An Optimal Solution to Infinite Horizon Nonlinear Control Problems

Mohamed Naveed Gul Mohamed, Raman Goyal, Suman Chakravorty

Abstract— In this paper, we consider the infinite horizon optimal control problem for nonlinear systems. Under the conditions of controllability of the linearized system around the origin, and nonlinear controllability of the system to a terminal set containing the origin, we establish an approximate regularized solution approach consisting of a "finite free final time" optimal transfer problem to the terminal set, and an infinite horizon linear regulation problem within the terminal set, that is shown to render the origin globally asymptotically stable. Further, we show that the approximations converge to the true optimal cost function as the size of the terminal set decreases to zero. The approach is empirically evaluated on the pendulum and cart-pole swing-up problems to show that the finite time transfer is far shorter than the effective horizon required to solve the infinite horizon problem without the proposed regularization.

Index Terms— Nonlinear control, Infinite horizon optimal control, Control Lyapunov function

I. INTRODUCTION

The goal of an optimal control problem is to find the control inputs that minimize a given cost function subject to constraints on the system dynamics. Further, it is desired that the optimal control problem results in a globally asymptotically stable closed-loop system. In order to satisfy the global asymptotic stability requirement, one has to pose an infinite horizon control problem, i.e., where the cost is an infinite sum. Although there exists infinite horizon optimal control law for specific, mostly linear, systems, obtaining an optimal solution for an infinite horizon nonlinear control problem is a very challenging task owing to the infinite horizon [1]. In this work, we propose an approximate solution to the infinite horizon optimal control problem by turning the problem into an equivalent finite horizon problem.

The infinite horizon optimal control problem may equivalently be posed as the stationary Dynamic Programming (DP) problem for discrete-time systems, or the stationary Hamilton-Jacobi-Bellman (HJB) equation in continuous time systems [1, 2] which turns the sequential decision making problem into a single step decision making problem given one knows the optimal cost function. It is also well known that the optimal solution to the stationary HJB equation is a globally asymptotically stabilizing control Lyapunov function [3, 4, 5]. However, the solution is computationally intractable owing to Bellman's dreaded "curse of dimensionality" [1, 2]. Thus, there is a very large literature on Approximate DP (ADP) and Reinforcement Learning that seeks to alleviate the curse of dimensionality. Approximate dynamic

programming methods [6, 7] typically use aggregation methods [8, 9] or various function approximation methods [10, 11] to give an approximately optimal policy/value function with a high confidence. Reinforcement learning [12] and approximate dynamic programming are typically used in the same sense, though the keyword reinforcement learning is often associated with a model-free approach where one does not have "explicit" knowledge of the model, and instead, seeks to improve the control policy by repeated interactions with the environment while observing the system's responses. The repeated interactions, or learning trials, allow these algorithms to compute the solution of the dynamic programming problem (optimal value/Q-value function or optimal policy) without explicitly constructing the model of the unknown dynamical system. Standard RL algorithms are broadly divided into value-based methods, like Q-learning [13], and policy-based methods, like policy gradient algorithms [14]. Recently, function approximation using deep neural networks has significantly improved the performance of reinforcement learning algorithms, leading to a growing class of literature on 'deep reinforcement learning' [15, 16, 17, 18]. Despite the success on relatively higher dimensional problems than previously possible, the inherent variance in the solution [19, 20] renders them unreliable and the training time required of these methods remain prohibitive.

An alternative "direct" approach to the HJB is to solve the underlying infinite horizon optimal control problem given a particular initial state. The field of Model Predictive Control (MPC) takes this approach to solving the infinite horizon problem, however, owing to the infinite horizon nature of the involved optimal control problem, MPC solves a "fixed final time" finite horizon problem in its stead, takes the first control action, and repeats the process once at the next state [21, 22]. The "stabilizing ingredient" for the asymptotic stability of the MPC approach is the use of a suitable terminal cost function in the optimization problem that is a control Lyapunov function for the system in some terminal set containing the origin [21]. Nonetheless, the domain of attraction of the MPC law can be undesirably small, and thus, different methods have been suggested to increase the domain of attraction [23, 24, 25]. Alternatively, one can eschew the use of a terminal cost function and set using a suitable long horizon [26], but this typically leads to intractability owing to very long prediction horizons [27]. Our approach to solving the infinite horizon problem is similar to MPC in that we "directly" solve the optimal control problem but the key difference is that given a state, we solve a "free-final time" problem for insertion into a terminal set in which the optimal linear feedback law is

The authors are with the Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843, USA. {naveed, ramaniitrgoyal92, schakrav}@tamu.edu

asymptotically stabilizing and/or a good approximation of the true optimal cost function. We can then establish the global asymptotic stability of the resulting feedback law under a mild nonlinear controllability assumption, and also show that the approximation converges to the true optimal cost function as the size of the terminal set is reduced to zero. Finally, there is no need for replanning in the approach. The limitation is that we do not consider state or control constraints in the problem.

