control $u_{j+1}$ corresponding to $x_{j+1}$.
4) The new set $W_{j+1} := W_{j} \cup \{(x_{j+1}, u_{j+1})\}$. Furthermore,

$$C_{j+1} : C_{j} \cap \left\{ c \mid V_{c}(x_{j+1}) > 0 \land \nabla V_{c} \cdot f(x_{j+1}, u_{j+1}) < 0 \right\}.$$ (6)

We now prove some core properties that guarantee the correctness of the proposed scheme. We assume that the learner and the verifier are implemented without any approximations/relaxations (as will be subsequently presented).

Theorem 2: The algorithmic learning framework as described above has the following property:

1) $c_{j} \notin C_{j+1}$. I.e., the candidate found at the $j^{th}$ step is eliminated from further consideration.
2) If the algorithm succeeds at iteration $j$, then the output function $V_{j}(x)$ is a valid CLF for stabilization.
3) The algorithm declares failure at iteration $j$ if and only if no linear combination of the basis functions is a CLF compatible with the demonstrator.

Proofs are available in [45].

B. Learner: Finding a Candidate

The FINDCANDIDATE function simply samples a point from the set $C_{j}$ defined in Eq. (3). Note that $V_{c}(x_{j}) : c' \cdot g(x_{j})$ is linear in $c$ and therefore $\nabla V_{c} \cdot f(x_{j}, u_{j})$ is linear in $c$ as well. The initial space of all candidates $C_{j}$ is assumed to be a hyper-rectangular open box. At each iteration, the candidate $c_{j} \in C_{j}$ is chosen. Suppose the algorithm does not terminate at this iteration. According to Eq. (6), the new set $C_{j+1}$ is obtained as $C_{j+1} : C_{j} \cap H_{j}$, wherein $H_{j}$ is defined by two linear inequalities $H_{j1} \land H_{j2} (H_{j1} : a_{j1}^{T}c < b_{j1}, H_{j2} : a_{j2}^{T}c < b_{j2})$. Let $C_{j}'$ represent the topological closure of the set $C_{j}$.

Lemma 3.2: For each $j \geq 0$, $C_{j}'$ is a convex polyhedron.

Thus the problem of implementing findCandidate is that of checking emptiness of a polyhedron with some strict inequality constraints. This is readily solved using a slight modification of standard linear programming algorithms using infinitesimals for strict inequalities.

Lemma 3.3: There exists a halfspace $H_{jk} : a_{jk}^{T}c \geq b_{jk}$ that passes through $c_{j}$, and $C_{j+1} \subseteq C_{j} \cap H_{jk}$. 
Choosing a Candidate $c_j$: The choice of $c_j \in C_j$ governs the overall convergence of the algorithm. Figure 5 demonstrates the importance of this choice by showing the candidate $c_j$, the hyperplane $H_{jk}$ and the new region $C_{j+1}$.

We wish to choose $c_j$ s.t. $\text{Vol}(C_{j+1}) \leq \alpha \text{Vol}(C_j)$, for a fixed $\alpha$, independent of $H_{jk}$.

**Theorem 3 (Tarasov et al.):** Let $c_j$ be chosen as the center of the maximum volume ellipsoid (MVE) of $C_j$. Then,

$$\text{Vol}(C_{j+1}) \leq \left(1 - \frac{1}{r}\right) \text{Vol}(C_j).$$

This leads us to a scheme that guarantees termination of the overall iterative scheme in finitely many steps under a robustness assumption.

**Definition 2 (Robust Candidate):** A candidate $c \in C$ is $\delta$-robust ($\delta > 0$), if for each $\hat{c} \in B_\delta(c)$, $V_{\hat{c}} : e' \cdot g(x)$ is a CLF, where $B_\delta(c)$ is a ball of radius $\delta$ around $c$.

