Polynomial Approximation of Value Functions and Nonlinear Controller Design with Performance Bounds

Morgan Jones, Matthew M. Peet

***Abstract*—For any suitable Optimal Control Problem (OCP) which satisfies the Principle of Optimality, there exists a value function, defined as the unique viscosity solution to a Hamilton Jacobi Bellman (HJB) equation, and which can be used to design an optimal feedback controller for the given OCP. Unfortunately, solving the HJB analytically is rarely possible, and existing numerical approximation schemes largely rely on discretization - implying that the resulting approximate value functions may not have the useful property of being uniformly less than or equal to the true value function (ie be sub-value functions). Furthermore, controllers obtained from such schemes currently have no associated bound on performance. To address these issues, for a given OCP, we propose a sequence of Sum-Of-Squares (SOS) programming problems, each of which yields a polynomial sub-solution to the HJB PDE, and show that the resulting sequence of polynomial sub-solutions converges to the value function of the OCP in the L^1 norm. Furthermore, for each polynomial sub-solution in this sequence we define an associated sublevel set, and show that the resulting sequence of sublevel sets converges to the sub-level set of the value function of the OCP in the volume metric. Next, for any approximate value function, obtained from an SOS program or any other method (e.g. discretization), we construct an associated feedback controller, and show that sub-optimality of this controller as applied to the OCP is bounded by the distance between the approximate and true value function of the OCP in the $W^{1,\infty}$ (Sobolev) norm. Finally, we demonstrate through several numerical examples how by solving our proposed SOS programing problem we are able to accurately approximate value functions, design controllers and estimate reachable sets.**

I. Introduction

Consider a nested family of Optimal Control Problems (OCPs), each initialized by $(x_0,t_0) \in \mathbb{R}^n \times [0,T]$, and each an optimization problem of the form

$$(\mathbf{u}^*,x^*) \in \arg\inf_{\mathbf{u},x} \left\{ \int_{t_0}^{T} c(x(t),\mathbf{u}(t),t)dt + g(x(T)) \right\} \text{ subject to,}$$
$$\dot{x}(t)=f(x(t),\mathbf{u}(t)),\ \mathbf{u}(t)\in U, \text{ for all } t\in[t_0,T],\ x(t_0)=x_0. \quad (1)$$

From the principle of optimality, if $(\mathbf{u}^*,x^*)$ solve the OCP for (x_0,t_0), then $(\mathbf{u}^*,x^*)$ also solve the OCP for $(x^*(t),t)$ for any $t \in [t_0,T]$. This can be used to show [1] that if a function, V, satisfies the Hamilton Jacobi Bellman (HJB) Partial Differential Equation (PDE), defined as

$$\nabla_t V(x,t) + \inf_{u\in U} \left\{ c(x,u,t) + \nabla_x V(x,t)^T f(x,u) \right\} = 0 \quad \text{for all } (x,t) \in \mathbb{R}^n \times (0,T), \quad (2)$$
$$V(x,T) = g(x) \quad \text{for all } x \in \mathbb{R}^n,$$

then necessary and sufficient conditions for $(\mathbf{u}^*,x^*)$ to solve OCP (1) initialized by (x_0,t_0) are

$$\mathbf{u}^*(t) = k(x^*(t),t),\ \dot{x}^*(t) = f(x^*(t),\mathbf{u}^*(t)), \text{ and } x^*(t_0) = x_0,$$
$$\text{where } k(x,t) \in \arg\inf_{u\in U} \left\{ c(x,u,t) + \nabla_x V(x,t)^T f(x,u) \right\}. \quad (3)$$

For a given family of OCPs of Form (1), if V satisfies (2), then V is called the Value Function (VF) of the OCP. If V is the VF, then for any (x,t), the value $V(x,t)$ determines the optimal objective value of OCP (1) initialized by (x,t). Furthermore, the VF yields a solution to the OCP (1) initialized by (x_0,t_0) through application of Eqn. (3). We call any $k : \Omega \times [0,T] \to U$ that satisfies Eqn. (3) a controller and we say this controller is the optimal controller for the OCP when V is the VF of the OCP.

Thus knowledge of the VF allows us to solve the nested family of OCPs in (1). Unfortunately, to find the VF, we must solve the HJB PDE, given in Eqn. (2), and this PDE has no analytic solution. In the absence of an analytic solution, we often parameterize a family of candidate VFs and search for one which satisfies the HJB PDE. However, this is a non-convex optimization problem since the HJB PDE is nonlinear. In this paper we view the search for a VF through the lens of convex optimization. Moreover, given an OCP, we are particularly interested in computing a sub-VF, a function that is uniformly less than or equal to the VF of the OCP (ie a function $\tilde{V}$ such that $\tilde{V}(x,t) \le V(x,t)$ for all $(x,t) \in \mathbb{R}^n \times [0,T]$ where V is the VF of the OCP). We consider what happens when we relax the nonlinear equality constraints imposed by the HJB PDE to linear inequality constraints and tighten the optimization problem's feasible set to polynomials. In this context, given an OCP, we consider the following questions.

Q1: Can we pose a sequence of convex optimization problems, each yielding a polynomial sub-VF that can be made arbitrarily "close" to the VF of the OCP?

Q2: Can we bound the sub-optimality in performance of a controller constructed from some function V by the "distance" between V and the VF of the OCP?

A. *Q1: Convex formulations of the problem of finding VFs*

Over the past few decades there has been a substantial amount of numerical schemes proposed that yield functions that closely approximate the VF of a given OCP. Current state of the art numerical methods for approximating the solution to the HJB PDE include grid based algorithms. Such algorithms generate a grid/mesh over the state space. In this class of algorithms we include (mixed) finite elements methods such as [2] that proposes a numerical scheme that yields an approximate VF with error bounds in terms of the first order mixed L^2 norm (assuming the Cordes condition holds). Other grid based methods include PDE discretization [3], [4] where [3] proposes a numerical scheme that yields an approximate VF with error bounds in terms of the L^∞ norm.

M. Jones is with the School for the Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85298 USA. e-mail: morgan.c.jones@asu.edu

M. Peet is with the School for the Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85298 USA. e-mail: mpeet@asu.edu

Alternative non-grid based algorithms include the method of characteristics [1] that can be used to compute evaluations of VF at fixed $(x,t) \in \mathbb{R}^n$, and max-plus methods [5]. In [5] a numerical scheme that approximates the VF of OCPs with linear dynamics and cost functions that are the point wise maximum of quadratic forms is proposed with a uniform error bound over the entire state space.

We note that none of the aforementioned state of the art numerical methods approximate the VF of a given OCP by a single polynomial sub-VF, unlike the methods later proposed in this paper. A major advantage of having a polynomial approximation, P, of the VF for some given OCP, is that $\nabla_x P$ can be efficiently computed (a useful property for solving the controller synthesis Eqn. (3)). Moreover, the advantage of having a sub-VF approximation is that the sublevel set of any sub-VF is guaranteed to contain the sublevel set of the true VF (see Cor. 1), leading to outer approximations of reachable sets (a useful property for safety analysis).

Alternative works on **Q1** includes the use of the moment-based dual to Sum-of-Squares (SOS) programming in [6], [7], [8]. Another duality-based approach, found in [9], considers a density-based dual to the VF and uses finite elements method to iteratively approximate the density and value function.

More narrowly focused work on **Q1** considers the problem of estimation of the reachable set (Defn. 10); the set of states that can be reached by the solution map from a given set of initial conditions. Specifically, value functions can be used to compute reachable sets. Likewise, candidate VFs can be used to approximate reachable sets. The link between solutions of the HJB PDE and reachable sets was first shown in [10] (which solved the HJB PDE via discretization).

Substantial work on SOS relaxations of the HJB PDE for reachable set estimation includes the carefully constructed optimization problems in [11], [12], [13] and includes, of course, our work in [14], [15]. However, there seems to be no prior work on using approximation theory to prove bounds on the sub-optimality of either controllers or corresponding reachable sets for SOS relaxations of the HJB PDE. We note, however, that [13] did establish *existence* of a close polynomial sub-solution the HJB in the framework of reachable sets.

In this paper we answer **Q1** by considering "sub-solutions" to the HJB PDE (2). Specifically, a "sub-solution", $\tilde{V}$, to the HJB PDE (2) satisfies the relaxed inequality constraint

$$\nabla_t \tilde{V}(x,t) + c(x,u,t) + \nabla_x \tilde{V}(x,t)^T f(x,u) \geq 0 \qquad (4)$$

for all $u \in U$ and $(x,t) \in \mathbb{R}^n \times [0,T]$, which implies that if V is a VF, $\tilde{V}(x,t) \leq V(x,t)$ - i.e. $\tilde{V}$ is a sub-VF. Then given an OCP (1) and based on this relaxed version of the HJB PDE (4), we propose a sequence of SOS programming problems, indexed by the degree $d \in \mathbb{N}$ of the polynomial variables, and given in Eqn. (61). The solution to each instance of the proposed sequence of optimization problems yields a polynomial P_d that is a sub-solution to the HJB PDE (2) (or sub-VF). We then show in Prop. 5 that for any value function V associated with the given OCP we have,

$$\lim_{d \to \infty} \|P_d - V\|_{L^1} = 0.$$

Furthermore, in Prop. 6 we show that this implies that the sublevel sets of $\{P_d\}_{d \in \mathbb{N}}$ converge to the sublevel sets of any value function, V (respect to the volume metric).

Our proposed method of approximately solving the HJB PDE by solving an SOS programming problem is implemented via Semi-Definite Programming (SDP). SDP problems can be solved to arbitrary accuracy in polynomial time using interior point methods [16]. However, the number of variables in the SDP problem associated with an n-dimensional and d-degree SOS problem is of the order n^{2d} [17], and therefore exponentially increases as $d \to \infty$. Fortunately there does exist several methods that improve the scalability of SOS [17], [18] but we do not discus such methods in this paper.

B. **Q2**: Performance bounds for controllers constructed from approximate VFs

To the best of our knowledge there exists no prior work on controller sub-optimality performance bounds, for OCPs of Form (1), in terms of the "closeness" of a candidate VF, used to construct the controller, and the true VF of the OCP. However, a performance bound for discrete time systems over infinite time horizons can be found in [19].

In Sec. VIII we answer **Q2** by showing that for any V, we can construct a candidate solution to the OCP (1), $\mathbf{u}(t) = k(x(t),t)$, given by the controller defined in Eqn. (3). We then show in Thm. 4 that the corresponding objective value of the OCP (1) evaluated at $\mathbf{u}$ is within $C\|V^* - V\|_{W^{1,\infty}}$ of the optimal objective, where V^* is the true VF of the OCP and $C > 0$ is given in Eqn. (70). Note, this result may be of broad interest since it does not require V to be a solution to our proposed SOS Problem (61) and hence provides a bound on the sub-optimality of controllers constructed from any approximate VF.

II. Notation

A. Standard Notation

We define $\text{sign}: \mathbb{R} \to \{-1,1\}$ and for $A \subset \mathbb{R}^n$ $\mathbb{1}_A : \mathbb{R}^n \to \mathbb{R}$ by $\text{sign}(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ -1 \text{ otherwise} \end{cases}$ and $\mathbb{1}_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ otherwise.} \end{cases}$ For two sets $A, B \subset \mathbb{R}^n$ we denote $A/B = \{x \in A : x \notin B\}$. For $B \subseteq \mathbb{R}^n$, $\mu(B) := \int_{\mathbb{R}^n} \mathbb{1}_B(x) dx$ is the Lebesgue measure of B, and for $X \subseteq \mathbb{R}^n$ and a function $f : X \to \mathbb{R}$ we denote the essential infimum by $\text{ess}\inf_{x \in X} f(x) := \sup\{a \in \mathbb{R} : \mu(\{x \in X : f(x) < a\}) = 0\}$. Similarly we denote the essential supremum by $\text{ess}\sup_{x \in X} f(x) := \inf\{a \in \mathbb{R} : \mu(\{x \in X : f(x) > a\}) = 0\}$. For $x \in \mathbb{R}^n$ we denote the Euclidean norm by $||x||_2 = \sqrt{\sum_{i=1}^n x_i^2}$. For $r > 0$ and $x \in \mathbb{R}^n$ we denote the ball $B(x,r) := \{y \in \mathbb{R}^n : ||x-y||_2 < r\}$. For an open set $\Omega \subset \mathbb{R}^n$ we denote the boundary of the set by $\partial \Omega$ and denote the closure of the set by $\overline{\Omega}$. Let $C(\Omega, \Theta)$ be the set of continuous functions with domain $\Omega \subset \mathbb{R}^n$ and image $\Theta \subset \mathbb{R}^m$. For an open set $\Omega \subset \mathbb{R}^n$ and $p \in [1, \infty)$ we denote the set of p-integrable functions by $L^p(\Omega, \mathbb{R}) := \{f : \Omega \to \mathbb{R} \text{ measurable } : \int_U |f|^p < \infty\}$, in the case $p = \infty$ we denote $L^\infty(\Omega, \mathbb{R}) := \{f : \Omega \to \mathbb{R} \text{ measurable } : \text{ess}\sup_{x \in \Omega} |f(x)| < \infty\}$. For $\alpha \in \mathbb{N}^n$ we denote the partial derivative $D^\alpha f(x) := \Pi_{i=1}^n \frac{\partial^{\alpha_i} f}{\partial x_i^{\alpha_i}}(x)$ where by convention if $\alpha = [0,..,0]^T$ we denote $D^\alpha f(x) := f(x)$. We denote the set of i'th continuously differentiable functions by $C^i(\Omega, \Theta) := \{f \in$

$C(\Omega,\Theta): D^\alpha f \in C(\Omega,\Theta)$ for all $\alpha \in \mathbb{N}^n$ such that $\sum_{j=1}^n \alpha_j \leq i\}$. For $k \in \mathbb{N}$ and $1 \leq p \leq \infty$ we denote the Sobolev space of functions with weak derivatives (Defn. 9) by $W^{k,p}(\Omega,\mathbb{R}) := \{u \in L^p(\Omega,\mathbb{R}) : D^\alpha u \in L^p(\Omega,\mathbb{R}) \text{ for all } |\alpha| \leq k\}$. For $u \in W^{k,p}(\Omega,\mathbb{R})$ we denote the Sobolev norm $||u||_{W^{k,p}(\Omega,\mathbb{R})} := \begin{cases} \left(\sum_{|\alpha|\leq k}\int_\Omega (D^\alpha u(x))^p dx\right)^{\frac{1}{p}} & \text{if } 1 \leq p < \infty \\ \sum_{|\alpha|\leq k} \operatorname{ess\,sup}_{x\in\Omega}\{|D^\alpha u(x)|\} & \text{if } p = \infty. \end{cases}$ In the case $k=0$ we have $W^{0,p}(\Omega,\mathbb{R}) = L^p(\Omega,\mathbb{R})$ and thus we use the notation $||\cdot||_{L^p(\Omega,\mathbb{R})} := ||\cdot||_{W^{0,p}(\Omega,\mathbb{R})}$. We denote the shift operator $\tau_s : L_2([0,T],\mathbb{R}^m) \to L_2([0,T-s],\mathbb{R}^m)$, where $s \in [0,T]$, and defined by $(\tau_s \mathbf{u})(t) := \mathbf{u}(s+t)$ for all $t \in [0,T-s]$.

B. Non-Standard Notation

We denote the set of locally and uniformly Lipschitz continuous functions on Θ_1 and Θ_2, Defn. 1, by $LocLip(\Theta_1,\Theta_2)$ and $Lip(\Theta_1,\Theta_2)$ respectively. Let us denote bounded subsets of $\mathbb{R}^n$ by $\mathscr{B} := \{B \subset \mathbb{R}^n : \mu(B) < \infty\}$. If M is a subspace of a vector space X we denote equivalence relation $\sim_M$ for $x,y \in X$ by $x \sim_M y$ if $x - y \in M$. We denote quotient space by $X \pmod M := \{\{y \in X : y \sim_M x\} : x \in X\}$. For an open set $\Omega \subset \mathbb{R}^n$ and $\sigma > 0$ we denote $<\Omega>_\sigma := \{x \in \Omega : B(x,\sigma) \subset \Omega\}$. For $V \in C^1(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ we denote $\nabla_x V := (\frac{\partial V}{\partial x_1},....,\frac{\partial V}{\partial x_n})^T$ and $\nabla_t V = \frac{\partial V}{\partial x_{n+1}}$. We denote the space of polynomials $p : \Omega \to \Theta$ by $\mathscr{P}(\Omega,\Theta)$ and polynomials with degree at most $d \in \mathbb{N}$ by $\mathscr{P}_d(\Omega,\Theta)$. We say $p \in \mathscr{P}_d(\mathbb{R}^n,\mathbb{R})$ is Sum-of-Squares (SOS) if there exists $p_i \in \mathscr{P}_d(\mathbb{R}^n,\mathbb{R})$ such that $p(x) = \sum_{i=1}^k (p_i(x))^2$. We denote $\sum_{SOS}^d$ to be the set SOS polynomials of at most degree $d \in \mathbb{N}$ and the set of all SOS polynomials as $\sum_{SOS}$. We denote $Z_d : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{\mathscr{N}_d}$ as the vector of monomials of degree $d \in \mathbb{N}$ or less and of size $\mathscr{N}_d := \binom{d+n}{d}$.

III. Optimal Control Problems

The nested family of finite-time Optimal Control Problems (OCPs), each initialized by $(x_0,t_0) \in \mathbb{R}^n \times [0,T]$, are defined as:

$$(\mathbf{u}^*,x^*) \in \arg\inf_{\mathbf{u},x}\left\{\int_{t_0}^T c(x(t),\mathbf{u}(t),t)dt + g(x(T))\right\} \text{ subject to,}$$
$$\dot{x}(t) = f(x(t),\mathbf{u}(t)) \quad \text{for all } t \in [t_0,T], \tag{5}$$
$$(x(t),\mathbf{u}(t)) \in \Omega \times U \quad \text{for all } t \in [t_0,T], \quad x(t_0) = x_0,$$

where $c : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ is referred to as the running cost; $g : \mathbb{R}^n \to \mathbb{R}$ is the terminal cost; $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is the vector field; $\Omega \subset \mathbb{R}^n$ is the state constraint set; $U \subset \mathbb{R}^m$ is the input constraint set; and T is the final time. For a given family of OCPs of Form (5) we associate the tuple $\{c,g,f,\Omega,U,T\}$.