Let $\text{Vol}(C) < \gamma \Delta^r$ for $\gamma > 0$ the volume of the unit $r$-ball, and $\Delta > 0$. Furthermore, the procedure terminates whenever $\text{Vol}(C_{j+1}) < \gamma \delta^r$ following the robustness assumption above.

**Theorem 4:** If at each step $c_j$ is chosen as the center of the maximum volume ellipsoid in $C_j$, the learning loop terminates in at most

$$\frac{r \left( \log(\Delta) - \log(\delta) \right)}{-\log \left( 1 - \frac{1}{r} \right)} = O(r^2) \text{ iterations}.$$

The maximum volume ellipsoid itself can be computed by solving a convex optimization problem [35].

**C. Implementing the Verifier**

The verifier given a candidate $V_j(x) : c_j' \cdot g(x)$ checks the CLF conditions in Eq. (3), split into two separate checks:

(A) Check if $V_j(x)$ is a positive definite polynomial over $x \in S$. Assuming that $g(0) = 0$ for the basis, this reduces to:

$$\exists x \in S \setminus \{0\}, V_j(x) \leq 0. \tag{7}$$

(B) Check if the Lie derivative of $V_j$ can be made negative for each $x \in S$ by a choice $u \in U$:

$$\exists x \in S \setminus \{0\}, \forall u \in U: \left( \nabla V_j \cdot f(x, u) \geq 0. \tag{8} \right)$$

This problem seems harder due to the presence of a quantifier alternation. Let $U$ be a polyhedral set defined by $U : \{u \in \mathbb{R}^m \mid Au \geq b\}$. Recall that $f(x, u)$ is control affine function $f_0(x) + \sum_{i=1}^m f_i(x)u_i$.

**Lemma 3.4:** Eq. (8) holds for some $x \in S$ iff

$$\exists x \in S \setminus \{0\}, \lambda \geq 0, \lambda' b \geq -\nabla V_j f_0(x)$$

$$A_i'\lambda = \nabla V_j f_i(x) \forall i \in \{1 \ldots m\}. \tag{9}$$

In other words, assuming that the dynamics and chosen bases are polynomials, the verification problem reduces to checking if a given semi-algebraic set defined by polynomial inequalities has a solution.

Failing the polynomial assumption, the problem of verification is in general undecidable. However, it can be approximated by techniques such as $\delta$-decision procedures proposed by Gao et al [10]. Solvers like dReal can thus be directly used for the verification problem. While dReal does a good job in adaptive space decomposition, in our experience, they do not scale reliably. Nevertheless, these solvers allow us to conveniently implement a verifier for small but hard problems involving rational and trigonometric functions.

The verification problem for polynomial systems and polynomial CLFs through a polynomial basis function $g(x)$ is NP-hard, in general. Exact approaches using semi-algebraic geometry and the branch-and-bound solvers (including the dReal approach cited above) can tackle this problem precisely. However, scalability is still an issue.

We sketch a relaxation using SDP solvers [35]. Let us fix a basis of monomial terms of degree up to $D_V$, $m^t = [1 \ x_1 \ldots \ x_{D_V}]$, wherein $D_V$ is chosen as at least half of the maximum degree in $x$ among all monomials in $g_j(x)$ and $\nabla g_j$.

$$f(x) : D_V \geq \frac{1}{2} \max \left( \sum_{j} \left( \text{deg}(g_j) \cup \text{deg}(\nabla g_j \cdot f) \right) \right).$$

Let $Z(x) : mm^t$. Thus, each polynomial of degree up to $2D_V$ may now be written as a trace inner product $p(x) = \langle P, Z(x) \rangle = \text{trace}(PZ(x))$, wherein the matrix $P$ is constant.

Let $S$ be the semi-algebraic set defined as

$$S : \{x \in \mathbb{R}^n \mid r_1(x) \leq 0, \ldots, r_k(x) \leq 0\},$$

for polynomials $r_1, \ldots, r_k$. The constraint in (7) is equivalent to solving the following optimization problem over $x$:

$$\max_{x} \langle I, Z(x) \rangle \tag{10}$$

s.t. $\forall i \in \{1, \ldots, k\}$

$$\forall j \in \{1, \ldots, m\}$$

$$\langle V_j, Z(x) \rangle \leq 0,$$

where $V_j(x) (r_j(x))$ is written as $\langle V_j, Z(x) \rangle (\langle R_i, Z(x) \rangle)$; and those in (9) are written as

$$\max_{x, \lambda} \langle I, Z(x) \rangle \tag{11}$$

s.t. $\forall i \in \{1, \ldots, k\}$

$$\langle R_i, Z(x) \rangle \leq 0, \lambda \geq 0,$$

$$\langle F_j, Z(x) \rangle = A_i' \lambda, \lambda \geq 0,$$

wherein the components $\nabla V_j \cdot f_i(x)$ defining the Lie derivatives of $V_j$ are now written in terms of $Z(x)$ as $\langle F_j, Z(x) \rangle$.

The SDP relaxation is used to solve these problems and provide an upper bound of the solution [10]. The result of the relaxation treats $Z(x)$ as a matrix variable $Z$ that will satisfy $Z \succeq 0$. Notice that each optimization problem is feasible simply by setting $Z$ and $\lambda$ to be zero. However, if the optimal solution of both problems is $Z = 0$ in the SDP relaxation, then we will conclude that the given candidate is a CLF.
Lemma 3.5: If the relaxed optimization problems in Eqs. (10) and (11) yield a zero solution, then the given candidate \( V_j(x) \) is in fact a CLF.

However, the converse is not true. It is possible for \( Z > 0 \) to be optimal for either relaxed condition, but no \( x \in \mathbb{R}^n \) corresponds to the solution. This happens because the relaxation drops two key constraints to convexify the conditions:

1) \( Z \) has to be a rank one matrix written as \( Z = mm^T \) and
2) there is an \( x \in \mathbb{R}^n \) such that \( m \) is the matrix of monomials corresponding to \( x \).

To deal with this, we adapt our learning framework to work with witnesses \( W_j : \{(Z_i, u_i)\}_{i=1}^j \) replacing states \( x_i \) by matrices \( Z_i \).

1) Each basis function \( g_j(x) \) in \( g \) is now written instead as \( (G_j, Z) \). The candidates are therefore, \( \sum_{j=1}^{n} c_j (G_j, Z) \).

Likewise, we write the components of its Lie derivative \( \nabla g_j : f_i \) in terms of \( Z \).

2) The learner maintains the set \( W \) as \( \{(Z_j, u_j)\} \), wherein \( Z_j \) is the feasible solution returned by the SDP solver while solving Eqs. (11) and (10). In other words, the CLF conditions are, in fact, taken to be the relaxed conditions.

3) We use a suitable projection operator \( \pi \) mapping each \( Z \) to a state \( x : \pi(Z) \), such that the demonstrator receives \( x \). In practice, since the vector of monomials used to define \( Z \) from \( x \) includes the terms \( 1, x_1, \ldots, x_n \), the projection operator simply selects a few entries from \( Z \) corresponding to the variables. Other more sophisticated projections are also possible.

The relaxed framework thus lifts counterexamples to work over matrices \( Z_j \). However, the candidate space begins with \( C \) and is refined each step as before. I.e., the relaxed framework continues to satisfy Lemmas 3.2, 3.3 Theorem 4 and Theorem 2 with the definition of (control) Lyapunov function changed to relaxed conditions.

IV. EXPERIMENTS

In this section, we describe numerical results on some example benchmark systems. The algorithmic framework is implemented using quadratic template forms for the CLFs with the tool Globoptipoly used to implement the verifier [13]. The demonstrator is implemented using a nonlinear MPC implemented using a gradient descent algorithm. For each benchmark, we tuned the time horizon, discretization step and the cost function until the control objectives were satisfied by the MPC over hundreds of simulations starting from randomly selected initial states.