In this paper we consider a special class of OCPs of Form (5), where U is compact and c,g,f are locally Lipschitz continuous. We next recall the definition of local Lipschitz continuity.

Definition 1. *Consider sets $\Theta_1 \subset \mathbb{R}^n$ and $\Theta_2 \subset \mathbb{R}^m$. We say the function $F : \Theta_1 \to \Theta_2$ is **locally Lipschitz continuous** on Θ_1 and Θ_2, denoted $F \in LocLip(\Theta_1,\Theta_2)$, if for every compact set $X \subseteq \Theta_1$ there exists $K_X > 0$ such that for all $x,y \in X$*

$$||F(x) - F(y)||_2 \leq K_X||x-y||_2. \tag{6}$$

*If there exists $K > 0$ such that Eqn. (6) holds for all $x,y \in \Theta_1$ we say F is **uniformly Lipschitz continuous**, denoted $F \in Lip(\Theta_1,\Theta_2)$.*

Definition 2. *We say the six tuple $\{c,g,f,\Omega,U,T\}$ is a Family of Lipschitz OCPs of Form (5) or $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ if:*
1) *$c \in LocLip(\Omega \times U \times [0,T],\mathbb{R})$.*
2) *$g \in LocLip(\Omega,\mathbb{R})$.*
3) *$f \in LocLip(\Omega \times U,\mathbb{R})$.*
4) *$U \subset \mathbb{R}^m$ is compact.*

For $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$, if $\Omega = \mathbb{R}^n$ we say the family of associated OCPs is *state unconstrained*, and if $\Omega \neq \mathbb{R}^n$ we say the associated family of OCPs is *state constrained*.

IV. Value Functions Can Solve OCP's

In the following subsections, we establish that for every family of Lipschitz OCPs, as defined in Section III, there exists a function, called the Value Function (VF), which:

(A) Is determined by the solution map - Eqn. (10).
(B) Solves the Hamilton-Jacobi-Bellman (HJB) Partial Differential Equation (PDE) - Eqn. (12).
(C) Can be used to construct a solution to the OCP.

A. Value Functions Are Determined By The Solution Map

Consider a nonlinear Ordinary Differential Equation (ODE) of the form

$$\dot{x}(t) = f(x(t),\mathbf{u}(t)), \quad x(0) = x_0, \tag{7}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $\mathbf{u} : \mathbb{R} \to \mathbb{R}^m$, and $x_0 \in \mathbb{R}^n$.

Definition 3. *We say the function ϕ_f is a solution map of the ODE given in Eqn. (7) on $[0,T] \subset \mathbb{R}$ if for all $t \in [0,T]$*

$$\frac{\partial \phi_f(x_0,t,\mathbf{u})}{\partial t} = f(\phi_f(x_0,t,\mathbf{u}),\mathbf{u}(t)), \text{ and } \phi_f(x_0,0,\mathbf{u}) = x_0.$$

Definition of Admissible Inputs: Given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and associated family of OCPs of Form (5), we now use the solution map to define the set of admissible input signals for the OCP initialized at $(x_0,t_0) \in \Omega \times [0,T]$. For this we use the shift operator, denoted $\tau_s : L_2([0,T],\mathbb{R}^m) \to L_2([0,T-s],\mathbb{R}^m)$, where $s \in [0,T]$, and defined by

$$(\tau_s \mathbf{u})(t) := \mathbf{u}(s+t) \text{ for all } t \in [0,T-s]. \tag{8}$$

Definition 4. *For any $(x_0,t_0) \in \mathbb{R}^n \times [0,T]$, we say $\mathbf{u}$ is admissible, denoted $\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x_0,t_0)$, if $\mathbf{u} : [t_0,T] \to U$ and there exists a unique solution map, ϕ_f, such that*

$$\frac{\partial \phi_f(x_0,t-t_0,\tau_{t_0}\mathbf{u})}{\partial t} = f(\phi_f(x_0,t-t_0,\tau_{t_0}\mathbf{u}),\mathbf{u}(t)) \text{ for } t \in [t_0,T],$$
$$\phi_f(x_0,t-t_0,\tau_{t_0}\mathbf{u}) \in \Omega \text{ for } t \in [t_0,T], \text{ and } \phi_f(x_0,0,\tau_{t_0}\mathbf{u}) = x_0. \tag{9}$$

For a given family of OCPs of Form (5), we now define the associated VF using the solution map, ϕ_f. Lemma 1 then shows that VFs are locally Lipschitz continuous.

Definition 5. *For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ we say $V^* : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a Value Function (VF) of the associated family of OCPs if for $(x,t) \in \Omega \times [0,T]$, the following holds*

$$V^*(x,t) = \inf_{\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x,t)}\left\{ \int_t^T c(\phi_f(x,s-t,\tau_t\mathbf{u}),\mathbf{u}(s),s)ds + g(\phi_f(x,T-t,\tau_t\mathbf{u}))\right\}, \tag{10}$$

where ϕ_f is as in Eqn. (9). By convention if $\mathscr{U}_{\Omega,U,f,T}(x,t) = \emptyset$ then $V^*(x,t) = \infty$.

Lemma 1 ([20], Local Lipschitz continuity of VF). *Consider some $\{c,g,f,\mathbb{R}^n,U,T\} \in \mathscr{M}_{Lip}$. Then if V^* satisfies Eqn.* (10)*, we have that $V^* \in LocLip(\mathbb{R}^n \times [0,T],\mathbb{R})$.*

B. Value Functions are Solutions to the HJB PDE

Consider the family of OCPs associated with $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$. As shown in [21], a sufficient condition for a function V^* to be a VF, is for V^* to satisfy the Hamilton Jacobi Bellman (HJB) PDE, given in Eqn. (12). However, for a general family of OCPs of form $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$, solutions to the HJB PDE may not be differentiable, and hence classical solutions to the HJB PDE may not exist. For this reason, one typically uses a generalized notion of a solution to the HJB PDE called a viscosity solution, which is defined in [22] as follows.

Definition 6. *Consider the first order PDE*

$$F(x,y(x),\nabla y(x)) = 0 \quad \text{for all } x \in \Omega, \tag{11}$$

where $\Omega \subset \mathbb{R}^n$ and $F \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n,\mathbb{R})$.

*We say $y \in C(\Omega)$ is a **viscosity sub-solution** of* (11) *if*

$$F(x,y(x),p) \le 0 \quad \text{for all } x \in \Omega \text{ and } p \in D^+y(x),$$

where $D^+y(x) := \{p \in \mathbb{R} : \exists \Phi \in C^1(\Omega,\mathbb{R})$ such that $\nabla\Phi(x) = p$ and $y-\Phi$ attains a local max at $x\}$.

*Similarly, $y \in C(\Omega)$ is a **viscosity super-solution** of* (11) *if*

$$F(x,y(x),p) \ge 0 \quad \text{for all } x \in \Omega \text{ and } p \in D^-y(x)$$

where $D^-y(x) := \{p \in \mathbb{R} : \exists \Phi \in C^1(\Omega,\mathbb{R})$ such that $\nabla\Phi(x) = p$ and $y-\Phi$ attains a local min at $x\}$.

*We say $y \in C(\Omega)$ is a **viscosity solution** of* (11) *if it is both a viscosity sub and super-solution.*

Theorem 1 ([20], Uniqueness of VF). *Consider the family of OCPs associated with the tuple $\{c,g,f,\mathbb{R}^n,U,T\} \in \mathscr{M}_{Lip}$. Any function satisfying Eqn.* (10) *is the unique viscosity solution of the HJB PDE*

$$\nabla_t V(x,t) + \inf_{u\in U}\left\{c(x,u,t) + \nabla_x V(x,t)^T f(x,u)\right\} = 0 \quad \text{for all } (x,t) \in \mathbb{R}^n \times [0,T] \tag{12}$$
$$V(x,T) = g(x) \quad \text{for all } x \in \mathbb{R}^n.$$

Note that Lemma 1 and Theorem 1 are only valid in the absence of state constraints ($\Omega = \mathbb{R}^n$). However, as we will show in Lemma 3, if the state constraints are sufficiently "loose", then the unconstrained and constrained solutions coincide.

C. VF's Can Construct Optimal Controllers

Given an OCP, we next show if a "classical" differentiable solution to the HJB PDE (12) associated with the OCP is known then a solution to the OCP can be constructed using Eqns. (13) and (14). We will refer to any $k : \Omega \times [0,T] \to U$ that satisfies Eqns. (13) and (14) for some V as a controller and say this is the optimal controller of the OCP if V is the VF of the OCP.

Theorem 2. *[1] Consider the family of OCPs associated with tuple $\{c,g,f,\mathbb{R}^n,U,T\} \in \mathscr{M}_{Lip}$. Suppose $V \in C^1(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ solves the HJB PDE* (12)*. Then $\mathbf{u}^* : [t_0,T] \to U$ solves the OCP associated with $\{c,g,f,\mathbb{R}^n,U,T\}$ initialized at $(x_0,t_0) \in \mathbb{R}^n \times [0,T]$ if and only if*

$$\mathbf{u}^*(t) = k(\phi_f(x_0,t,\mathbf{u}^*),t) \text{ for all } t \in [t_0,T], \tag{13}$$
$$\text{where } k(x,t) \in \arg\inf_{u\in U}\{c(x,u,t) + \nabla_x V(x,t)^T f(x,u)\}. \tag{14}$$

If the function V in Eqn. (14) is not a VF the resulting controller may no longer construct a solution to the OCP. In Section VIII we will provide a bound on the performance of a constructed controller from a candidate VF based on how "close" the candidate VF is to the true VF under the Sobolev norm.

V. The Feasibility Problem Of Finding VF's

Consider a family of OCPs associated with some $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$. Previously it was shown in Theorem 2 that if $V \in C^1(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ is a solution to the HJB PDE (12) then V may be used to solve the family of OCPs using Eqns. (13) and (14). The question, now, is how to find such a V.

Let us consider the problem of finding a value function as an optimization problem subject to constraints imposed by the HJB PDE (12). This yields the following feasibility problem:

$$\text{Find } V \in C^1(\mathbb{R}^n \times \mathbb{R},\mathbb{R}), \tag{15}$$
$$\text{such that } V \text{ satisfies (12).}$$

Note that our optimization problem of Form (15) is non-convex and may not even have a solution with sufficient regularity. For these reasons, we next propose a convex relaxation of Problem (15). We first define sub-VFs and super-VFs that uniformly bound VFs either from above or bellow.

Definition 7. *We say the function $J : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a sub-VF to the family of OCPs associated with $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ if*

$$J(x,t) \le V^*(x,t) \text{ for all } t \in [0,T] \text{ and } x \in \Omega,$$

for any V^* satisfying Eqn.(10)*. Moreover if*

$$J(x,t) \ge V^*(x,t) \text{ for all } t \in [0,T] \text{ and } x \in \Omega,$$

for any V^* satisfying Eqn. (10)*, we say J is a super-VF.*

A. A Sufficient Condition For A Function To Be A Sub-VF

We now propose "dissipation" inequalities and show that if a differentiable function satisfies such inequalities then it must be a sub-value function.

Proposition 1. *For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ suppose $J \in C^1(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ satisfies for all $(x,u,t) \in \Omega \times U \times (0,T)$*

$$\nabla_t J(x,t) + c(x,u,t) + \nabla_x J(x,t)^T f(x,u) \ge 0, \tag{16}$$
$$J(x,T) \le g(x). \tag{17}$$

Then J is a sub-value function of the family of OCPs associated with $\{c,g,f,\Omega,U,T\}$.

Proof. Suppose $J \in C^1(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ satisfies Eqns. (16) and (17). Consider an arbitrary $(x_0,t_0) \in \Omega \times [0,T]$. If $\mathscr{U}_{\Omega,U,f,T}(x_0,t_0) = \emptyset$ then $V^*(x_0,t_0) = \infty$. Clearly in this case $J(x_0,t_0) < V^*(x_0,t_0)$ as J is continuous and therefore is finite over the compact region $\Omega \times [0,T]$. Alternatively if

$\mathscr{U}_{\Omega,U,f,T}(x_0,t_0) \neq \emptyset$, then for any $\tilde{\mathbf{u}} \in \mathscr{U}_{\Omega,U,f,T}(x_0,t_0)$, we have the following by Defn. 4:

$$\phi_f(x_0, t-t_0, \tau_{t_0}\tilde{\mathbf{u}}) \in \Omega \text{ for all } t \in [t_0, T],$$
$$\tilde{\mathbf{u}}(t) \in U \text{ for all } t \in [t_0, T].$$

Therefore (using the shorthand $\tilde{x}(t) := \phi_f(x_0, t-t_0, \tau_{t_0}\tilde{\mathbf{u}})$), by Eqn. (16) we have for all $t \in [t_0, T]$

$$\nabla_t J(\tilde{x}(t),t) + c(\tilde{x}(t), \tilde{\mathbf{u}}(t), t) + \nabla_x J(\tilde{x}(t),t)^T f(\tilde{x}(t), \tilde{\mathbf{u}}(t)) \geq 0.$$

Now, using the chain rule we deduce

$$\frac{d}{dt} J(\tilde{x}(t),t) + c(\tilde{x}(t), \tilde{\mathbf{u}}(t), t) \geq 0 \text{ for all } t \in [t_0, T].$$

Then, integrating over $t \in [t_0, T]$, and since $J(\tilde{x}(T),T) \leq g(\tilde{x}(T))$ by Eqn. (17), we have

$$J(x_0,t_0) \leq \int_{t_0}^{T} c(\tilde{x}(t), \tilde{\mathbf{u}}(t), t)dt + g(\tilde{x}(T)). \tag{18}$$

Since Eqn. (18) holds for all $\tilde{\mathbf{u}} \in \mathscr{U}_{\Omega,U,f,T}(x_0,t_0)$, we may take the infimum over $\mathscr{U}_{\Omega,U,f,T}(x_0,t_0)$ to show that $J(x_0,t_0) \leq V^*(x_0,t_0)$. As this argument can be used for any $(x_0,t_0) \in \Omega \times [0,T]$ it follows J is a sub-value function. □

Definition 8. *For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ we say a function $J \in C^1(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ is dissipative if it satisfies Inequalities* (16) *and* (17)*.*

Dissipative functions are viscosity sub-solutions (as per Defn. 6) to the HJB PDE (12). Moreover, by Prop. 1 a dissipative function is a sub-VF. However, a sub-VF need not be dissipative or a viscosity sub-solution to the HJB PDE.

B. A Convex Relaxation Of The Problem Of Finding VF's

The set of functions satisfying Eqns. (16) and (17) is convex as Eqns. (16) and (17) are linear in terms of the unknown variable/function J. Furthermore, for given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$, any function which satisfies the HJB PDE (12) also satisfies Eqns. (16) and (17). This allows us to propose the following convex relaxation of the problem of finding a VF (Problem (15)):

$$\text{Find } J \in C^1(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}), \qquad \text{such that } J \text{ satisfies (16) and (17).} \tag{19}$$

C. A Polynomial Tightening Of The Problem Of Finding VF's

Problem (19) is convex. However, a function J, feasible for Problem (19) (and hence dissipative), may be arbitrarily far from the VF. For instance, in the case $c(x,u,t) \geq 0$ and $0 \leq g(x) < M$, the constant function $J(x,t) \equiv -C$ is dissipative for any $C > M$. Thus, by selecting sufficiently large enough $C > M$, we can make $||J - V||$ arbitrary large, regardless of the chosen norm, $||\cdot||$.

To address this issue, we propose a modification of Problem (19), wherein we include an objective of Form $\int_{\Lambda \times [0,T]} w(x,t)J(x,t)dxdt$, parameterized by a compact domain of interest $\Lambda \subset \mathbb{R}^n$ and weight $w \in L^1(\Lambda \times [0,T], \mathbb{R}^+)$ (we use the weight, w, in Prop. 6). Specifically, for given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and $d \in \mathbb{N}$, consider the optimization problem:

$$\begin{aligned} J_d \in \arg & \max_{J \in \mathscr{P}_d(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})} \int_{\Lambda \times [0,T]} w(x,t)J(x,t)dxdt \\ & \text{subject to: } \nabla_t J(x,t) + c(x,u,t) + \nabla_x J(x,t)^T f(x,u) > 0 \\ & \qquad \text{for all } x \in \Omega, t \in (0,T), u \in U, \\ & \qquad J(x,T) < g(x) \text{ for all } x \in \Omega. \end{aligned} \tag{20}$$

Minimizing $\int_{\Lambda \times [0,T]} w(x,t)J(x,t)dxdt$ minimizes the weighted L^1 norm $\int_{\Lambda \times [0,T]} w(x,t)|V(x,t) - J(x,t)|dxdt$. The restriction to polynomial solutions $J \in \mathscr{P}_d(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ makes the problem finite-dimensional.

VI. A Sequence Of Dissipative Polynomials That Converge To The VF In Sobolev Space

For a given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$, in Eqn. (20), we proposed a sequence of optimization problems, indexed by $d \in \mathbb{N}$, each instance of which yields a dissipative function $J_d \in \mathscr{P}_d(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$. In this section, we prove that $\lim_{d \to \infty} ||J_d - V||_{L_1} \to 0$ where V is the VF associated with the OCP $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$. To accomplish this proof, we divide the section into three subsections, wherein we find the following.

(A) In Prop. 3 we show that for every $\varepsilon > 0$ there exists a dissipative function $J_\varepsilon \in C^\infty(\Omega \times [0,T], \mathbb{R})$ such that $||J_\varepsilon - V||_{W^{1,p}(\Omega \times [0,T], \mathbb{R})} < \varepsilon$ for any $V \in Lip(\Omega \times [0,T], \mathbb{R})$ which satisfies the dissipation-type inequality in Eqn. (23).
(B) In Theorem 3 we show that for every $\varepsilon > 0$, there exists $d \in \mathbb{N}$ and dissipative $P_\varepsilon \in \mathscr{P}_d(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ such that $||P_\varepsilon - V||_{W^{1,p}(\Omega \times [0,T], \mathbb{R})} < \varepsilon$, for any value function, V, associated with $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$.
(C) For any positive weight w, Prop. 4 shows that if J_d solves (20) for $d \in \mathbb{N}$, then $\lim_{d \to \infty} ||w(J_d - V)||_{L^1} = 0$ for any VF, V, associated with $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$.