Most of the benchmarks consider a reach-while-stay problem, wherein the goal is to reach target set \( T \), starting initial set \( I \), while remaining in the safe set \( S \). We also illustrate an example involving a trajectory stabilization problem. All the computations are performed on a Mac Book Pro with 2.9 GHz Intel Core i7 processor and 16GB of RAM. The reported CLFs are rounded to 2 decimal points. The summary of results is provided in Table 1.

**Bicycle Model:** This system is two-wheeled mobile robot modeled with five states \([x, y, v, \theta, \gamma] \) and two control inputs \([u, \tau] \). Each basis function \( g_j(x) \) in \( g \) is now written instead as \( (G_j, Z) \). The candidates are therefore, \( \sum_{j=1}^{n} c_j (G_j, Z) \). Likewise, we write the components of its Lie derivative \( \nabla g_j : f_i \) in terms of \( Z \).

The learner maintains the set \( W \) as \( \{(Z_j, u_j)\} \), wherein \( Z_j \) is the feasible solution returned by the SDP solver while solving Eqs. (11) and (10). In other words, the CLF conditions are, in fact, taken to be the relaxed conditions.

3) We use a suitable projection operator \( \pi \) mapping each \( Z \) to a state \( x : \pi(Z) \), such that the demonstrator receives \( x \). In practice, since the vector of monomials used to define \( Z \) from \( x \) includes the terms \( 1, x_1, \ldots, x_n \), the projection operator simply selects a few entries from \( Z \) corresponding to the variables. Other more sophisticated projections are also possible.

The relaxed framework thus lifts counterexamples to work over matrices \( Z_j \). However, the candidate space begins with \( C \) and is refined each step as before. I.e., the relaxed framework continues to satisfy Lemmas 3.2, 3.3 Theorem 4 and Theorem 2 with the definition of (control) Lyapunov function changed to relaxed conditions.

Table I: Results on the Numerical Examples. \( n: \# \text{variables, } m: \# \text{control inputs, } \tau: \text{MPC time step, } H: \text{MPC time horizon, } D_V: \text{SDP relaxation degree bound for the Verifier, } \# \text{itr: } \# \text{of iterations, Time: total computation time (minutes).}

<table>
<thead>
<tr>
<th>System Name</th>
<th>( n )</th>
<th>( m )</th>
<th>( \tau )</th>
<th>( H )</th>
<th>( D_V )</th>
<th># itr</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Torus</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>4</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>Bicycle</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
<td>8</td>
<td>3</td>
<td>51</td>
<td>9</td>
</tr>
<tr>
<td>Bicycle \times 2</td>
<td>8</td>
<td>4</td>
<td>0.4</td>
<td>8</td>
<td>3</td>
<td>536</td>
<td>303</td>
</tr>
<tr>
<td>Inverted Pendulum</td>
<td>4</td>
<td>1</td>
<td>0.04</td>
<td>2</td>
<td>5</td>
<td>85</td>
<td>31</td>
</tr>
<tr>
<td>Forward Flight</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
<td>16</td>
<td>5</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>Hover Mode</td>
<td>6</td>
<td>2</td>
<td>0.4</td>
<td>16</td>
<td>4</td>
<td>213</td>
<td>163</td>
</tr>
<tr>
<td>Unicycle (Seg.1)</td>
<td>3+1*</td>
<td>2</td>
<td>0.1</td>
<td>2</td>
<td>4</td>
<td>41</td>
<td>15</td>
</tr>
<tr>
<td>Unicycle (Seg.2)</td>
<td>3+1*</td>
<td>2</td>
<td>0.1</td>
<td>3</td>
<td>4</td>
<td>31</td>
<td>7</td>
</tr>
</tbody>
</table>

* +1 refers to the time variable.

Fig. 6. A Schematic View of the Bicycle Model for Stabilizing to the Road.

puts \( g \). The goal is to stabilize the car to a target velocity \( v^* = 5 \), as shown in Fig. 6. We drop the variable \( x \) (since it is immaterial to our stabilization problem) and obtain a model with four state variables:

\[
\begin{bmatrix}
\dot{y} \\
\dot{\theta} \\
\dot{\gamma}
\end{bmatrix}
= \begin{bmatrix}
u \sin(\theta) \\
1 \\
u \sigma
\end{bmatrix}
U : \left[ -10, 10 \right] \times \left[ -10, 10 \right]
S : \left[ -2, 2 \right] \times \left[ 3, 7 \right] \times \left[ -1, 1 \right]^{2}
I : B_{0.4}(0)
T : B_{0.1}(0),
\]

where \( \sigma = \tan(\gamma) \) (see Fig. 6). The function is approximated with a polynomial of degree 1. The method finds the following CLF:

\[
V = + 0.42y^2 + 0.59y\theta + 2.57\theta^2 + 0.79y\sigma + 4.64\sigma\theta + 4.06\sigma^2 - 0.38y\theta + 1.46\nu\theta + 1.18\nu\sigma + 2.39\nu^2.
\]

**Inverted Pendulum on a Cart:** This example has applications in balancing two-wheeled robots [6] (cf. [1] for list of such robots). The system has four state variables \([x, \dot{x}, \theta, \dot{\theta}] \) and one input \( u \). The dynamics, after partial linearization, have the following form [24]:

\[
\begin{bmatrix}
\frac{\dot{x}}{\dot{\theta}}
\end{bmatrix}
= \begin{bmatrix}
4u + \frac{4l(M + m)g \tan(\theta) - 3mg \sin(\theta) \cos(\theta)}{4(M + m)^2 \sin^2(\theta)} \\
4l(M + m)g \sin(\theta) \cos(\theta)
\end{bmatrix}
\]

where \( m = 0.21, M = 0.815, g = 9.8 \) and \( l = 0.305 \). The trigonometric and rational functions are approximated with polynomials of degree 3. The sets are \( S : [-1, 1]^4, U : [-20, 20], I : B_{0.2}(x), T : B_{0.1}(x) \). We obtain the following
CLF:
\[ V = + 11.64\ddot{\theta}^2 + 45.93\dot{\theta}\theta + 85.47\dot{\theta}^2 + 12.15x\dot{\theta} + 36.57x\theta + 6.44x^2 + 15.07\ddot{x} + 33.06\dot{x}\theta + 8.98\dot{x}x + 6.09x^2. \]

**Forward Flight for Caltech Ducted Fan:** The Caltech ducted fan models the aerodynamics of a single wing of a thrust vectored, fixed wing aircraft [16]. This problem is to design forward flight control in which the angle of attack needs to be changed. The model of the system is carefully calibrated through wind tunnel experiments. The system has 4 states \([v, \gamma, \theta, q]\) and two control inputs: \(u\) and \(\delta_u\). The dynamics are:
\[
\begin{bmatrix}
\dot{m}\dot{v} \\
\dot{m}v\dot{\gamma} \\
\dot{\theta} \\
J\dot{\gamma}
\end{bmatrix}
= 
\begin{bmatrix}
-D(v, \alpha) - W\sin(\gamma) + u\cos(\alpha + \delta_u) \\
L(v, \alpha) - W\cos(\gamma) + u\sin(\alpha + \delta_u) \\
q \\
M(v, \alpha) - ulT\sin(\delta_u)
\end{bmatrix},
\]
where \(\alpha = \theta - \gamma\), and \(D, L,\) and \(M\) are polynomials in \(v\) and \(\alpha\). For full list of parameters, see [16]. The goal is to reach \(x^* = [6, 0, 0.1771, 0]\) as the steady state. We perform a translation so that the \(x^*\) is the origin in the new coordinate system.