A. Existence Of Smooth Dissipative Functions That Approximate The VF Arbitrarily Well Under The $W^{1,p}$ Norm

In this section we create a sequence of smooth (elements of $C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$) functions that converges, with respect to the $W^{1,p}$ norm, to any Lipschitz function, V, satisfying the dissipation-type inequality in Eqn. (23). This subsection uses some aspects of mollification theory. For an overview of this field, we refer to [23].

a) Mollifiers: The standard mollifier, $\eta \in C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ is defined as

$$\eta(x,t) := \begin{cases} C \exp\left(\frac{1}{||(x,t)||_2^2 - 1}\right) & \text{when } ||(x,t)||_2 < 1, \\ 0 \quad \text{when } ||(x,t)||_2 \geq 1, \end{cases} \tag{21}$$

where $C > 0$ is chosen such that $\int_{\mathbb{R}^n \times \mathbb{R}} \eta(x,t)dxdt = 1$.

For $\sigma > 0$ we denote the scaled standard mollifier by $\eta_\sigma \in C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ such that

$$\eta_\sigma(x,t) := \frac{1}{\sigma^{n+1}} \eta\left(\frac{x}{\sigma}, \frac{t}{\sigma}\right).$$

Note, clearly $\eta_\sigma(x,t) = 0$ for all $(x,t) \notin B(0,\sigma)$.

b) Mollification of a Function (Smooth Approximation): Recall from Section II-B that for open sets $\Omega \subset \mathbb{R}^n$, $(0,T) \subset \mathbb{R}$, and $\sigma > 0$ we denote $< \Omega \times (0,T) >_\sigma := \{x \in \Omega \times (0,T) : B(x,\sigma) \subset \Omega \times (0,T)\}$. Now, for each $\sigma > 0$ and function $V \in L^1(\Omega \times (0,T), \mathbb{R})$ we denote the *σ-mollification* of V by $[V]_\sigma :< \Omega \times (0,T) >_\sigma \to \mathbb{R}$, where

$$[V]_\sigma(x,t) := \int_{\mathbb{R}^n \times \mathbb{R}} \eta_\sigma(x-z_1, t-z_2) V(z_1,z_2) dz_1 dz_2 \quad (22)$$
$$= \int_{B(0,\sigma)} \eta_\sigma(z_1,z_2) V(x-z_1, t-z_2) dz_1 dz_2.$$

To calculate the derivative of a mollification we next introduce the concept of weak derivatives.

Definition 9. *For $\Omega \subset \mathbb{R}^n$ and $F \in L^1(\Omega,\mathbb{R})$ we say any $H \in L^1(\Omega,\mathbb{R})$ is the weak $i \in \{1,..,n\}$-partial derivative of F if*

$$\int_\Omega F(x) \frac{\partial}{\partial x_i} \alpha(x) dx = -\int_\Omega H(x) \alpha(x) dx, \text{ for } \alpha \in C^\infty(\mathbb{R}^n, \mathbb{R}).$$

Weak derivatives are "essentially unique". That is if H_1 and H_2 are both weak derivatives of a function F then the set of points where $H_1(x) \neq H_2(x)$ has measure zero. If a function is differentiable then its weak derivative is equal to its derivative in the "classical" sense. We will use the same notation for the derivative in the "classical" sense and in the weak sense.

In the next proposition we state some useful properties about Sobolev spaces and mollifications taken from [23].

Proposition 2 ([23])**.** *For $1 \le p < \infty$ and $k \in \mathbb{N}$ we consider $V \in W^{k,p}(E,\mathbb{R})$, where $E \subset \mathbb{R}^n$ is an open bounded set, and its σ-mollification $[V]_\sigma$. Recalling from Section II-B that for an open set $\Omega \subset \mathbb{R}^n$ and $\sigma > 0$ we denote $< \Omega >_\sigma := \{x \in \Omega : B(x,\sigma) \subset \Omega\}$, the following holds:*

1) *For all $\sigma > 0$ we have $[V]_\sigma \in C^\infty(< E >_\sigma, \mathbb{R})$.*
2) *For all $\sigma > 0$ we have $\nabla_t [V]_\sigma(x,t) = [\nabla_t V]_\sigma(x,t)$ and $\nabla_x [V]_\sigma(x,t) = [\nabla_x V]_\sigma(x,t)$ for $(x,t) \in < E >_\sigma$, where $\nabla_t V$ and $\nabla_x V$ are weak derivatives.*
3) *If $V \in C(E,\mathbb{R})$ then for any compact set $K \subset E$ we have $\lim_{\sigma \to 0} \sup_{(x,t)\in K} |V(x,t) - [V]_\sigma(x,t)| = 0$.*
4) *(Meyers-Serrin Local Approximation) For any compact set $K \subset E$ we have $\lim_{\sigma \to 0} ||[V]_\sigma - V||_{W^{k,p}(K,\mathbb{R})} = 0$.*

c) Approximation of Lipschitz functions satisfying a dissipation-type inequality: We now show that for any Lipschitz function, V, satisfying the dissipation-type inequality in Eqn. (23), V can be approximated arbitrarily well by a smooth function, J_ε, that also satisfies the dissipation-type inequality in Eqn. (23). We use a similar proof strategy to (Lemma B5) [24].

Lemma 2. *Let $E \subset \mathbb{R}^{n+1}$ be an open bounded set, $\Omega \subset \mathbb{R}^n$ be such that $\Omega \times (0,T) \subseteq E$, where $T > 0$, $U \subset \mathbb{R}^m$ be a compact set, $f \in Lip(\Omega \times U, \mathbb{R}^n)$, $c \in Lip(\Omega \times U \times [0,T], \mathbb{R})$, and $V \in Lip(E,\mathbb{R})$ such that*

$$\operatorname*{ess\,inf}_{(x,t)\in\Omega\times(0,T)} \{\nabla_t V(x,t) + \nabla_x V(x,t)^T f(x,u) + c(x,u,t)\} \ge 0, \quad (23)$$

where the derivatives, $\nabla_t V$ and $\nabla_x V$, are weak derivatives.

Then for any compact set $K \subset E$, $1 \le p < \infty$ and for all $\varepsilon > 0$ there exits $J_\varepsilon \in C^\infty(K,\mathbb{R})$ such that

$$||V - J_\varepsilon||_{W^{1,p}(K,\mathbb{R})} < \varepsilon \text{ and } \sup_{(x,t)\in K} |V(x,t) - J_\varepsilon(x,t)| < \varepsilon, \quad (24)$$

and for all $(x,t) \in K \cap (\Omega \times (0,T))$ and $u \in U$

$$\nabla_t J_\varepsilon(x,t) + \nabla_x J_\varepsilon(x,t)^T f(x,u) + c(x,u,t) \ge -\varepsilon. \quad (25)$$

Proof. Suppose V satisfies Eqn. (23), $K \subset E$ is a compact set, $1 \le p < \infty$, and $\varepsilon > 0$. By Rademacher's Theorem (Theorem 7) V is weakly differentiable with essentially bounded derivative. Therefore $V \in W^{1,\infty}(E,\mathbb{R})$ and hence $V \in W^{1,p}(E,\mathbb{R})$. Now Prop. 2 (Statements 3 and 4) can be used to show there exists $\sigma_1 > 0$ such that for any $0 \le \sigma < \sigma_1$ we have

$$||V - [V]_{\sigma_1}||_{W^{1,p}(K,\mathbb{R})} < \varepsilon \text{ and } \sup_{(x,t)\in K} |V(x,t) - [V]_{\sigma_1}(x,t)| < \varepsilon. \quad (26)$$

Select $\sigma_2 > 0$ small enough so $K \subset< E >_{\sigma_2}$ (which can be done as E is open). Select $0 < \sigma_3 < \frac{\varepsilon}{L_V L_f + 2L_c}$, where $L_V, L_f, L_c > 0$ are the Lipschitz constant of the functions V, f, and c respectively. We now have the following for all $\sigma_4 < \min\{\sigma_3, \sigma_2\}$, $u \in U$ and $(x,t) \in K \cap (\Omega \times (0,T))$,

$$\begin{aligned}
&\nabla_t [V]_{\sigma_4}(x,t) + \nabla_x [V]_{\sigma_4}(x,t)^T f(x,u) + c(x,u,t) \quad (27)\\
&= [\nabla_t V]_{\sigma_4}(x,t) + [\nabla_x V]_{\sigma_4}(x,t)^T f(x,u) + c(x,u,t)\\
&= \int_{B(0,\sigma_4)} \eta_{\sigma_4}(z_1,z_2) \Big(\nabla_t V(x-z_1,t-z_2)\\
&\qquad + \nabla_x V(x-z_1,t-z_2)^T f(x-z_1,u) + c(x-z_1,u,t-z_2) \Big) dz_1 dz_2\\
&\qquad - \int_{B(0,\sigma_4)} \eta_{\sigma_4}(z_1,z_2) \nabla_x V(x-z_1,t-z_2)^T\\
&\qquad\qquad \Big(f(x-z_1,u) - f(x,u) \Big) dz_1 dz_2\\
&\quad - \int_{B(0,\sigma_4)} \eta_{\sigma_4}(z_1,z_2) \Big(c(x-z_1,u,t-z_2) - c(x,u,t) \Big) dz_1 dz_2\\
&\ge \operatorname*{ess\,inf}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ \nabla_t V(x-z_1,t-z_2)\\
&\qquad + \nabla_x V(x-z_1,t-z_2)^T f(x-z_1,u) + c(x-z_1,u,t-z_2) \Big\}\\
&\quad - \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ ||\nabla_x V(x-z_1,t-z_2)||_2 \Big\}\\
&\qquad \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ ||f(x-z_1,u) - f(x,u)||_2 \Big\}\\
&\quad - \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ |c(x-z_1,u,t-z_2) - c(x,u,t)| \Big\}\\
&\ge -L_V \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ ||f(x-z_1,u) - f(x,u)||_2 \Big\}\\
&\qquad - \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ |c(x-z_1,u,t-z_2) - c(x,u,t)| \Big\}\\
&\ge -L_V L_f \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ ||z_1||_2 \Big\} - L_c \operatorname*{ess\,sup}_{(z_1,z_2)\in B(0,\sigma_4)} \Big\{ ||z_1||_2 + |z_2| \Big\}\\
&= -(L_V L_f + 2L_c)\sigma_4 \ge -\varepsilon.
\end{aligned}$$

The first equality of Eqn. (27) follows since $\nabla_t [V]_{\sigma_4}(x,t) = [\nabla_t V]_{\sigma_4}(x,t)$ and $\nabla_x [V]_{\sigma_4}(x,t) = [\nabla_x V]_{\sigma_4}(x,t)$ for all $(x,t) \in K \subset< E >_{\sigma_4}$ by Prop. 2 (Statement 2). The first inequality follows by the monotonicity property of integration and the Cauchy Swartz inequality. Since V is Lipschitz $\operatorname{ess\,sup}_{(x,t)\in E} ||\nabla_x V(x,t)||_2 < L_V$ by Rademacher's Theorem (Theorem 7). Now the second inequality follows by using (23) together with $\operatorname{ess\,sup}_{(x,t)\in E} ||\nabla_x V(x,t)||_2 < L_V$. The third inequality follows by the Lipschitz continuity of f and c. Finally the fourth inequality follows by the fact $\sigma_4 < \sigma_3 < \frac{\varepsilon}{L_V L_f + L_c}$.

Now define $J_\varepsilon(x,t) := [V]_\sigma(x.t)$ where $0 < \sigma < \min\{\sigma_1, \sigma_4\}$. It follows that $J_\varepsilon \in C^\infty(K,\mathbb{R})$ by Prop. 2 (Statement 1). Moreover J_ε satisfies Eqns. (24) and (25) by Eqns. (26) and (27). ☐

In Lemma 2 we showed that for any given function, $V :\in Lip(E,\mathbb{R})$, any compact subsets $K \subset E$, any $\varepsilon > 0$, and any $1 \le p < \infty$, there exists a smooth function, J_ε, satisfying Eqn. (25), such that $||V - J_\varepsilon||_{W^{1,p}(K,\mathbb{R})} < \varepsilon$. We next show this "local" result over compact subsets, K, can be extended to a "global" results over the entire domain, E. To do this we use Theorem 9, stated in Section XIII. Given an open cover of E, Theorem 9 states that there exists a family of functions, called a partition of unity. In the next proposition we use partitions of unity together with the "local" approximates of the Lipschitz function, V, to construct a smooth "global" approximation of V over the entire domain E.

Proposition 3. *Let $E \subset \mathbb{R}^{n+1}$ be an open bounded set, $\Omega \subset \mathbb{R}^n$ be such that $\Omega \times (0,T) \subseteq E$, where $T > 0$, $U \subset \mathbb{R}^m$ be a compact set, $f \in Lip(\Omega \times U, \mathbb{R}^n)$, $c \in Lip(\Omega \times U \times [0,T], \mathbb{R})$, and $V \in Lip(E,\mathbb{R})$ satisfies Eqn. (23). Then for all $1 \le p < \infty$ and $\varepsilon > 0$ there exits $J \in C^\infty(E,\mathbb{R})$ such that*

$$||V - J||_{W^{1,p}(E,\mathbb{R})} < \varepsilon \text{ and } \sup_{(x,t)\in E} |V(x,t) - J(x,t)| < \varepsilon, \quad (28)$$

and for all $(x,u,t) \in \Omega \times U \times (0,T)$

$$\nabla_t J(x,t) + \nabla_x J(x,t)^T f(x,u) + c(x,u,t) \ge -\varepsilon. \quad (29)$$

Proof. Let us consider the family of sets $E_i = \{x \in E : \sup_{y \in \partial E} ||x - y||_2 < \frac{1}{i}\}$ for $i \in \mathbb{N}$. It follows $\{E_i\}_{i=1}^\infty$ is an open cover (Defn. 14) for E and thus by Theorem 9 there exists a smooth partition of unity, $\{\psi_i\}_{i=1}^\infty \subset C^\infty(E,\mathbb{R})$, that satisfies Statements 1 to 4 of Theorem 9.

For $\varepsilon > 0$ Lemma 2 shows that for each $i \in \mathbb{N}$ there exists a function $J_i \in C^\infty(\overline{E_i},\mathbb{R})$ such that

$$\sup_{(x,t)\in E_i} |V(x,t) - J_i(x,t)| < \frac{\varepsilon}{2^{i+1}(1+\tau_i+\theta_i)}, \quad (30)$$

$$||V - J_i||_{W^{1,p}(\overline{E_i},\mathbb{R})} < \frac{\varepsilon}{2^{i+1}(1+\tau_i+\theta_i)}, \quad (31)$$

$$\nabla_t J_i(x,t) + \nabla_x J_i(x,t)^T f(x,u) + c(x,u,t) \ge -\frac{\varepsilon}{2^{i+1}(1+\tau_i+\theta_i)}$$
$$\forall (x,t) \in \overline{E_i} \cap (\Omega \times (0,T)), u \in U, \quad (32)$$

where we denote $\tau_i := \sup_{(x,u,t)\in\Omega\times U\times(0,T)}\{|\nabla_t \psi_i(x,t) + \nabla_x \psi_i(x,t)^T f(x,u)|\} \ge 0$ and $\theta_i := \left(\max_{|\alpha|\le 1} \sup_{(x,t)\in E} |D^\alpha \psi_i(x,t)|^p\right)^p \ge 0$; which is well defined and finite as $\Omega \times U \times (0,T)$ is bounded and ψ_i is smooth.

Now, let us define $J(x,t) := \sum_{i=1}^\infty \psi_i(x,t) J_i(x,t)$, we will show $J \in C^\infty(E,\mathbb{R})$ and that J satisfies Eqns. (28) and (29).

It follows $J \in C^\infty(E,\mathbb{R})$ by Theorem 9. To see this we note for each $i \in \mathbb{N}$ we have $\psi_i \in C^\infty(E,\mathbb{R})$ and $\psi_i(x,t) = 0$ outside E_i implying $\psi_i J_i \in C^\infty(E,\mathbb{R})$. Moreover, for each $(x,t) \in E$ there exists an open set, $S \subseteq E$, where only a finite number of ψ_i are nonzero. Therefore it follows that the function J is a finite sum of infinitely differentiable functions and thus J is also infinitely differentiable.

We now show J satisfies Eqn. (28). We first show $\|V - J\|_{W^{1,p}(E,\mathbb{R})} < \varepsilon$:

$$\begin{aligned}\|V - J\|_{W^{1,p}(E,\mathbb{R})} &= \|V - \sum_{i=1}^\infty \psi_i J_i\|_{W^{1,p}(E,\mathbb{R})} \quad (33)\\ &= \|\sum_{i=1}^\infty \psi_i (V - J_i)\|_{W^{1,p}(E,\mathbb{R})} \le \sum_{i=1}^\infty \|\psi_i (V - J_i)\|_{W^{1,p}(E,\mathbb{R})}\\ &= \sum_{i=1}^\infty \|\psi_i (V - J_i)\|_{W^{1,p}(\bar{E}_i,\mathbb{R})} \le \sum_{i=1}^\infty \theta_i \|V - J_i\|_{W^{1,p}(\bar{E}_i,\mathbb{R})}\\ &< \sum_{i=1}^\infty \left(\frac{\varepsilon + \theta_i}{2^{i+1}(1+\tau_i+\theta_i)}\right) < \varepsilon.\end{aligned}$$

The second equality of Eqn. (33) follows since partitions of unity satisfy $\sum_{i=1}^\infty \psi_i(x,t) \equiv 1$ by Theorem 9. The first inequality follows by the triangle inequality. The third equality follows since partitions of unity satisfy $\psi_i(x,t) = 0$ outside of E_i for all $i \in \mathbb{N}$ by Theorem 9. The third inequality follows by Eqn. (31). The fourth inequality follows as $\sum_{i=1}^\infty \frac{1}{2^i} = 1$. Now, by a similar augment to Eqn. (33), using Eqn. (30) rather than Eqn. (31), it also follows $\sup_{(x,t)\in E} |V(x,t) - J(x,t)| < \varepsilon$ and thus J satisfies Eqn. (28).

Next we will show J satisfies Eqn. (29). Before doing this we first prove a preliminary identity. Specifically,

$$\sum_{i=1}^\infty \left(\nabla_t \psi_i(x,t) + \nabla_x \psi_i(x,t)^T f(x,u)\right) = 0, \quad (34)$$

for all $(x,t) \in \Omega \times (0,T) \subseteq E$ and $u \in U$. This follows because $\sum_{i=1}^\infty \psi_i(x,t) \equiv 1$ and therefore $\nabla_t \left(\sum_{i=1}^\infty \psi_i(x,t)\right) = \sum_{i=1}^\infty \nabla_t \psi_i(x,t) = 0$ and similarly for each $j \in \{1,...,n\}$ we have $\sum_{i=1}^\infty \frac{\partial \psi_i(x,t)}{\partial x_j} = 0$ which implies $\sum_{i=1}^\infty \nabla_x \psi_i(x,t) = 0 \in \mathbb{R}^n$.

Now, it follows J satisfies Eqn. (29) since

$$\begin{aligned}&\nabla_t J(x,t) + \nabla_x J(x,t)^T f(x,u) + c(x,u,t) \quad (35)\\ &= \sum_{i=1}^\infty \left(\psi_i(x,t)(\nabla_t J_i(x,t) + \nabla_x J_i(x,t)^T f(x,u) + c(x,u,t))\right)\\ &\quad + \sum_{i=1}^\infty \left(J_i(x,t)(\nabla_t \psi_i(x,t) + \nabla_x \psi_i(x,t)^T f(x,u))\right)\\ &\ge \frac{-\varepsilon}{2} + \sum_{i=1}^\infty (J_i(x,t) - V(x,t))(\nabla_t \psi_i(x,t) + \nabla_x \psi_i(x,t)^T f(x,u))\\ &\ge -\varepsilon,\end{aligned}$$

for all $(x,t) \in \Omega \times (0,T) \subseteq E$ and $u \in U$. The first equality of Eqn. (35) follows by the chain rule and the fact $\sum_{i=1}^\infty \psi_i(x,t) \equiv 1$. The first inequality follows by Eqns. (32) and (34). The second inequality follows by Eqn. (30) and $\sum_{i=1}^\infty \frac{1}{2^i} = 1$. ☐

B. Existence Of Dissipative Polynomials That Can Approximate The VF Arbitrarily well Under The $W^{1,p}$ Norm

Previously, in Prop. 3, we showed for any $V \in Lip(\Omega \times [0,T],\mathbb{R})$ satisfying Eqn. (23) there exists a smooth function J that also satisfies Eqn. (23) and approximates V with arbitrary accuracy under the Sobolev norm. We now use this result to show for any VF, associated with some family OCPs $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$, there exists a dissipative polynomial, V_l, that approximates the VF arbitrarily well with respect to the Sobolev norm. Our proof uses Theorem 6, found in Appendix XIII, that shows differentiable functions, such as J, can be approximated up to their first order derivatives over compact sets arbitrarily well by polynomials. Prop. 3 only gives the existence of a smooth approximation, J, when the VF is Lipschitz continuous. Lemma 1 shows the VF, associated with a family of OCPs, is locally Lipschitz when $\Omega = \mathbb{R}^n$ (which is not a compact set). Unfortunately, Theorem 6 can only be used for polynomial approximation over compact sets. Thus, before proceeding we first give a sufficient condition for a VF, associated with a family of OCPs with compact state constraints, to be Lipschitz continuous over some set $\Lambda \subset \Omega$.

a) Lipschitz continuity of VF's associated with a family of state constrained OCPs: Consider the family of OCPs $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$. If the state is constrained ($\Omega \neq \mathbb{R}^n$), the associated VF can be discontinuous and is no longer uniquely defined as the viscosity solution of the HJB PDE. Next, in Lemma 3, we give a sufficient condition that when satisfied implies VFs, associated with a family of state constrained OCPs, are equal to the unique locally Lipschitz continuous VF of the state unconstrained OCP over some subset $\Lambda \subseteq \Omega$. To state Lemma 3 we first define the forward reachable set.

Definition 10. *For $X_0 \subset \mathbb{R}^n$, $\Omega \subseteq \mathbb{R}^n$, $U \subset \mathbb{R}^m$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $S \subset \mathbb{R}^+$, define*

$$FR_f(X_0,\Omega,U,S) := \Big\{ y \in \mathbb{R}^n \ : \text{there exists}\, x \in X_0, T \in S, \text{and} \\ \mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x,0) \text{ such that } \phi_f(x,T,\mathbf{u}) = y \Big\}.$$

Lemma 3. *Consider $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and any function $V_1 : \Omega \times [0,T] \to \mathbb{R}$ that satisfies Eqn. (10). Let $V_2 : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ be the VF for the unconstrained problem $\{c,g,f,\mathbb{R}^n,U,T\}$. If $\Lambda \subseteq \Omega$ is such that*

$$FR_f(\Lambda,\mathbb{R}^n,U,[0,T]) \subseteq \Omega, \tag{36}$$

then $V_1(x,t) = V_2(x,t)$ for all $(x,t) \in \Lambda \times [0,T]$.

Proof. To show $V_1(x,t) = V_2(x,t)$ for all $(x,t) \in \Lambda \times [0,T]$ we must prove $\mathscr{U}_{\Omega,U,f,T}(x,t) = \mathscr{U}_{\mathbb{R}^n,U,f,T}(x,t)$ for all $(x,t) \in \Lambda \times [0,T]$.

For any $(x,t) \in \Lambda \times [0,T]$ if $\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x,t)$ then clearly $\mathbf{u} \in \mathscr{U}_{\mathbb{R}^n,U,f,T}(x,t)$, thus $\mathscr{U}_{\Omega,U,f,T}(x,t) \subseteq \mathscr{U}_{\mathbb{R}^n,U,f,T}(x,t)$. On the other hand if $\mathbf{u} \in \mathscr{U}_{\mathbb{R}^n,U,f,T}(x,t)$ then by Defn. 4 it follows $\mathbf{u}(s) \in U$ for all $s \in [t,T]$ and that there exists a unique map, denoted by $\phi_f(x,s,\mathbf{u})$, that satisfies the following for all $s \in [t,T]$

$$\frac{\partial \phi_f(x,s-t,\tau_t\mathbf{u})}{\partial s} = f(\phi_f(x,s-t,\tau_t\mathbf{u}),\mathbf{u}(s)),\ \phi_f(x,0,\tau_t\mathbf{u}) = x.$$

To show $\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x,t)$ we need $\phi_f(x,s-t,\tau_t\mathbf{u}) \in \Omega$ for all $s \in [t,T]$, which is equivalent to

$$\phi_f(x,s,\tilde{\mathbf{u}}) \in \Omega \text{ for all } s \in [0,T-t], \tag{37}$$

where $\tilde{\mathbf{u}} = \tau_t\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T-t}(x,0)$. Eqn. (37) then follows trivially by Eqn. (36). □

Alternative sufficient conditions that imply a VF, associated with some family of state constrained OCPs, is Lipschitz continuous and the unique viscosity solution of the HJB PDE include: the Inward Pointing Constraint Qualification (IPCQ) [25] [26], the Outward Pointing Constraint Qualification (OPCQ) [27], and epigraph characterization of VF's [28].

b) Approximation of VF's by dissipative polynomials: Considering a family of OCP's $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$, and assuming there exists a set $\Lambda \subseteq \Omega$ that satisfies Eqn. (36), we now prove the existence of dissipative polynomial functions that can approximate the any VF of $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ arbitrarily well under the Sobolev norm.

Theorem 3. *For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ suppose $\Lambda \subseteq \Omega$ is a bounded set that satisfies (36), then for any function V satisfying Eqn. (10), $1 \leq p < \infty$, and $\varepsilon > 0$ there exists $V_l \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ such that*

$$||V - V_l||_{W^{1,p}(\Lambda\times[0,T],\mathbb{R})} < \varepsilon, \tag{38}$$

$$\sup_{(x,t)\in\Lambda\times[0,T]} |V(x,t) - V_l(x,t)| < \varepsilon, \tag{39}$$

$$V_l(x,t) \leq V(x,t) \text{ for all } t \in [0,T] \text{ and } x \in \Omega, \tag{40}$$

$$\nabla_t V_l(x,t) + c(x,u,t) + \nabla_x V_l(x,t)^T f(x,u) > 0 \tag{41}$$
$$\text{for all } x \in \Omega, t \in (0,T), u \in U,$$

$$V_l(x,T) < g(x) \text{ for all } x \in \Omega. \tag{42}$$

Proof. Let $\varepsilon > 0$. Suppose V satisfies Eqn. (10). Rather than approximating V, defined for a family of OCPs on the compact set Ω, we instead approximate the unique VF, denoted by V^*, associated with the family of OCPs where $\Omega = \mathbb{R}^n$. It is easier to approximate V^* compared to V as V^* has the following useful properties: By Lemma 1, V^* is locally Lipschitz continuous; and by Theorem 1, V^* is the unique viscosity solution of the HJB PDE (12). Furthermore, as Λ satisfies Eqn. (36), Lemma 3 implies

$$V^*(x,t) = V(x,t) \text{ for all } (x,t) \in \Lambda \times [0,T]. \tag{43}$$

This proof is structured as follows. We first use Prop. 3 to approximate V^* by an infinitely differentiable function denoted as J_δ. Then using Theorem 6, found in Appendix XIII, we approximate J_δ by a polynomial P_δ. Finally, to ensure Inequalities (41) and (42) are satisfied, a correction term ρ is subtracted from P_δ, creating the function $V_l(x,t) := P_\delta(x,t) - \rho(t)$ that we show satisfies Eqns. (38) to (42).

Since Ω is compact, there exists some open bounded set $E \subset \mathbb{R}^{n+1}$ of finite measure which contains $\overline{\Omega \times (0,T)}$. Since $V^* \in LocLip(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ (by Lemma 1) and $E \subset \mathbb{R}^n$ is bounded it follows $V^* \in Lip(E \times [0,T],\mathbb{R})$. Then by Rademacher's theorem (See Theorem 7 in Section XIII), V^* is differentiable almost everywhere in E. Moreover, as V^* is the unique viscosity solution to the HJB PDE, the following holds for all $u \in U$ and almost everywhere in $(x,t) \in \Omega \times (0,T) \subset E$.

$$\nabla_t V^*(x,t) + c(x,u,t) + \nabla_x V^*(x,t)^T f(x,u) \\ \geq \nabla_t V^*(x,t) + \inf_{u\in U}\{c(x,u,t) + \nabla_x V^*(x,t)^T f(x,u)\} = 0$$

This implies that the following holds for all $u \in U$

$$\operatorname*{ess\,inf}_{(x,t)\in\Omega\times(0,T)} \Big\{ \nabla_t V^*(x,t) + \nabla_x V^*(x,t)^T f(x,u) + c(x,u,t) \Big\} \geq 0.$$

Therefore, we conclude that V^* satisfies Eqn. (23). Thus, by Prop. 3, for any $\delta > 0$ there exists $J_\delta \in C^\infty(E,\mathbb{R})$ such that

$$||V^* - J_\delta||_{W^{1,p}(E,\mathbb{R})} < \delta, \tag{44}$$

$$\nabla_t J_\delta(x,t) + \nabla_x J_\delta(x,t)^T f(x,u) + c(x,u,t) \geq -\delta \tag{45}$$
$$\text{for all} (x,t) \in \Omega \times (0,T).$$

In particular, let us choose $\delta > 0$ such that

$$\delta < \frac{\varepsilon}{2 + (2 + 4T + 2MT)(T\mu(\Lambda))^{\frac{1}{p}}}, \tag{46}$$

where $M := \sup_{(x,u)\in\Omega\times U} ||f(x,u)||_2 < \infty$ and $\mu(\Lambda) < \infty$ is the Lebesgue measure of Λ.

We now approximate $J_\delta \in C^\infty(E,\mathbb{R})$ by a polynomial function. Theorem 6, found in Section XIII, shows there exists $P_\delta \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R},\mathbb{R})$ such that for all $(x,t) \in E$

$$|J_\delta(x,t) - P_\delta(x,t)| < \delta. \tag{47}$$
$$|\nabla_t J_\delta(x,t) - \nabla_t P_\delta(x,t)| < \delta. \tag{48}$$
$$||\nabla_x J_\delta(x,t) - \nabla_x P_\delta(x,t)||_2 < \delta. \tag{49}$$
$$||J_\delta - P_\delta||_{W^{1,p}(E,\mathbb{R})} < \delta. \tag{50}$$

Now, $||V^* - P_\delta||_{W^{1,p}(E,\mathbb{R})} = ||V^* - J_\delta + J_\delta - P_\delta||_{W^{1,p}(E,\mathbb{R})}$

$$\leq ||V^* - J_\delta||_{W^{1,p}(E,\mathbb{R})} + ||J_\delta - P_\delta||_{W^{1,p}(E,\mathbb{R})} < 2\delta, \tag{51}$$

where the first inequality follows by the triangle inequality, and the second inequality follows from Eqn. (44) and Eqn. (50).

By a similar argument to Inequality (51) we deduce,

$$\sup_{(x,t)\in E} |V^*(x,t) - P_\delta(x,t)| < 2\delta. \tag{52}$$

Furthermore,

$$\begin{aligned}
&\nabla_t P_\delta(x,t) + \nabla_x P_\delta(x,t)^T f(x,u) + c(x,u,t)\\
&\geq \left(\nabla_t P_\delta(x,t) + \nabla_x P_\delta(x,t)^T f(x,u) + c(x,u,t)\right)\\
&\qquad - \delta - \left(\nabla_t J_\delta(x,t) + \nabla_x J_\delta(x,t)^T f(x,u) + c(x,u,t)\right)\\
&= -\delta + \left(\nabla_t P_\delta(x,t) - \nabla_t J_\delta(x,t)\right)\\
&\qquad - \left(\nabla_x J_\delta(x,t) - \nabla_x P_\delta(x,t)\right)^T f(x,u)\\
&> -\delta - \delta - ||\nabla_x J_\delta(x,t) - \nabla_x P_\delta(x,t)||_2 ||f(x,u)||_2\\
&> -(2+M)\delta \text{ for all } (x,t) \in \Omega \times (0,T),
\end{aligned} \tag{53}$$

The first inequality of Eqn. (53) follows by Inequality Eqn. (45). The second inequality follows by Eqn. (48) and the Cauchy Schwarz inequality. The third inequality follows by Eqn. (49).

Moreover, we have that

$$\begin{aligned}
P_\delta(x,T) &= P_\delta(x,T) - V^*(x,T) + V^*(x,T)\\
&< g(x) + 2\delta \text{ for all } x \in \Omega.
\end{aligned} \tag{54}$$

This inequality follows from the fact that $V^*(x,T) = g(x)$ since V^* satisfies the boundary condition in the HJB PDE (12), and Eqn. (52).

We now construct V_l from P_δ. Let us denote the correction function $\rho(t) := (2+M)(T-t)\delta + 2\delta$, where $M = \sup_{(x,u)\in\Omega\times U} ||f(x,u)||_2$. We define V_l as

$$V_l(x,t) := P_\delta(x,t) - \rho(t). \tag{55}$$

We now find that V_l satisfies Inequality (41) since we have

$$\begin{aligned}
&\nabla_t V_l(x,t) + c(x,u,t) + \nabla_x V_l(x,t)^T f(x,u)\\
&= \left(\nabla_t P_\delta(x,t) + \nabla_x P_\delta(x,t)^T f(x,u) + c(x,u,t)\right) + (2+M)\delta\\
&> 0, \qquad \text{for all } (x,t) \in \Omega \times (0,T),
\end{aligned}$$

where the above inequality follows from Eqn. (53).

We next show V_l satisfies Inequality (42):

$$V_l(x,T) = P_\delta(x,T) - 2\delta < g(x) \text{ for all } x \in \Omega,$$

where the above inequality follows by Eqn. (54).

Now, since V_l satisfies Eqns. (41) and (42) it follows V_l satisfies Eqn. (40) by Prop. 1.

To show that V_l satisfies Inequality (38), we first we derive a bound on the norm of the correction function ρ.

$$\begin{aligned}
||\rho||_{W^{1,p}(\Lambda\times[0,T],\mathbb{R})} &= \left(\int_{\Lambda\times[0,T]} |(2+M)(T-t)\delta + 2\delta|^p dxdt\right)^{\frac{1}{p}}\\
&+ \left(\int_{\Lambda\times[0,T]} |(2+M)\delta|^p dxdt\right)^{\frac{1}{p}} \leq (2+4T+2MT)(T\mu(\Lambda))^{\frac{1}{p}}\delta.
\end{aligned}$$

Now, by Eqns. (43), (46) and (51),

$$\begin{aligned}
&||V - V_l||_{W^{1,p}(\Lambda\times[0,T],\mathbb{R})} = ||V^* - V_l||_{W^{1,p}(\Lambda\times[0,T],\mathbb{R})}\\
&= ||V^* - P_\delta - \eta||_{W^{1,p}(\Lambda\times[0,T],\mathbb{R})}\\
&\leq ||V^* - P_\delta||_{W^{1,p}(E,\mathbb{R})} + ||\eta||_{W^{1,p}(\Lambda\times[0,T],\mathbb{R})}\\
&\leq 2\delta + (2+4T+2MT)(T\mu(\Lambda))^{\frac{1}{p}}\delta < \varepsilon.
\end{aligned} \tag{56}$$

By a similar argument to Eqn. (56) we deduce V_l satisfies Eqn. (39)

We conclude that V_l, defined in Eqn. (55), satisfies Eqns. (39), (40), (41), and (42) thus completing the proof. □

C. Our Family Of SOS Problems Yield A Sequence Of Polynomials That Converge To A VF With Respect To The L^1 Norm

Consider some $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and suppose the sequence $\{J_d\}_{d\in\mathbb{N}}$ solves each instance of the optimization problem given in Eqn. (20) for $d \in \mathbb{N}$. We next use Theorem 3 to show that the sequence, $\{J_d\}_{d\in\mathbb{N}}$ converges to any VF associated with the family of OCP's $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ with respect to the weighted L^1 norm as $d \to \infty$.