\[
U : [0, 13.5] \times [-0.45, 0.45] ;\]
\[
S : [3, 9] \times [-0.75, 0.75] \times [-0.75, 0.75] \times [-2, 2] ;
\]
\[
I : \{(v, \gamma, \theta, q) | (4v)^2 + (10\gamma)^2 + (10\theta)^2 + (10q)^2 < 4^2\} ;
\]
\[
T : \{(v, \gamma, \theta, q) | (4v)^2 + (10\gamma)^2 + (10\theta)^2 + (10q)^2 < 0.5^2\}.
\]

First, we approximate \(v^{-1}\), \(\sin\) and \(\cos\) with polynomials of degree 1, 3 and 3 respectively, to get polynomial dynamics. However, the system is not affine in control. We replace \(u\) and \(\delta_u\) input variables with \(u_1 = u\sin(\delta_u)\) and \(u_2 = u\cos(\delta_u)\) and \(U\) is under-approximated with a polytope. These changes yield a polynomial control affine dynamics, which fits the description of our model. The procedure finds the following CLF:
\[
V = + 3.21q^2 + 2.18q\theta + 3.90\dot{\theta}^2 + 0.40q\dot{\gamma} - 0.15v\theta + 0.56q^2 + 1.78q\gamma - 1.42\gamma\theta - 0.11v\gamma + 3.90\gamma^2.
\]

**Hover Mode for Caltech Ducted Fan:** This problem is again taken from [16]. The goal is to keep the ducted fan in a hover mode. The system has 6 variables \(x, y, \theta, \dot{x}, \dot{y}, \dot{\theta}\) and two control inputs \(u_1, u_2\). The dynamics are
\[
\begin{bmatrix}
\dot{m}\ddot{x} \\
\dot{m}\ddot{y} \\
\dot{J}\ddot{\theta}
\end{bmatrix}
= 
\begin{bmatrix}
-d_c\ddot{x} + u_1\cos(\theta) - u_2\sin(\theta) \\
-d_c\ddot{y} + u_2\cos(\theta) + u_1\sin(\theta) - mg \\
r_{u_1}
\end{bmatrix},
\]
where \(m = 11.2, \quad g = 0.28, \quad J = 0.0462, \quad r = 0.156\) and \(d_c = 0.1\). The sets are:
\[
S : [-1, 1] \times [-1, 1] \times [-0.7, 0.7] \times [-1, 1]^3 ;
\]
\[
U : [-10, 10] \times [0, 10], \quad I : B_{0.25}(0), \quad T : B_{0.05}(0).
\]

**Synthesis of Lyapunov Functions from Data:** The problem of synthesizing Lyapunov functions for a control system by observing the states of the system in simulation has been investigated in the past by Topcu et al to learn Lyapunov functions along with the resulting basin of attraction [54]. Whereas the original problem is bilinear, the use of simulation data makes it easier to postulate states that belong to the region of attraction and therefore find Lyapunov functions that belong to this region by solving LMIs in each case. The application of this idea to larger black-box systems is demonstrated by Kapinski et al [20]. Our approach focuses on controller synthesis through learning a control Lyapunov function in a
Learning CLFs is well known to be hard, involving bilinear Artstein and the construction of a feedback law given a CLF concept learning in discrete spaces [11].