Proposition 4. *For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and positive integrable function $w \in L^1(\Omega\times[0,T],\mathbb{R}^+)$ suppose $\Lambda \subseteq \Omega$ satisfies Eqn.* (36) *then*

$$\lim_{d\to\infty} \int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - J_d(x,t)|dxdt = 0, \tag{57}$$

where V is any function satisfying Eqn. (10)*, and $J_d \in \mathscr{P}_d(\mathbb{R}^n\times\mathbb{R},\mathbb{R})$ is any solution to Optimization Problem* (20) *for $d \in \mathbb{N}$.*

Proof. Suppose V satisfies the theorem statement. To show Eqn. (57) we must show that for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - J_d(x,t)|dxdt < \varepsilon \text{ for all } d \geq N.$$

Since by assumption Λ satisfies Eqn. (36), we can use Theorem 3 (from Section VI-B) to show that for any $\delta > 0$ there exists dissipative $V_l \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ feasible to Optimization Problem (20) and is such that

$$\operatorname*{ess\,sup}_{(x,t)\in\Lambda\times[0,T]} |V(x,t) - V_l(x,t)| < \delta.$$

For our given $\varepsilon > 0$, by selecting $\delta < \varepsilon / \int_{\Lambda\times[0,T]} w(x,t)dxdt$ (Note if $\int_{\Lambda\times[0,T]} w(x,t)dxdt = 0$, Eqn. (57) already holds and the proof is complete) we have a V_l such that

$$\begin{aligned} &\int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - V_l(x,t)|dxdt \qquad (58)\\ &\leq \int_{\Lambda\times[0,T]} w(x,t)dxdt \operatorname*{ess\,sup}_{(x,t)\in\Lambda\times[0,T]} |V(x,t) - V_l(x,t)| \\ &< \delta \int_{\Lambda\times[0,T]} w(x,t)dxdt < \varepsilon. \end{aligned}$$

Now define $N := \deg(V_l)$ and denote the solution to Problem (20) for $d \geq N$ as $J_d \in \mathscr{P}_N(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$. As V_l is feasible to Problem (20) for all $d \geq N$, it follows the objective function evaluated at J_d is greater than or equal to the objective function evaluated at V_l; that is

$$\int_{\Lambda\times[0,T]} w(x,t)J_d(x,t)dxdt \geq \int_{\Lambda\times[0,T]} w(x,t)V_l(x,t)dxdt \text{ for } d \geq N. \quad (59)$$

$$\begin{aligned} \text{Now, } &\int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - J_d(x,t)|dxdt \qquad (60)\\ &= \int_{\Lambda\times[0,T]} w(x,t)V(x,t) - w(x,t)J_d(x,t)dxdt \\ &\leq \int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - V_l(x,t)|dxdt < \varepsilon \text{ for all } d \geq N. \end{aligned}$$

The equality in Eqn. (60) follows since $J_d(x,t) \leq V(x,t)$ for all $(x,t) \in \Omega \times [0,T]$ (Prop. 1). The first inequality follows by a combination of Eqn. (59) and the inequality $V_l(x,t) \leq V(x,t)$ for all $(x,t) \in \Omega \times [0,T]$. Finally, the second inequality follows by Eqn. (58). □

VII. A Family Of SOS Problems That Yield Polynomials That Converge To The VF

Consider some $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and denote $\{J_d\}_{d\in\mathbb{N}}$ as the sequence of solutions to the optimization problem found in Eqn. (20). We have shown in Prop. 5 that the sequence of functions, $\{J_d\}_{d\in\mathbb{N}}$, converge to any VF associated with the family of OCP's $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ with respect to the L^1 norm. The indexed polynomial optimization problems in Eqn. (20) may now be readily tightened to more tractable SOS optimization problems.

Specifically, for each $d \in \mathbb{N}$, we tighten the polynomial optimization problem in Eqn. (20) to the SOS optimization problem given in Eqn. (61). We will show in Prop. 5 that the sequence of solutions to the SOS problem given in Eqn. (61) yield polynomials, $\{P_d\}_{d\in\mathbb{N}}$, indexed by degree $d \in \mathbb{N}$, that converge to the VF (with respect to the L^1 norm) as $d \to \infty$.

For our SOS implementation we consider a special class of OCPs, given next in Defn. 11.

Definition 11. *We say the six tuple $\{c,g,f,\Omega,U,T\}$ is a polynomial optimal control problem or $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Poly}$ if the following holds*

1) $c \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}, \mathbb{R})$ *and* $g \in \mathscr{P}(\mathbb{R}^n, \mathbb{R})$.
2) $f \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^n)$.
3) *There exists* $h_\Omega \in \mathscr{P}(\mathbb{R}^n, \mathbb{R})$ *such that* $\Omega = \{x \in \mathbb{R}^n : h_\Omega(x) \geq 0\}$.
4) *There exists* $h_U \in \mathscr{P}(\mathbb{R}^m, \mathbb{R})$ *such that* $U = \{u \in \mathbb{R}^m : h_U(u) \geq 0\}$.

Note polynomials are locally Lipschitz continuous, that is $\mathscr{P}(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}) \subset LocLip(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$. Therefore $\mathscr{M}_{Poly} \subset \mathscr{M}_{Lip}$.

For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Poly}$, $d \in \mathbb{N}$, $\Lambda \subset \mathbb{R}^n$ and $w \in L^1(\Lambda \times [0,T], \mathbb{R}^+)$, we thus propose an SOS tightening of Optimization Problem (20) as follows:

$$\begin{aligned} &P_d \in \arg \max_{P\in\mathscr{P}_d(\mathbb{R}^n\times\mathbb{R},\mathbb{R})} c^T\alpha \qquad (61)\\ &\text{subject to: } k_0, k_1 \in \sum_{SOS}^{d}, \quad s_i \in \sum_{SOS}^{d} \text{ for } i = 0,1,2,3,\\ &P(x,t) = c^T Z_d(x,t),\\ &k_0(x) = g(x) - V_l(x,T) - s_0(x)h_\Omega(x),\\ &k_1(x,u,t) = \nabla_t P(x,t) + c(x,u,t) + \nabla_x P(x,t)^T f(x,u)\\ &\quad - s_1(x,u,t)h_\Omega(x) - s_2(x,u,t)h_U(u) - s_3(x,u,t)\cdot(Tt - t^2), \end{aligned}$$

where $\alpha_i = \int_{\Lambda\times[0,T]} w(x,t)Z_{d,i}(x,t)dxdt$, and recalling $Z_d : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{\mathscr{N}_d}$ is the vector of monomials of degree $d \in \mathbb{N}$.

Proposition 5. *For given $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Poly}$ and positive integrable function $w \in L^1(\Omega \times [0,T], \mathbb{R}^+)$ suppose $\Lambda \subseteq \Omega$ satisfies Eqn.* (36) *then*

$$\lim_{d\to\infty} \int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - P_d(x,t)|dxdt = 0, \qquad (62)$$

where V is any function satisfying Eqn. (10) *and $P_d \in \mathscr{P}_d(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ is any solution to Problem* (61) *for $d \in \mathbb{N}$.*

Proof. To show Eqn. (62) we show that for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $d \geq N$

$$\int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - P_d(x,t)|dxdt < \varepsilon.$$

As it is assumed Λ satisfies Eqn. (36) we are able to use Prop. 4 that shows for any $\varepsilon > 0$ there exists $N_1 \in \mathbb{N}$ such that for all $d \geq N_1$

$$\int_{\Lambda\times[0,T]} w(x,t)|V(x,t) - J_d(x,t)|dxdt < \varepsilon, \qquad (63)$$

where J_d is a solution to Optimization Problem (20) for $d \in \mathbb{N}$.

In particular let us fix some $d_1 \geq N_1$. Since J_{d_1} solves Problem (20) it must satisfy the constraints of Problem (20). Thus we have

$$\begin{aligned} &k_0(x) := g(x) - J_{d_1}(x,T) > 0 \text{ for all } x \in \Omega,\\ &k_1(x,u,t) := \nabla_t J_{d_1}(x,t) + c(x,u,t) + \nabla_x J_{d_1}(x,t)^T f(x,u) > 0\\ &\qquad\qquad \text{for all } (x,u,t) \in \Omega \times U \times [0,T]. \end{aligned}$$

Since k_0 and k_1 are strictly positive functions over the compact semialgebriac set $\Omega \times U \times [0,T] = \{(x,u,t) \in \mathbb{R}^{n+m+1} : h_\Omega(x) \geq 0, h_U(u) \geq 0, t(T-t) \geq 0\}$, Putinar's Positivstellesatz (stated in Theorem 8, Appendix XIII) shows that there exist $s_0, s_1, s_2, s_3, s_4, s_5 \in \sum_{SOS}$ such that

$$k_0 - h_\Omega s_0 = s_1, \tag{64}$$
$$k_1 - h_\Omega s_2 - h_U s_3 - h_T s_4 = s_5,$$

where $h_T(t) := (t)(T-t)$.

Let us denote $N_2 := \max_{i \in \{0,1,2,3,4,5\}} deg(s_i)$. By Eqn. (64) it follows that if J_{d_1} is feasible to Problem (61) for $d \geq \max\{d_1, N_2\}$. Therefore, for $d \geq \max\{d_1, N_2\}$, the objective function evaluated at the solution to Problem (61) must be greater than or equal to objective function evaluated at J_{d_1}. That is by writing the solution to Problem (61) as $P_d(x,t) = c_d^T Z_d(x,t)$ and writing J_{d_1} as $J_{d_1} = b_{d_1}^T Z_{d_1}(x,t)$ we get that for $d \geq \max\{d_1, N_2\}$

$$c_d^T \alpha \geq b_{d_1}^T \alpha. \tag{65}$$

Now using Eqn's. (65) and (63) it follows for all $d \geq \max\{d_1, N_2\}$

$$\begin{aligned}
&\int_{\Lambda \times [0,T]} w(x,t)|V(x,t) - P_d(x,t)|dxdt \\
&= \int_{\Lambda \times [0,T]} w(x,t)V(x,t)dxdt - \int_{\Lambda \times [0,T]} w(x,t)P_d(x,t)dxdt \\
&= \int_{\Lambda \times [0,T]} w(x,t)V(x,t)dxdt - c_d^T \alpha \\
&\leq \int_{\Lambda \times [0,T]} w(x,t)V(x,t)dxdt - b_{d_1}^T \alpha \\
&= \int_{\Lambda \times [0,T]} w(x,t)|V(x,t) - J_{d_1}(x,t)|dxdt < \varepsilon,
\end{aligned}$$

where the above inequality follows using Prop. 1, which shows $P_d(x,t) \leq V(x,t)$ and $J_d(x,t) \leq V(x,t)$ for all $(x,t) \in \Omega \times [0,T]$, as P_d and J_d satisfy Inequalities (16) and (17) (since they both satisfy the constraints of Optimization Problem (20)). □

A. We Can Numerically Construct A Sequence Of Sublevel Sets That Converge To The VF's Sublevel Set

For a given family of OCPs, Prop. 5 shows the SOS optimization problem, given in Eqn. (61), yields a sequence of polynomials, $\{P_d\}_{d \in \mathbb{N}}$, a sequence that converges to the VF (denoted by V), where convergence is with respect to the L^1 norm, and where the VF is associated with the given family of OCPs. We next extend this convergence result by showing that, for any $\gamma \in \mathbb{R}$, the sequence $\{P_d\}_{d \in \mathbb{N}}$ yields a sequence of γ-sublevel sets, where the sequence of γ-sublevel sets converges to the γ-sublevel set of the value function, V, where convergence is with respect to the volume metric.

For sets $A, B \subset \mathbb{R}^n$, we denote the volume metric as $D_V(A,B)$, where

$$D_V(A,B) := \mu((A/B) \cup (B/A)), \tag{66}$$

where we recall $\mu(A) := \int_{\mathbb{R}^n} \mathbb{1}_A(x)dx$ is the Lebesgue measure. Note that Lemma 4 (Appendix XI) shows that D_V is a metric.

Proposition 6. *Consider $\{c, g, f, \Omega, U, T\} \in \mathscr{M}_{Poly}$ and $w(x,t) = \delta(t-s)$ where $s \in [0,T]$ and δ is the Dirac delta function. Suppose $\Lambda \subseteq \Omega$ satisfies Eqn.* (36)*. Then we have the following for all $\gamma \in \mathbb{R}$:*

$$\lim_{d \to \infty} D_V\left(\{x \in \Lambda : V(x,s) \leq \gamma\}, \{x \in \Lambda : P_d(x,s) \leq \gamma\}\right) = 0, \tag{67}$$

where V is any function satisfying Eqn. (10)*, and P_d is any solution to Problem* (61) *for $d \in \mathbb{N}$.*

Proof. To show Eqn. (67) we use Prop. 7, found in Appendix XI. Let us consider the family of functions, $\{P_d \in \mathscr{P}_d(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}) : d \in \mathbb{N}\}$, where P_d solves the optimization problem given in Eqn. (61) for $d \in \mathbb{N}$ and $w(x,t) = \delta(t-s)$.

From the definition of Problem (61), we have that P_d satisfies the Inequalities in (16) and (17). Therefore, by Prop. 1, we have that $P_d(x,t) \leq V(x,t)$ for all $(x,t) \in \Omega \times [0,T]$, where V is any function satisfying Eqn. (10). Since $\Lambda \subseteq \Omega$ satisfies Eqn. (36), and although the Dirac Delta function is not a member of $L^1(\Omega \times [0,T], \mathbb{R})$, a similar argument to Prop. 5 implies that

$$\begin{aligned}
&\lim_{d \to \infty} \int_\Lambda |V(x,s) - P_d(x,s)|dxdt \\
&\qquad = \lim_{d \to \infty} \int_{\Lambda \times [0,T]} \delta(t-s)|V(x,t) - P_d(x,t)|dxdt = 0.
\end{aligned}$$

We now apply Prop. 7 (Section XI) to deduce Eqn. (67). □

VIII. A Performance Bound On Controllers Constructed Using Approximation to the VF

Given an OCP, if an associated differentiable VF is known then a solution to the OCP can be constructed using Theorem 2. However, in general, it is challenging to find a VF analytically. Rather than computing a true VF, we consider a candidate VF which is "close" to a true VF under some norm. This motivates us to ask the question: how well will a controller constructed from a candidate VF perform? To answer this question we next define the loss/performance of an input. For $\{c, g, f, \Omega, U, T\} \in \mathscr{M}_{Lip}$ we denote the loss/performance function as,

$$\begin{aligned}
L(x_0, \mathbf{u}) := &\int_0^T c(\phi_f(x_0, s, \mathbf{u}), \mathbf{u}(s), s)ds + g(\phi_f(x_0, T, \mathbf{u})) \qquad (68)\\
&- \inf_{\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x_0,0)} \left\{\int_0^T c(\phi_f(x_0, s, \mathbf{u}), \mathbf{u}(s), s)ds + g(\phi_f(x_0, T, \mathbf{u}))\right\}.
\end{aligned}$$

Clearly, $L(x_0, \mathbf{u}) \geq 0$ for all $(x_0, \mathbf{u}) \in \Omega \times \mathscr{U}_{\Omega,U,f,T}(x_0, 0)$.

Theorem 4. *Consider $\{c, g, f, \mathbb{R}^n, U, T\} \in \mathscr{M}_{Lip}$. Suppose $J \in C^2(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ and $\Omega \subset \mathbb{R}^n$ is an open set such that for $x_0 \in \mathbb{R}^n$ we have $FR_f(x_0, \mathbb{R}^n, U, [0,T]) \subseteq \Omega$. In this case we have*

$$L(x_0, \mathbf{u}_J) \leq C||J - V^*||_{W^{1,\infty}(\Omega \times [0,T], \mathbb{R})}, \tag{69}$$

$$\text{where } C := 2\max\left\{1, T, T \max_{1 \leq i \leq n} \sup_{(x,t) \in \Omega \times U} |f_i(x,u)|\right\}, \tag{70}$$

V^* is the unique viscosity solution to the HJB PDE (12)*,*

$$\mathbf{u}_J(t) = k_J(\phi_f(x_0, t, \mathbf{u}_J), t), \tag{71}$$

and k_J is any function such that

$$k_J(x,t) \in \arg\inf_{u \in U}\{c(x,u,t) + \nabla_x J(x,t)^T f(x,u)\}. \tag{72}$$

Proof. Now, for any $J \in C^2(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}) \subset LocLip(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$, we wish to show that Eqn. (69) holds. To do this, we will show that J is the true VF for some modified OCP. Before constructing this modified OCP, for any $F \in LocLip(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$, let us define

$$H_F(x,t,u) := \nabla_t F(x,t) + c(x,u,t) + \nabla_x F(x,t)^T f(x,u),$$
$$\tilde{H}_F(x,t) := \inf_{u \in U} H_F(x,t,u),$$

where $\nabla_t F$ and $\nabla_x F$ are weak derivatives, known to exist by Rademacher's Theorem (Thm. 7).

Then, by construction, J satisfies the following PDE

$$\nabla_t J(x,t) + \inf_{u\in U}\left\{c(x,u,t) - \tilde{H}_J(x,t) + \nabla_x J(x,t)^T f(x,u)\right\} = 0$$
$$\text{for all } (x,t) \in \mathbb{R}^n \times [0,T]. \quad (73)$$

Eqn. (73) implies that J satisfies the HJB PDE associated with $\{\tilde{c}, \tilde{g}, f, \mathbb{R}^n, U, T\}$, where $\tilde{c}(x,u,t) := c(x,u,t) - \tilde{H}_J(x,t)$ and $\tilde{g}(x) := J(x,T)$. Note that since $c \in LocLip(\Omega \times U \times [0,T], \mathbb{R})$, $f \in LocLip(\Omega \times U, \mathbb{R})$, and $\frac{\partial}{\partial x_i} J \in LocLip(\Omega \times [0,T], \mathbb{R})$ for all $i \in \{1, ... n+1\}$ (since $J \in C^2(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$) it follows that $H_J \in LocLip(\Omega \times U \times [0,T], \mathbb{R})$. By Lemma 7 we then deduce $\tilde{H}_J \in LocLip(\Omega \times [0,T], \mathbb{R})$ and thus $\{\tilde{c}, \tilde{g}, f, \mathbb{R}^n, U, T\} \in \mathscr{M}_{Lip}$.

Since H_J is independent of $u \in U$, we have that

$$\arg\inf_{u\in U}\{\tilde{c}(x,u,t) + \nabla_x J(x,t)^T f(x,u)\}$$
$$= \arg\inf_{u\in U}\{c(x,u,t) + \nabla_x J(x,t)^T f(x,u)\},$$

and therefore we are able to deduce by Theorem 2 that $\mathbf{u}_J$ (given in Eqn. (71)) solves the modified OCP associated with $\{\tilde{c}, \tilde{g}, f, \mathbb{R}^n, U, T\}$ with initial condition $x_0 \in \mathbb{R}^n$. Thus for all $\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x_0, 0)$ we have that

$$\int_0^T \tilde{c}(\phi_f(x_0,s,\mathbf{u}_J), \mathbf{u}_J(s), s)ds + \tilde{g}(\phi_f(x_0,T,\mathbf{u}_J)) \quad (74)$$
$$\leq \int_0^T \tilde{c}(\phi_f(x_0,s,\mathbf{u}), \mathbf{u}(s), s)ds + \tilde{g}(\phi_f(x_0,T,\mathbf{u})).$$

By substituting $\tilde{c}(x,u,t) = c(x,u,t) - \tilde{H}_J(x,t)$ and $\tilde{g}(x) = J(x,T)$ into Inequality (74) and noting that $V^*(x,T) = g(x)$, we have the following for all $\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x_0, 0)$:

$$\int_0^T c(\phi_f(x_0,s,\mathbf{u}_J), \mathbf{u}_J(s), s)ds + g(\phi_f(x_0,T,\mathbf{u}_J)) \quad (75)$$
$$- \int_0^T c(\phi_f(x_0,s,\mathbf{u}), \mathbf{u}(s), s)ds - g(\phi_f(x_0,T,\mathbf{u}))$$
$$\leq \int_0^T \tilde{H}_J(\phi_f(x_0,s,\mathbf{u}_J), s) - \tilde{H}_J(\phi_f(x_0,s,\mathbf{u}), s)ds$$
$$+ V^*(\phi_f(x_0,T,\mathbf{u}_J), T) - J(\phi_f(x_0,T,\mathbf{u}_J), T)$$
$$+ J(\phi_f(x_0,T,\mathbf{u}), T) - V^*(\phi_f(x_0,T,\mathbf{u}), T)$$
$$< T \operatorname*{ess\,sup}_{s\in[0,T]}\{\tilde{H}_J(\phi_f(x_0,s,\mathbf{u}_J), s) - \tilde{H}_J(\phi_f(x_0,s,\mathbf{u}), s)\}$$
$$+ 2\sup_{y\in\Omega}\{|V^*(y,T) - J(y,T)|\}$$
$$\leq T\left(\operatorname*{ess\,sup}_{(y,s)\in\Omega\times[0,T]}\{\tilde{H}_J(y,s)\} - \operatorname*{ess\,inf}_{(y,s)\in\Omega\times[0,T]}\{\tilde{H}_J(y,s)\}\right)$$
$$+ 2 \operatorname*{ess\,sup}_{(y,s)\in\Omega\times[0,T]}\{|V^*(y,s) - J(y,s)|\}.$$

The second and third inequalities of Eqn. (75) follow because $\phi_f(x_0,t,\mathbf{u}) \in \Omega$ for all $(t,\mathbf{u}) \in [0,T] \times \in \mathscr{U}_{\Omega,U,f,T}(x_0,0)$ (since it is assumed $FR_f(x_0, \mathbb{R}^n, U, [0,T]) \subseteq \Omega$), and because $\sup_{y\in\Omega}\{|V^*(y,T) - J(y,T)|\} = \operatorname{ess\,sup}_{y\in\Omega}\{|V^*(y,T) - J(y,T)|\}$ holds by Lemma 9 (since V^* and J are both continuous, and Ω is open).

We now split the remainder of the proof into three parts. In Part 1, we derive an upper bound for $\operatorname{ess\,sup}_{(y,s)\in\Omega\times[0,T]}\{\tilde{H}_J(y,s)\}$. In Part 2, we find a lower bound for $\operatorname{ess\,inf}_{(y,s)\in\Omega\times[0,T]}\{\tilde{H}_J(y,s)\}$. In Part 3 we use these two bounds, combined with Inequality (75) to verify Eqn. (69) and complete the proof.

Before proceeding with Parts 1 to 3 we introduce some notation for the set of points where the VF is differentiable,

$$S_{V^*} := \{(x,t) \in \Omega \times [0,T] : V^* \text{ is differentiable at } (x,t)\}.$$

Lemma 1 shows that $V^* \in Lip(\Omega \times [0,T], \mathbb{R}) \subset LocLip(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ and Rademacher's Theorem (Thm. 7) states that Lipschitz functions are differentiable almost everywhere. It follows, therefore, that $\mu((\Omega \times [0,T])/S_{V^*}) = 0$, where μ is the Lebesgue measure.

Part 1 of Proof: For each $(y,s) \in S_{V^*}$ let us consider some family of points $k^*_{y,s} \in U$ such that

$$k^*_{y,s} \in \arg\inf_{u\in U}\left\{c(y,u,s) + \nabla_x V^*(y,s)^T f(y,u)\right\}.$$

Note, $k^*_{y,s}$ exists for each fixed $(y,s) \in S_{V^*}$ by the extreme value theorem since $U \subset \mathbb{R}^m$ is compact, c, f are continuous, and $\nabla_x V^*$ is independent of $u \in U$ and bounded by Rademacher's Theorem (Thm. 7).

Now for all $(y,s) \in S_{V^*}$ it follows that

$$\tilde{H}_J(y,s) = \inf_{u\in U} H_J(y,s,u) \leq H_J(y,s,k^*_{y,s}). \quad (76)$$

Moreover, since V^* is the viscosity solution to the HJB PDE by Theorem 1, we have that

$$H_{V^*}(y, k^*_{y,s}, s) = 0 \quad \text{for all } (y,s) \in S_{V^*}. \quad (77)$$

Combing Eqns. (76) and (77) it follows that

$$\tilde{H}_J(y,s) \leq H_J(y,s,k^*_{y,s}) - H_{V^*}(y, k^*_{y,s}, s)$$
$$= \nabla_t J(y,s) - \nabla_t V^*(y,s) + (\nabla_x J(y,s) - \nabla_x V^*(y,s))^T f(y, k^*_{y,s})$$
$$\leq |\nabla_t J(y,s) - \nabla_t V^*(y,s)|$$
$$+ \max_{1\leq i\leq n}|f_i(y,k^*_{y,s})| \sum_{i=1}^n \left|\frac{\partial}{\partial x_i}(J(y,s) - V^*(y,s))\right| \quad (78)$$

for all $(y,s) \in S_{V^*}$. As Eqn. (78) is satisfied for all $(y,s) \in S_{V^*}$ and $\mu((\Omega \times [0,T])/S_{V^*}) = 0$ it follows Eqn. (78) holds almost everywhere. Therefore

$$\operatorname*{ess\,sup}_{(y,s)\in\Omega\times[0,T]} \tilde{H}_J(y,s) \quad (79)$$
$$\leq \max\left\{1, \max_{1\leq i\leq n} \sup_{(x,t)\in\Omega\times U} |f_i(x,u)|\right\} ||V^* - J||_{W^{1,\infty}(\Omega\times[0,T])}.$$

Part 2 of Proof: If k_J satisfies Eqn. (72), then

$$\tilde{H}_J(y,s) = \inf_{u\in U} H_J(y,s,u) = H_J(y,s,k_J(y,s)) \text{ for all } (y,s) \in S_{V^*}. \quad (80)$$

Moreover, since V^* is a viscosity solution to the HJB PDE (12), we have by Theorem 1 that

$$H_{V^*}(y,s,k_J(y,s)) \geq \inf_{u\in U} H_{V^*}(y,s,u) = 0 \text{ for all } (y,s) \in S_{V^*}. \quad (81)$$

Combining Eqn's (80) and (81) it follows that

$$\begin{aligned}&\tilde{H}_J(y,s) \geq H_J(y,s,k_J(y,s)) - H_{V^*}(y,s,k_J(y,s))\\ &= \nabla_t J(y,s) - \nabla_t V^*(y,s) + (\nabla_x J(y,s) - \nabla_x V^*(y,s))^T f(y,k_J(y,s))\\ &\geq -|\nabla_t J(y,s) - \nabla_t V^*(y,s)|\\ &\quad - \max_{1\leq i\leq n} |f_i(y,k_J(y,s))| \sum_{i=1}^n \left| \frac{\partial}{\partial x_i}(J(y,s) - V^*(y,s)) \right| \qquad (82)\end{aligned}$$

for all $(y,s) \in S_{V^*}$. Therefore, since $\mu((\Omega \times [0,T])/S_{V^*}) = 0$, we have that Inequality (82) holds almost everywhere. Thus

$$\begin{aligned}&\operatorname*{ess\,inf}_{(y,s)\in\Omega\times[0,T]} \tilde{H}_J(y,s) \qquad (83)\\ &\geq -\max\left\{1, \max_{1\leq i\leq n} \sup_{(x,t)\in\Omega\times U} |f_i(x,u)|\right\} ||V^* - J||_{W^{1,\infty}(\Omega\times[0,T])}.\end{aligned}$$

Part 3 of Proof:

Combining Inequalities (75), (79) and (83) it follows that

$$\begin{aligned}&\int_0^T c(\phi_f(x_0,s,\mathbf{u}_J),\mathbf{u}_J(s),s)ds + g(\phi_f(x_0,T,\mathbf{u}_J)) \qquad (84)\\ &\quad - \int_0^T c(\phi_f(x_0,s,\mathbf{u}),\mathbf{u}(s),s)ds - g(\phi_f(x_0,T,\mathbf{u}))\\ &< C||J - V^*||_{W^{1,\infty}} \text{ for all } \mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x_0,0),\end{aligned}$$

where $C := 2\max\left\{1, T, T\max_{1\leq i\leq n}\sup_{(x,t)\in\Omega\times U}|f_i(x,u)|\right\}$. Now as Inequality (84) holds for all $\mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(x_0,0)$ we can take the infimum and deduce Inequality (69). □

IX. Numerical Examples: Using Our SOS Algorithm To Approximate VF's

In this section we use the SOS programming problem as defined in Eqn. (61) to numerically approximate the VFs associated with several different OCPs. We first approximate a known VF. Then, in Subsection IX-A, we approximate an unknown VF and use this approximation to construct a controller and analyze its performance. Then, in Subsection IX-B, we approximate another unknown VF for reachable set estimation.

Example 1. *Let us consider the tuple* $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Poly}$, *where* $c(x,u,t) \equiv 0$, $g(x) = x$, $f(x,u) = xu$, $\Omega = (-R,R) = \{x \in \mathbb{R} : x^2 < R^2\}$, $U = (-1,1) = \{u \in \mathbb{R} : u^2 < 1\}$, *and* $T = 1$. *It was shown in [1] that the VF associated with* $\{c,g,f,\mathbb{R}^n,U,T\}$ *can be analytically found as*

$$V^*(x,t) = \begin{cases} \exp(t-1)x \text{ if } x > 0,\\ \exp(1-t)x \text{ if } x < 0,\\ 0 \text{ if } x = 0.\end{cases} \qquad (85)$$

We note that V^* *is not differentiable at* $x = 0$. *However,* V^* *satisfies the HJB PDE away from* $x = 0$. *This problem shows that the VF can be non-smooth even for simple OCP's with polynomial vector field and cost functions.*

In Fig. 1 we have plotted the function $F(d) := ||V^* - V_d||_{L^1(\Lambda,\mathbb{R})}$ *where* V^* *is given in Eqn. (85) and* V_d *is the solution to the SOS Optimization Problem (61) for* $d = 4$ *to* 20, *where* $\Lambda = [-0.5, 0.5]$, $w(x,t) \equiv 1$, $h_\Omega(x) = 2.4^2 - x^2$ *and* $h_U(u) = 1 - u^2$. *All solutions,* V_d, *of Problem (61) were sub-value functions as expected. Moreover, the figure shows by increasing the degree* $d \in \mathbb{N}$ *the resulting sub-VF,* V_d, *better approximates* V^*; *where the* L^1 *error appears to decay at a faster rate than* $1/d^{1.63}$.

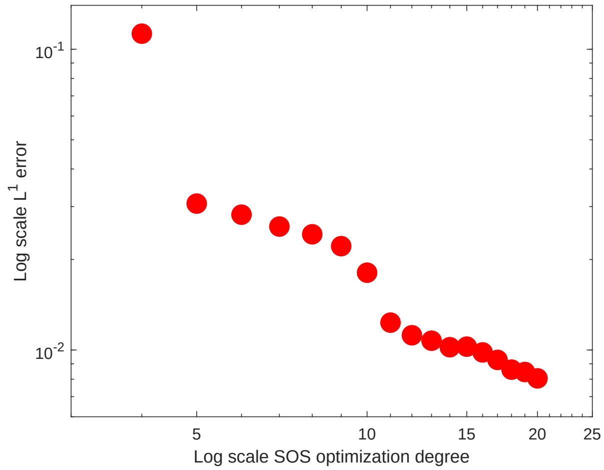

Figure 1. Scatter plot showing $||V^* - V_d||_{L^1(\Lambda,\mathbb{R})}$, where V^* is given in Eqn. (85) and V_d solves the SOS Problem (61) for $d = 4$ to 20.

A. Using SOS Programming To Construct Polynomial Sub-Value Functions For Controller Construction

Given an OCP, in Theorem 4 we showed that the performance of a controller constructed from a candidate VF is bounded by the $W^{1,\infty}$ norm between the true VF of the OCP and the candidate VF. We next demonstrate through numerical examples that the performance of a controller constructed from a typical solutions to the SOS Problem (61) is significantly higher than that predicted by this bound.

Consider tuple $\{c,g,f,\mathbb{R}^n,U,T\} \in \mathscr{M}_{Poly}$, where the cost function is of the form $c(x,u,t) = c_0(x,t) + \sum_{i=1}^m c_i(x,t)u_i$, the dynamics are of the form $f(x,u) = f_0(x) + \sum_{i=1}^m f_i(x)u_i$, and the input constraints are of the form $U = [a_1,b_1] \times ... \times [a_m,b_m]$. Since any rectangular set can be represented as $U = [-1,1]^m$ using the substitution $\tilde{u}_i = \frac{2u_i - 2b_i}{b_i - a_i}$ for $i \in \{1,...,m\}$, without loss of generality we assume $U = [-1,1]^m$. Now, given an OCP associated with $\{c,g,f,\mathbb{R}^n,U,T\} \in \mathscr{M}_{Poly}$ suppose $V \in C^1(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})$ solves the HJB PDE (12), then by Theorem 2 a solution to the OCP initialized at $x_0 \in \mathbb{R}^n$ can be found as

$$\mathbf{u}^*(t) := k(\phi_f(x_0,t,\mathbf{u}^*),t), \text{ where} \qquad (86)$$

$$k(x,t) \in \arg\inf_{u\in[-1,1]^m}\left\{\sum_{i=1}^m c_i(x,t)u_i + \nabla_x V(x,t)^T f_i(x)u_i\right\}. \qquad (87)$$

Since the objective function in Eqn. (87) is linear in the decision variables $u \in \mathbb{R}^m$, and since the constraints have the form $u_i \in [-1,-1]$, it follows that Eqns. (86) and (87) can be reformulated as

$$\mathbf{u}^*(t) := k(\phi_f(x_0,t,\mathbf{u}^*),t), \text{ where} \qquad (88)$$

$$k_i(x,t) = -\operatorname{sign}(c_i(x,t) + \nabla_x V(x,t)^T f_i(x)). \qquad (89)$$

In the following numerical examples we approximately solve OCPs by constructing a controller from the solution P to the SOS Problem (61). We construct such controllers by replacing

V with P in Eqns. (86) and (87). We will consider OCPs with no state constraints and initial conditions inside some set $\Lambda \subseteq \mathbb{R}^n$. We select $\Omega = B(0,R)$ with $R>0$ sufficiently large enough so Eqn. (36) is satisfied. That is, no matter what control we use, the solution map starting from any $x_0 \in \Lambda$ will not be able to leave the state constraint set Ω. In this case the solution to the state constrained problem, $\{c,g,f,\Omega,U,T\}$, is equivalent to the solution of the state unconstrained problem, $\{c,g,f,\mathbb{R}^n,U,T\}$.

Example 2. *Let us consider the following OCP from [29]:*

$$\min_{\mathbf{u}} \int_0^5 x_1(t)dt \tag{90}$$
$$\text{subject to: } \begin{bmatrix}\dot{x_1}(t)\\ \dot{x_2}(t)\end{bmatrix} = \begin{bmatrix}x_2(t)\\ \mathbf{u}(t)\end{bmatrix},\ \mathbf{u}(t)\in[-1,1] \text{ for all } t\in[0,5].$$

We associate this problem with the tuple $\{c,g,f,\Omega,U,T\} \in \mathscr{M}_{Poly}$ *where* $c(x,t)=x_1$, $g(x)\equiv 0$, $f(x,u)=[x_2,u]^T$, $U=[-1,1]$, *and* $T=5$. *By solving the SOS Optimization Problem* (61) *for* $d=3$, $\Lambda=[-0.6,0.6]\times[-1,1]$, $w(x,t)\equiv 1$, $h_\Omega(x)=10^2-x_1^2-x_2^2$ *and* $h_U(u)=1-u^2$, *it is possible to obtain a polynomial sub-value function P. By replacing V with P in Eqns.* (88) *and* (89) *it is then possible to construct a controller,* k_P, *that yields a candidate solution to the OCP as* $\tilde{\mathbf{u}}(t)=k_P(x(t),t)$.