bounds on the number of queries, paralleling some results on simple yet powerful ideas from convex optimization to place attempt to validate the current candidate. (c) Finally, we use counterexample to query the demonstrator comes from a failed adversarial model. (b) Our framework is the CLF is correct with respect to the underlying dynamical problems to ensure that the resulting policy learned through many ways: (a) We represent our policies by CLFs which are well as how the policy is inferred. Our approach stands out in the interaction between the learner and the demonstrator; as the CLF is correct with respect to the underlying dynamical system and a specification of the correctness in temporal logic to solve the problem of controller design to ensure that the resulting trajectories of the closed loop satisfy the temporal specifications. Majority of the approaches are based on discretization of the state-space into cells to compute a discrete abstraction of the overall system [56, 29, 46, 36, 37, 22, 50, 43, 44, 15]. These techniques focus on a given dynamical system and a specification of the correctness in temporal logic to solve the problem of controller design to ensure that the resulting trajectories of the closed loop satisfy the temporal specifications. Majority of the approaches are based on discretization of the state-space into cells to compute a discrete abstraction of the overall system [56, 29, 46, 36, 37, 22]. A smaller set of approaches synthesize Lyapunov-like functions by solving nonlinear constraints either through branch-and-bound techniques or by sum-of-square (SOS) relaxation techniques [50, 43, 44]. In this paper, we use the Lyapunov function approach to synthesizing controllers. An alternative is to use occupation measures [42, 40, 26, 32]. These methods formulate an infinite dimensional problem to maximize the region of attraction and obtain a corresponding control law. This is relaxed to a sequence of finite dimensional SDPs [25]. Note however that the approach computes an over approximation of the finite time backward reachable set from the target and a corresponding control. Our framework here instead seeks an under approximation that yields a guaranteed controller.

Control Lyapunov functions were originally introduced by Artstein and the construction of a feedback law given a CLF was first given by Sontag [3, 49]. As such, the problem of learning CLFs is well known to be hard, involving bilinear matrix inequalities (BMIs) [51]. An equivalent approach involves solving bilinear problems simultaneously for a control law and a Lyapunov function certifying it [8, 51]. BMIs are well known to be NP-hard, and hard to solve for a feasible solution [13]. The common approach is to perform an alternating minimization by fixing one set of bilinear variables while minimizing in the other. Such an approach has poor guarantees in practice, often “getting stuck” on a saddle point that does not allow the technique to make progress in finding a feasible solution. To combat this, Majumdar et al. (ibid) use LQR controllers and their associated Lyapunov functions for the linearization of the dynamics as good initial seed solutions [31]. In contrast, our approach simply assumes a demonstrator in the form of a MPC controller that can be used to resolve the bilinearity. Furthermore, our approach does not encounter the local saddle point problem.

The use of the learning framework with a demonstrator distinguishes the approach in this paper from recently developed ideas based on formal synthesis [56, 29, 46, 36, 37, 22, 50, 43, 44, 15]. In the setting of this paper, however, the use of a MPC scheme as a demonstrator allows us to use faster LP solvers and provide convergence guarantees.

Learning from Demonstration: The idea of learning from demonstration has had a long history in robotics [2]. The overall framework uses a demonstrator that can in fact be a human operator [21] or a complex MPC-based control law [19, 57, 58, 34]. The approaches differ on the nature of the interaction between the learner and the demonstrator; as well as how the policy is inferred. Our approach stands out in many ways: (a) We represent our policies by CLFs which are polynomial. On one hand, these are much less powerful than approaches that use neural networks [57, 19], for instance. However, the advantage lies in our ability to solve verification problems to ensure that the resulting policy learned through the CLF is correct with respect to the underlying dynamical model. (b) Our framework is adversarial. The choice of the counterexample to query the demonstrator comes from a failed attempt to validate the current candidate. (c) Finally, we use simple yet powerful ideas from convex optimization to place bounds on the number of queries, paralleling some results on concept learning in discrete spaces [11].

VI. CONCLUSION AND FUTURE WORK

We have proposed an algorithmic learning framework for synthesizing CLFs using a demonstrator and demonstrated our approach on challenging numerical examples with 4-8 state variables. As future work, we are considering many directions including the extensions to noisy/erroneous demonstrations, using output feedback (rather than full state feedback) synthesis and allowing disturbances in our framework. We are also working on integrating our control framework with RRT-based path planning and implementing it on board robotic vehicles.

ACKNOWLEDGMENTS

This work was funded in part by NSF under award numbers SHF 1527075 and CPS 1646556. All opinions expressed are those of the authors and not necessarily of the NSF.
REFERENCES


[34] Igor Mordatch and Emanuel Todorov. Combining the benefits of function approximation and trajectory optimization. In Robotics: Science and Systems (RSS), 2014.