Matlab's function, `ode45`, *can be used to approximately solve the ODE* $\dot{x}(t)=f(x(t),\mathbf{u}(t))$ *for each given* $\mathbf{u}$. *We denote the solution map (Defn. 4) of this ODE by* ϕ_f. *For initial condition* $x_0=[0,1]^T$ *the set* $\{\phi_{\tilde{f}}(x_0,t,\tilde{\mathbf{u}})\in\mathbb{R}^2 : t\in[0,T]\}$ *is shown in the phase plot in Figure 2. We approximate the objective function of Problem* (90) *evaluated at* $\mathbf{u}$ *(ie the cost associated with* $\mathbf{u}$*) using the Riemann sum:*

$$\int_0^T c(\phi_f(x_0,t,\mathbf{u}),t)dt \approx \sum_{i=1}^{N-1} c(\phi_f(x_0,t_i,\mathbf{u}),t_i)\Delta t_i, \tag{91}$$

where $0=t_0<...<t_N=T$ *and* $\Delta t_i=t_{i+1}-t_i$ *for all* $i\in\{1,...,N-1\}$. *For* $N=10^8$ *Eqn.* (91) *was used to find the cost associated with a fixed input,* $\mathbf{u}(t)\equiv 1$, *as* 354.17, *whereas the cost of using* $\mathbf{u}(t)\equiv -1$ *was found to be* 41.67. *The cost of using our derived input* $\tilde{\mathbf{u}}$ *was found to be* 0.2721, *an improvement when compared to the cost* 0.2771 *found in [29].*

Example 3. *Consider an OCP found in [29] and [30] which has the same dynamics as Eqn.* (90) *but a different cost function. The associated tuple is* $\{c,g,f,\Omega,U,T\}\in\mathscr{M}_{Poly}$ *where* $c(x,t)=x_1^2+x_2^2$, $g(x)\equiv 0$, $f(x,u)=[x_2,u]^T$, $U=[-1,1]$, *and* $T=5$. *By solving the SOS Optimization Problem* (61) *for* $d=4$, $\Lambda=[-0.5,1.1]\times[-1.1,.5]$, $w(x,t)\equiv 1$, $h_\Omega(x)=10^2-x_1^2-x_2^2$ *and* $h_U(u)=1-u^2$, *we obtain the polynomial sub-VF P. Similarly to Example 2 we construct a controller from the polynomial sub-VF P using Eqn's* (88) *and* (89). *Using Eqn.* (91), *the fixed input* $\mathbf{u}(t)\equiv +1$ *was found to have cost* 446.03. *The fixed input* $\mathbf{u}(t)\equiv -1$ *cost was found to be* 67.48. *The controller derived from P was found to have cost* 0.7255, *an improvement compared to a cost of* 0.75041 *found in [30] and* 0.8285 *found in [29].*

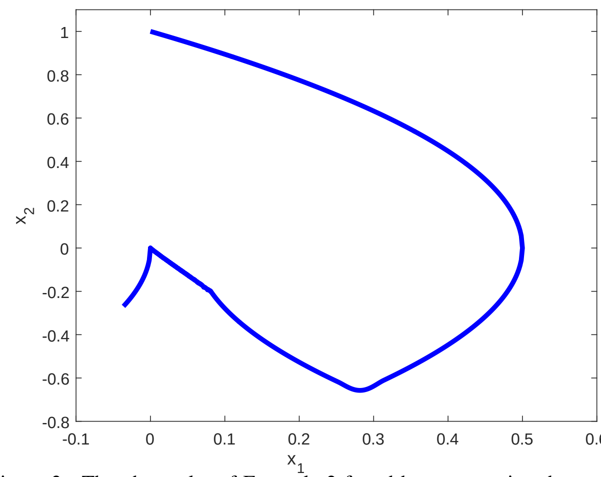

Figure 2. The phase plot of Example 2 found by constructing the controller given in Eqn. (86) using the solution to the SOS Problem (61).

B. Using SOS Programming to Construct Polynomial Sub-Value Functions For Reachable Sets Estimation

Appendix XII shows the sublevel sets of VFs characterize reachable sets. We now numerically solve the SOS programming problem in Eqn. (61) obtaining an approximate VF that can be used to estimate the reachable set of the Lorenz system. The problem of estimating the Lorenz attractor has previously been studied in [31], [15], [32], [33], [34].

Example 4. *Let us consider the Lorenz system defined by the three dimensional second order nonlinear ODE:*

$$\dot{x}_1(t)=\sigma(x_2(t)-x_1(t)),\ \dot{x}_2(t)=x_1(t)(\rho-x_3(t))-x_2(t),$$
$$\dot{x}_3(t)=x_1(t)x_2(t)-\beta x_3(t), \tag{92}$$

where $\sigma=10$, $\beta=8/3$, $\rho=28$. *We make a coordinate change so the Lorenz attractor is located in a unit box by defining*

$$\bar{x}_1 := 50x_1,\quad \bar{x}_2 := 50x_2,\quad \bar{x}_3 := 50x_3+25. \tag{93}$$

The ODE (92) *can then be written in the form* $\dot{x}(t)=\tilde{f}(x(t),\mathbf{u}(t))$ *using* $\tilde{f}(x)=[50\sigma(x_2-x_1),50x_1(\rho-50x_3-50(25))-50x_2,50^2x_1x_2-50\beta x_3-25\beta]^T$. *Note, as* $\tilde{f}$ *is independent of any input* $u\in U$ *without loss of generality we will set* $U=\emptyset$. *The problem of estimating the Lorenz attractor is then equivalent to the problem of estimating* $FR_{\tilde{f}}(\mathbb{R}^n,\mathbb{R}^n,U,\{\infty\})$. *In this section we estimate* $FR_{\tilde{f}}(\mathbb{R}^n,\mathbb{R}^n,U,\{\infty\})$ *by estimating* $FR_{\tilde{f}}(X_0,\Lambda,U,\{T\})$ *for some* $T<\infty$, $\Lambda\subset\mathbb{R}^3$, $X_0:=\{x\in\mathbb{R}^3 : g(x)<0\}$, *and* $g\in\mathscr{P}(\mathbb{R}^n,\mathbb{R})$.

Figure 3 shows the set $\{x\in\mathbb{R}^3 : P(x,0)<0\}$ *where P is the solution to the SOS Optimization Problem* (61) *for* $d=4$, $T=0.5$, $f=-\tilde{f}$, $h_U\equiv 0$, $h_\Omega(x)=1-x_1^2-x_2^2-x_3^2$, $c\equiv 0$, $g(x)=(x_1+0.6)^2+(x_2-0.6)^2+(x_3-0.2)^2-0.1^2$, $\Lambda=[-0.4,0.4]\times[-0.5,0.5]\times[-0.4,0.6]$, *and* $w(x,t)=\delta(t)$ *where* δ *is the Dirac delta function. Prop. 1 shows P is a sub-VF. Then Cor. 1 shows* $BR_f(X_0,\Lambda,U,\{T\})\subseteq\{x\in\mathbb{R}^3 : P(x,0)<0\}$ *and hence* $FR_{\tilde{f}}(X_0,\Lambda,U,\{T\})=BR_f(X_0,\Lambda,U,\{T\})\subseteq\{x\in\mathbb{R}^3 : P(x,0)<0\}$ *by Lem. 6. Thus the* 0*-sublevel set of P contains the forward reachable set. Moreover, Figure 3 provides numerical evidence that the* 0*-sublevel set of P approximates the Lorenz attractor accurately.*

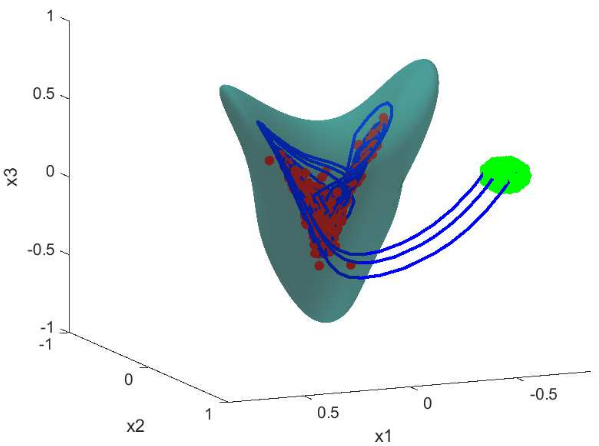

Figure 3. Forward reachable set estimation from Example 4. The transparent cyan set represents the 0-sublevel set of the solution to the SOS Problem (61), the 20^3 green points represent initial conditions, the 20^3 orange points represent where initial conditions transition to after $t = 0.5$ under scaled dynamics from the ODE (92) (found using Matlab's `ODE45` function), and the three blue curves represents three sample trajectories terminated at $t = 0.5$ and initialized at three randomly selected green initial conditions.

Note, given an OCP with VF denoted by V^*, Prop. 5 shows that the sequence of polynomial solutions to the SOS Problem (61), indexed by $d \in \mathbb{N}$, converges to V^* with respect to the L^1 norm as $d \to \infty$. Moreover, Prop. 6 shows that this sequence of polynomial solutions yields a sequence of sublevel sets that converges to $\{x \in \mathbb{R}^n : V^*(x,0) \le 0\}$ with respect to the volume metric as $d \to \infty$. However, Theorem. 5 shows reachable sets are characterized by the "strict" sublevel sets of VFs, $\{x \in \mathbb{R}^n : V^*(x,0) < 0\}$. Counterexample 1 (Appendix XI) shows that a sequence of functions that converges to some function V with respect to the L^1 norm may not yield a sequence of "strict" sublevel sets that converges to the "strict" sublevel set of V. Therefore we conclude that the sequence of "strict" sublevel sets obtained by solving the SOS Problem (61) may in general not converge to the desired reachable set. However, in practice there is often little difference between the sets $\{x \in \mathbb{R}^n : V^*(x,0) \le 0\}$ and $\{x \in \mathbb{R}^n : V^*(x,0) < 0\}$. Example 4 shows how accurate estimates of reachable sets can be obtained by solving the SOS Problem (61). Moreover, these reachable set estimations are guaranteed to contain the true reachable set by Cor. 1, a property useful in safety analysis [12].

X. Conclusion

For a given optimal control problem, we have proposed a sequence of SOS programming problems, each instance of which yields a polynomial, and where the polynomials become increasingly tight approximations to the true value function of the optimal control problem respect to the L^1 norm. Moreover, the sublevel sets of these polynomials become increasingly tight approximations to the sublevel sets of the true value function with respect to the volume metric. Furthermore, we have also shown that a controller can be constructed from a candidate value function that performs arbitrarily close to optimality when the candidate value function approximates the true value function arbitrarily well with respect to the $W^{1,\infty}$ norm. We would like to emphasize that our performance bound, for controllers constructed from candidate value functions, can be applied independently of our proposed SOS algorithm for value function approximation, and therefore is maybe of broader interest.

References

[1] D. Liberzon, *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press, 2011.

[2] D. Gallistl, T. Sprekeler, and E. Süli, "Mixed finite element approximation of periodic Hamilton–Jacobi–Bellman problems with application to numerical homogenization," *arXiv preprint arXiv:2010.01647*, 2020.

[3] Y. Achdou, F. Camilli, and I. Capuzzo Dolcetta, "Homogenization of Hamilton–Jacobi equations: numerical methods," *Mathematical models and methods in applied sciences*, vol. 18, no. 07, pp. 1115–1143, 2008.

[4] D. Kalise and K. Kunisch, "Polynomial approximation of high-dimensional Hamilton–Jacobi–Bellman equations and applications to feedback control of semilinear parabolic PDEs," *SIAM Journal on Scientific Computing*, vol. 40, no. 2, pp. A629–A652, 2018.

[5] W. M. McEneaney, "A curse-of-dimensionality-free numerical method for solution of certain HJB PDEs," *SIAM journal on Control and Optimization*, vol. 46, no. 4, pp. 1239–1276, 2007.

[6] A. Kamoutsi, T. Sutter, P. M. Esfahani, and J. Lygeros, "On infinite linear programming and the moment approach to deterministic infinite horizon discounted optimal control problems," *IEEE control systems letters*, vol. 1, no. 1, pp. 134–139, 2017.

[7] A. Pakniyat and R. Vasudevan, "A convex duality approach to optimal control of killed markov processes," *CDC*, 2019.

[8] M. Korda, D. Henrion, and C. N. Jones, "Controller design and value function approximation for nonlinear dynamical systems," *Automatica*, vol. 67, pp. 54–66, 2016.

[9] Y. Chen and A. D. Ames, "Duality between density function and value function with applications in constrained optimal control and markov decision process," *arXiv preprint arXiv:1902.09583*, 2019.

[10] I. M. Mitchell, A. M. Bayen, and C. J. Tomlin, "A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games," *IEEE Transactions on automatic control*, vol. 50, no. 7, pp. 947–957, 2005.

[11] E. Summers, A. Chakraborty, W. Tan, U. Topcu, P. Seiler, G. Balas, and A. Packard, "Quantitative local L2-gain and reachability analysis for nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 10, pp. 1115–1135, 2013.

[12] H. Yin, A. Packard, M. Arcak, and P. Seiler, "Reachability analysis using dissipation inequalities for nonlinear dynamical systems," *arXiv preprint arXiv:1808.02585*, 2018.

[13] B. Xue, M. Fränzle, and N. Zhan, "Inner-approximating reachable sets for polynomial systems with time-varying uncertainties," *IEEE Transactions on Automatic Control*, 2019.

[14] M. Jones and M. M. Peet, "Relaxing the Hamilton Jacobi Bellman equation to construct inner and outer bounds on reachable sets," *arXiv preprint arXiv:1903.07274*, 2019.

[15] M. Jones and M. M. Peet, "Using SOS and sublevel set volume minimization for estimation of forward reachable sets," *arXiv preprint arXiv:1901.11174*, 2019.

[16] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[17] A. A. Ahmadi and A. Majumdar, "DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization," *SIAM Journal on Applied Algebra and Geometry*, vol. 3, no. 2, pp. 193–230, 2019.

[18] Y. Zhang, M. E. Raoufat, K. Tomsovic, and S. M. Djouadi, "Set theory-based safety supervisory control for wind turbines to ensure adequate frequency response," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 680–692, 2019.

[19] D. P. Bertsekas and J. N. Tsitsiklis, "Neuro-dynamic programming: an overview," in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, vol. 1, pp. 560–564, IEEE, 1995.

[20] A. Bressan, "Viscosity solutions of Hamilton-Jacobi equations and optimal control problems,"

[21] D. P. Bertsekas, *Dynamic programming and optimal control*, vol. 1. Athena scientific Belmont, MA, 2005.

[22] M. G. Crandall, "Viscosity solutions: a primer," in *Viscosity solutions and applications*, pp. 1–43, Springer, 1997.

[23] L. C. Evans, *Partial differential equations*, vol. 19. American Mathematical Society, 2010.

[24] Y. Lin, E. D. Sontag, and Y. Wang, "A smooth converse Lyapunov theorem for robust stability," *SIAM Journal on Control and Optimization*, vol. 34, no. 1, pp. 124–160, 1996.

[25] H. M. Soner, "Optimal control with state-space constraint i," *SIAM Journal on Control and Optimization*, vol. 24, no. 3, pp. 552–561, 1986.
[26] H. Frankowska and M. Mazzola, "Discontinuous solutions of Hamilton–Jacobi–Bellman equation under state constraints," *Calculus of Variations and Partial Differential Equations*, vol. 46, no. 3-4, pp. 725–747, 2013.
[27] H. Frankowska and R. Vinter, "Existence of neighboring feasible trajectories: applications to dynamic programming for state-constrained optimal control problems," *Journal of Optimization Theory and Applications*, vol. 104, no. 1, pp. 20–40, 2000.
[28] A. Altarovici, O. Bokanowski, and H. Zidani, "A general Hamilton-Jacobi framework for non-linear state-constrained control problems," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 2, pp. 337–357, 2013.
[29] D. Jacobson, S. Gershwin, and M. Lele, "Computation of optimal singular controls," *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 67–73, 1970.
[30] S. Dadebo, K. McAuley, and P. McLellan, "On the computation of optimal singular and bang–bang controls," *Optimal Control Applications and Methods*, vol. 19, no. 4, pp. 287–297, 1998.
[31] M. Jones and M. M. Peet, "Using SOS for optimal semialgebraic representation of sets: Finding minimal representations of limit cycles, chaotic attractors and unions," *arXiv preprint arXiv:1809.10308*, 2018.
[32] D. Li, J. Lu, X. Wu, and G. Chen, "Estimating the bounds for the Lorenz family of chaotic systems," *Chaos, Solitons and Fractals*, vol. 23, pp. 529–534, 2005.
[33] T. Wang, S. Lall, and M. West, "Polynomial level-set method for attractor estimation," *Journal of The Franklin Institute*, vol. 349, pp. 2783–2798, 2012.
[34] D. Goluskin, "Bounding extreme values on attractors using sum-of-squares optimization, with application to the Lorenz attractor," *arXiv*, 2018.
[35] J. B. Lasserre, "Tractable approximations of sets defined with quantifiers," *Mathematical Programming*, vol. 151, no. 2, pp. 507–527, 2015.
[36] M. Peet, "Exponentially stable nonlinear systems have polynomial Lyapunov functions on bounded regions," *IEEE Transactions on Automatic Control*, 2009.
[37] J. Malỳ and W. P. Ziemer, *Fine regularity of solutions of elliptic partial differential equations*. No. 51, American Mathematical Soc., 1997.
[38] P. H. Lasz, "Geometric analysis," 2014.
[39] M. Putinar, "Positive polynomials on compact semialgebriac sets.," *Math J*, 1993.
[40] M. Spivak, *Calculus on Manifolds*. Addison-Wesley, 1965.
[41] D. Fischer, "Essential supremum with the continuous function?." Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/618126 (version: 2015-06-09).

Morgan Jones received the B.S. and Mmath in mathematics from The University of Oxford, England in 2016. He is a research associate with Cybernetic Systems and Controls Laboratory (CSCL) in the School for Engineering of Matter, Transport and Energy (SEMTE) at Arizona State University. His research primarily focuses on the estimation of reachable sets, attractors and regions of attraction for nonlinear ODE's. Furthermore, he has an interest in extensions of the dynamic programing framework to non-separable cost functions.

Matthew M. Peet received the B.S. degree in physics and in aerospace engineering from the University of Texas, Austin, TX, USA, in 1999 and the M.S. and Ph.D. degrees in aeronautics and astronautics from Stanford University, Stanford, CA, in 2001 and 2006, respectively. He was a Postdoctoral Fellow at INRIA, Paris, France from 2006 to 2008. He was an Assistant Professor of Aerospace Engineering at the Illinois Institute of Technology, Chicago, IL, USA, from 2008 to 2012. Currently, he is an Associate Professor of Aerospace Engineering at Arizona State University, Tempe, AZ, USA. Dr. Peet received a National Science Foundation CAREER award in 2011.

XI. Appendix A: sublevel set approximation

In this appendix we show that the volume metric (D_V in Eqn. (66)) is indeed a metric. Moreover, in Prop. 7 we show that if $\lim_{d\to\infty} ||J_d - V||_{L^1} = 0$ then for any $\gamma \in \mathbb{R}$ we have $\lim_{d\to\infty} D_V(\{x\in\Lambda : V(x)\le\gamma\},\{x\in\Lambda : J_d(x)\le\gamma\}) = 0$. The sublevel approximation results presented in this appendix are required in the proof of Prop. 6.

Definition 12. *$D : X\times X \to \mathbb{R}$ is a metric if the following is satisfied for all $x,y\in X$,*

- $D(x,y)\ge 0$,
- $D(x,y) = D(y,x)$,
- $D(x,y) = 0$ *iff* $x = y$,
- $D(x,z)\le D(x,y)+D(y,z)$.

Lemma 4 ([31])**.** *Consider the quotient space,*

$$X := \mathscr{B} \pmod{\{X\subset\mathbb{R}^n : X\ne\emptyset, \mu(X)=0\}},$$

recalling $\mathscr{B} := \{B\subset\mathbb{R}^n : \mu(B)<\infty\}$ is the set of all bounded sets. Then $D_V : X\times X\to\mathbb{R}$, defined in Eqn. (66)*, is a metric.*

Lemma 5 ([31])**.** *If $A,B\in\mathscr{B}$ and $B\subseteq A$ then*

$$D_V(A,B) = \mu(A/B) = \mu(A)-\mu(B).$$

Inspired by an argument used in [35] we now show if two functions are close in the L^1 norm then it follows their sublevel sets are close with respect to the volume metric.

Proposition 7. *Consider a set $\Lambda\in\mathscr{B}$, a function $V\in L^1(\Lambda,\mathbb{R})$, and a family of functions $\{J_d\in L^1(\Lambda,\mathbb{R}) : d\in\mathbb{N}\}$ that satisfies the following properties:*

1) *For any $d\in\mathbb{N}$ we have $J_d(x)\le V(x)$ for all $x\in\Lambda$.*
2) $\lim_{d\to\infty} ||V-J_d||_{L^1(\Lambda,\mathbb{R})} = 0$.

Then for all $\gamma\in\mathbb{R}$

$$\lim_{d\to\infty} D_V\bigg(\{x\in\Lambda : V(x)\le\gamma\},\{x\in\Lambda : J_d(x)\le\gamma\}\bigg) = 0. \quad (94)$$

Proof. To prove Eqn. (94) we show for all $\varepsilon>0$ there exists $N\in\mathbb{N}$ such that for all $d\ge N$

$$D_V\bigg(\{x\in\Lambda : V(x)\le\gamma\},\{x\in\Lambda : J_d(x)\le\gamma\}\bigg) < \varepsilon. \quad (95)$$

In order to do this we first denote the following family of sets for each $n\in\mathbb{N}$

$$A_n := \left\{x\in\Lambda : V(x)\le\gamma+\frac{1}{n}\right\}.$$

Since $J_d(x)\le V(x)$ for all $x\in\Lambda$ and $d\in\mathbb{N}$ we have

$$\{x\in\Lambda : V(x)\le\gamma\}\subseteq\{x\in\Lambda : J_d(x)\le\gamma\} \text{ for all } d\in N. \quad (96)$$

Moreover, since $\{x\in\Lambda: V(x)\le\gamma\}\subseteq\Lambda$, $\{x\in\Lambda: J_d(x)\le\gamma\}\subseteq\Lambda$ and $\Lambda\in\mathscr{B}$ it follows $\{x\in\Lambda: V(x)\le\gamma\}\in\mathscr{B}$ and $\{x\in\Lambda: J_d(x)\le\gamma\}\in\mathscr{B}$.

Now for $d\in\mathbb{N}$

$$\begin{aligned}
&D_V\left(\{x\in\Lambda: V(x)\le\gamma\},\{x\in\Lambda: J_d(x)\le\gamma\}\right) \quad (97)\\
&=\mu(\{x\in\Lambda: J_d(x)\le\gamma\})-\mu(\{x\in\Lambda: V(x)\le\gamma\})\\
&=\mu(\{x\in\Lambda: J_d(x)\le\gamma\})-\mu(A_n\cap\{x\in\Lambda: J_d(x)\le\gamma\})\\
&\qquad+\mu(A_n\cap\{x\in\Lambda: J_d(x)\le\gamma\})-\mu(\{x\in\Lambda: V(x)\le\gamma\})\\
&\le\mu(\{x\in\Lambda: J_d(x)\le\gamma\})-\mu(A_n\cap\{x\in\Lambda: J_d(x)\le\gamma\})\\
&\qquad+\mu(A_n)-\mu(\{x\in\Lambda: V(x)\le\gamma\})\\
&=\mu(\{x\in\Lambda: J_d(x)\le\gamma\}/A_n)+\mu(A_n/\{x\in\Lambda: V(x)\le\gamma\}).
\end{aligned}$$

The first equality of Eqn. (97) follows by Lemma 5 (since the sublevel sets of V and J_d are bounded and satisfy Eqn. (96)). The first inequality follows as $A_n\cap\{x\in\Lambda: J_d(x)\le\gamma\}\subseteq A_n$ which implies $\mu(A_n\cap\{x\in\Lambda: J_d(x)\le\gamma\})\le\mu(A_n)$. The third equality follows using Lemma 5 and since $A_n\cap\{x\in\Lambda: J_d(x)\le\gamma\}\subseteq\{x\in\Lambda: J_d(x)\le\gamma\}$ and $\{x\in\Lambda: V(x)\le\gamma\}\subseteq A_n$.

To show that Eqn. (95) holds for any $\varepsilon>0$ we will split the remainder of the proof into two parts. In Part 1 we show that there exists $N_1\in\mathbb{N}$ such that $\mu(A_n/\{x\in\Lambda: V(x)\le\gamma\})<\frac{\varepsilon}{2}$ for all $n\ge N_1$. In Part 2 we show that for any $n\in\mathbb{N}$ there exists $N_2\in\mathbb{N}$ such that $\mu(\{x\in\Lambda: J_d(x)\le\gamma\}/A_n)<\frac{\varepsilon}{2}$ for all $d\ge N_2$.

Part 1 of proof: In this part of the proof we show that there exists $N_1\in\mathbb{N}$ such that $\mu(A_n/\{x\in\Lambda: V(x)\le\gamma\})<\frac{\varepsilon}{2}$ for all $n>N_1$.

Since $\cap_{n=1}^{\infty}A_n=\{x\in\Lambda: V(x)\le\gamma\}$ and $A_{n+1}\subseteq A_n$ for all $n\in\mathbb{N}$ we have that $\mu(\{x\in\Lambda: V(x)\le\gamma\})=\mu(\cap_{n=1}^{\infty}A_n)=\lim_{n\to\infty}\mu(A_n)$ (using the "continuity from above" property of measures). Thus there exists $N_1\in\mathbb{N}$ such that

$$|\mu(A_n)-\mu(\{x\in\Lambda: V(x)\le\gamma\})|<\frac{\varepsilon}{2}\text{ for all }n>N_1.$$

Therefore it follows

$$\begin{aligned}
&\mu(A_n/\{x\in\Lambda: V(x)\le\gamma\})\\
&\qquad=\mu(A_n)-\mu(\{x\in\Lambda: V(x)\le\gamma\})<\frac{\varepsilon}{2}\text{ for all }n>N_1.
\end{aligned}$$

Part 2 of proof: For fixed $n>N_1$ we now show there exists $N_2\in\mathbb{N}$ such that $\mu(\{x\in\Lambda: J_d(x)\le\gamma\}/A_n)<\frac{\varepsilon}{2}$ for all $d\ge N_2$.

Now

$$\{x\in\Lambda: J_d(x)\le\gamma\}/A_n\subseteq\{x\in\Lambda: n|J_d(x)-V(x)|\ge1\}\quad(98)$$
$$\text{for all } d\in\mathbb{N}.$$

The set containment in Eqn. (98) follows since if $y\in\{x\in\Lambda: J_d(x)\le\gamma\}/A_n$ then $y\in\Lambda$, $J_d(y)\le\gamma$ and $y\notin A_n$. Since $y\notin A_n$ we have $V(y)>\gamma+\frac{1}{n}$. Thus

$$n|J_d(y)-V(y)|\ge n(V(y)-J_d(y))\ge n\left(\gamma+\frac{1}{n}-\gamma\right)=1,$$

which implies $y\in\{x\in\Lambda: n|J_d(x)-V(x)|\ge1\}$.

Since $\lim_{d\to\infty}\int_\Lambda|V(x)-J_d(x)|dx=0$ there exists $N_2\in\mathbb{N}$ such that

$$\int_\Lambda|V(x)-J_d(x)|dx<\frac{\varepsilon}{2n}\text{ for all }d\ge N_2.\quad(99)$$

Therefore,

$$\begin{aligned}
\mu(\{x\in\Lambda: J_\delta(x)\le\gamma\}/A_n)&\le\mu(\{x\in\Lambda: n|J_\delta(x)-V(x)|\ge1\})\\
&\le\int_\Lambda n|J_\delta(x)-V(x)|dx<\frac{\varepsilon}{2}\text{ for }d\ge N_2.\quad(100)
\end{aligned}$$

The first inequality in Eqn. (100) follows by Eqn. (98). The second inequality follows by Chebyshev's inequality (Lemma 8). The third inequality follows by Eqn. (99). ☐

Prop. 7 shows if a sequence of functions $\{J_d\}_{d\in\mathbb{N}}$ converges from bellow to some function V with respect to the L^1 norm then the sequence sublevel sets $\{x\in\Lambda: J_d(x)\le\gamma\}$ converge to $\{x\in\Lambda: V(x)\le\gamma\}$ with respect to the volume metric. However, this does not imply the sequence of "strict" sublevel sets $\{x\in\Lambda: J_d(x)<\gamma\}$ converge to $\{x\in\Lambda: V(x)<\gamma\}$ (even if $\{J_d\}_{d\in\mathbb{N}}$ converges from bellow to V with respect to the L^∞ norm). To see this we next consider a counterexample where $\{J_d\}_{d\in\mathbb{N}}$ is a family of functions that can uniformly approximate some given $V:\in Lip((0,1),\mathbb{R})$ but $\{x\in\Lambda: J_d(x)<\gamma\}$ does not converge to $\{x\in\Lambda: V(x)<\gamma\}$.

Counterexample 1. *We show there exists $\gamma\in\mathbb{R}$, $\Lambda\subset\mathbb{R}$, $V\in Lip(\Lambda,\mathbb{R})$ and $\{J_d\}_{d\in\mathbb{N}}\subset Lip(\Lambda,\mathbb{R})$ such that $J_d(x)\le V(x)$ for all $x\in\Lambda$ and $\lim_{d\to\infty}\int_\Lambda|V(x)-J_d(x)|dx=0$ but*

$$\lim_{d\to\infty}D_V\left(\{x\in\Lambda: V(x)<\gamma\},\{x\in\Lambda: J_d(x)<\gamma\}\right)\neq0$$

Let

$$\Lambda=(0,1),\quad V(x)=\begin{cases}0 \text{ if } x\in(0,0.25]\\ 2(x-0.25)\text{ if }x\in(0.25,0.75),\\ 1\text{ if }x\in[0.75,1)\end{cases}$$

$$J_d(x)=\begin{cases}0\text{ if }x\in(0,0.25]\\ 2(x-0.25)\text{ if }x\in(0.25,0.75),\\ 1-\frac{1}{d}\text{ if }x\in[0.75,1)\end{cases}\quad\gamma=1.$$

Now for all $d\in\mathbb{N}$ it is clear that we have $J_d(x)\le V(x)$ and $V(x)-J_d(x)<\frac{1}{d}$ for all $x\in\Lambda$. This implies

$$\lim_{d\to\infty}\int_\Lambda V(x)-J_d(x)dx\le\lim_{d\to\infty}\sup_{x\in\Lambda}(V(x)-J_d(x))\le\lim_{d\to\infty}\frac{1}{d}=0.$$

However $\{x\in\Lambda: V(x)<\gamma\}=(0,0.75)$ and for all $d\in\mathbb{N}$ $\{x\in\Lambda: J_d(x)<\gamma\}=(0,1)$. Therefore

$$\begin{aligned}
&D_V(\{x\in\Lambda: V(x)<\gamma\},\{x\in\Lambda: J_d(x)<\gamma\})\\
&\qquad=D_V((0,0.75),(0,1))=0.25\text{ for all }d\in\mathbb{N}.
\end{aligned}$$

Hence,

$$\lim_{d\to\infty}D_V\left(\{x\in\Lambda: V(x)<\gamma\},\{x\in\Lambda: J_d(x)<\gamma\}\right)=0.25\neq0.$$

XII. Appendix B: Value Functions Characterize Reachable Sets

In this appendix we present several reachable set results required in our numerical approximation of the Lorenz attractor (Example 4). Similarly to forward reachable sets (Defn. 10) we now define backward reachable sets.

Definition 13. *For $X_0\subset\mathbb{R}^n$, $\Omega\subseteq\mathbb{R}^n$, $U\subset\mathbb{R}^m$, $f:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$ and $S\subset\mathbb{R}^+$, let*

$$BR_f(X_0,\Omega,U,S) := \Bigg\{ y \in \mathbb{R}^n \, : \, \text{there exists } x \in X_0, T \in S,$$
$$\text{and } \mathbf{u} \in \mathscr{U}_{\Omega,U,f,T}(y,0) \text{ such that } \phi_f(y,T,\mathbf{u}) = x \Bigg\}.$$

Theorem 5 (VFs characterize backward reachable sets [14])**.** *Given $\{0,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ define $X_0 := \{x \in \mathbb{R}^n : g(x) < 0\}$. Then*

$$BR_f(X_0,\Omega,U,\{T\}) = \{x \in \Omega : V^*(x,0) < 0\}, \tag{101}$$

where $V^* : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is any function that satisfies Eqn. (10).

Corollary 1 (Sub-VFs contain reachable sets)**.** *Given $\{0,g,f,\Omega,U,T\} \in \mathscr{M}_{Lip}$ and suppose $V_l : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is a sub-VF (Defn. 7), then*

$$BR_f(X_0,\Omega,U,\{T\}) \subseteq \{x \in \Omega : V_l(x,0) < 0\}, \tag{102}$$

where $X_0 := \{x \in \mathbb{R}^n : g(x) < 0\}$.

Lemma 6 (Equivalence of computation of backward and forward reachable sets [14])**.** *Suppose $X_0 \subset \mathbb{R}^n$, $\Omega \subset \mathbb{R}^n$, $U \subset \mathbb{R}^m$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, and $T \in \mathbb{R}^+$. Then*

$$FR_{-f}(X_0,\Omega,U,\{T\}) = BR_f(X_0,\Omega,U,\{T\}).$$

XIII. Appendix C

In this appendix we present several miscellaneous results required in various places throughout the paper and not previously found in any of the other appendices.

Theorem 6 (Polynomial Approximation [36])**.** *Let $E \subset \mathbb{R}^n$ be an open set and $f \in C^1(E,\mathbb{R})$. For any compact set $K \subseteq E$ and $\varepsilon > 0$ there exists $g \in \mathscr{P}(\mathbb{R}^n,\mathbb{R})$ such that*

$$\sup_{x \in K} |D^\alpha f(x) - D^\alpha g(x)| < \varepsilon \text{ for all } |\alpha| \le 1.$$

Theorem 7 (Rademacher's Theorem [37] [23])**.** *If $\Omega \subset \mathbb{R}^n$ is an open subset and $V \in Lip(\Omega,\mathbb{R})$, then V is differentiable almost everywhere in Ω with point-wise derivative corresponding to the weak derivative almost everywhere; that is the set of points in Ω where V is not differentiable has Lebesgue measure zero. Moreover,*

$$\operatorname*{ess\,sup}_{x \in \Omega} \left| \frac{\partial}{\partial x_i} V(x) \right| \le L_V \text{ for all } 1 \le i \le n,$$

where $L_V > 0$ is the Lipschitz constant of V and $\frac{\partial}{\partial x_i} V(x)$ is the weak derivative of V.

Lemma 7 (Infimum of family of Lipshitz functions is Lipschitz [38])**.** *Suppose $\{h_\alpha\}_{\alpha \in I} \subset LocLip(\mathbb{R}^n,\mathbb{R})$ is a family of locally Lipschitz continuous functions. Then $h : \mathbb{R}^n \to \mathbb{R}$ defined as $h(x) := \inf_{\alpha \in I} h_\alpha(x)$ is such that $h \in LocLip(\mathbb{R}^n,\mathbb{R})$ provided there exists $x \in \mathbb{R}^n$ such that $h(x) < \infty$.*

Theorem 8 (Putinar's Positivstellesatz [39])**.** *Consider the semialgebriac set $X = \{x \in \mathbb{R}^n : g_i(x) \ge 0 \text{ for } i = 1,...,k\}$. Further suppose $\{x \in \mathbb{R}^n : g_i(x) \ge 0\}$ is compact for some $i \in \{1,..,k\}$. If the polynomial $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $f(x) > 0$ for all $x \in X$, then there exists SOS polynomials $\{s_i\}_{i \in \{1,..,m\}} \subset \Sigma_{SOS}$ such that,*

$$f - \sum_{i=1}^m s_i g_i \in \sum_{SOS}.$$

Definition 14. *Let $\Omega \subset \mathbb{R}^n$. We say $\{U_i\}_{i=1}^\infty$ is an open cover for Ω if $U_i \subset \mathbb{R}^n$ is an open set for each $i \in \mathbb{N}$ and $\Omega \subseteq \{U_i\}_{i=1}^\infty$.*

Theorem 9 (Existence of Partitions of Unity [40])**.** *Let $E \subseteq \mathbb{R}^n$ and let $\{E_i\}_{i=1}^\infty$ be an open cover of E. Then there exists a collection of $C^\infty(E,\mathbb{R})$ functions, denoted by $\{\psi\}_{i=1}^\infty$, with the following properties:*

1) *For all $x \in E$ and $i \in \mathbb{N}$ we have $0 \le \psi_i(x) \le 1$.*
2) *For all $x \in E$ there exists an open set $S \subseteq E$ containing x such that all but finitely many ψ_i are 0 on S.*
3) *For each $x \in E$ we have $\sum_{i=1}^\infty \psi_i(x) = 1$.*
4) *For each $i \in \mathbb{N}$ we have $\{x \in E : \psi_i(x) \neq 0\} \subseteq E_i$.*

Lemma 8 (Chebyshev's Inequality)**.** *Let (X,Σ,μ) be a measurable space and $f \in L^1(X,\mathbb{R})$. For any $\varepsilon > 0$ and $0 < p < \infty$,*

$$\mu(\{x \in X : |f(x)| > \varepsilon\}) \le \frac{1}{\varepsilon^p} \int_X |f(x)|^p dx.$$

Lemma 9 (Equivalence of essential supremum and supremum [41])**.** *Let $E \subset \mathbb{R}^n$ be an open set and $f \in C(E,\mathbb{R})$. Then $\operatorname{ess\,sup}_{x \in E} |f(x)| = \sup_{x \in E} |f(x)|$.*