The primary contribution of this paper is a tractable direct approach for the solution of infinite horizon optimal control problems that is globally asymptotically stabilizing for nonlinear systems under the conditions of controllability of the system linearization around the origin, and a nonlinear controllability assumption into a terminal set containing the origin. The rest of the paper is organized as follows: we introduce the problem in Section II, the solution approach is detailed in Section III including the heuristic idea behind the construction, and the method is tested on empirically on several nonlinear systems in Section IV.

II. PRELIMINARIES

Consider the following problem which we want to solve:

$$J_\infty^*(x) = \min_{\{u_t\}} \sum_{t=0}^{\infty} c(x_t, u_t); \quad \text{given } x_0 = x \quad (1a)$$

subject to the dynamics: $x_{t+1} = f(x_t, u_t),$ (1b)

where $x_t \in \mathbb{R}^n$ represents the state of the dynamical system, $u_t \in \mathbb{R}^p$ represents the control input to the dynamical system, and $c(x_t, u_t)$ is the incremental cost incurred in taking control action u_t at state x_t. The above problem is an infinite horizon optimal control problem (IH-OCP), and thus, solving the problem is, in general, intractable owing to the infinite horizon of the problem.

Our goal in this work is to develop a tractable approach to solving the above problem by transforming the problem into a suitable finite horizon problem.

Given that we can obtain a solution to the IH-OCP Eq. (1), it is well known that the infinite horizon cost-to-go $J_\infty(\cdot)$ satisfies Bellman's equation [1, Ch.7]:

$$J_\infty^*(x) = \min_u \{c(x, u) + J_\infty^*(f(x, u))\}. \quad (2)$$

Given that $c(x, u) > 0 \ \forall \ (x, u) \neq (0, 0)$, one may see that $J_\infty^*(\cdot)$ is a Control Lyapunov Function (CLF) for the dynamical system in Eq. (1b), and thus, the control feedback policy implicitly defined by the optimal cost-to-go function $J_\infty^*(\cdot)$, globally asymptotically stabilizes the dynamical system Eq. (1b). The proof is straightforward and shown below.

Corollary 1: Let $J_\infty^*(x)$ satisfy the Bellman equation Eq. (2), then it is a control Lyapunov function for the system in Eq. (1b) that renders the origin globally asymptotically stable.

Proof: Let u_t^* be the control action at state x_t according to the Bellman equation, where (x_t, u_t^*) is the system

trajectory given $x_0 = x$. Then:

$$J_\infty^*(x_t) = c(x_t, u_t^*) + J_\infty^*(f(x_t, u_t^*)).$$
$$\Rightarrow J_\infty^*(x_t) > J_\infty^*(x_{t+1}) \quad (\text{since } c(x_t, u_t^*) > 0).$$
$$\Rightarrow J_\infty^*(x_t) \to 0 \text{ as } t \to \infty \quad ,$$

where the last line follows from the fact that $J_\infty^*(x_t)$ is bounded below by zero, and thus, strict monotonicity of the sequence $\{J_\infty^*(x_t)\}$ implies that the sequence has to converge to zero. This holds for any choice of x. Since $J_\infty^*(x') = 0$ only at the origin, this implies the trajectory $x_t \to 0$ as $t \to \infty$ and the result follows. ∎

Further, suppose that if there exists a $J_\infty(\cdot)$ such that it satisfies the Bellman Eq. (not necessarily optimal)

$$J_\infty(x) = \min_u \{c(x, u) + J_\infty(f(x, u))\}, \quad (3)$$

then $J_\infty(\cdot)$ also is a CLF that renders the origin globally asymptotically stable (GAS).

Thus, another goal for us in solving Eq. (1) is to construct CLFs as in Eq. (2)/ (3), such that they render the origin GAS.

III. SOLUTION TO THE INFINITE HORIZON OPTIMAL CONTROL PROBLEM

Define the following finite-horizon optimal control problem (FH-OCP):

$$J_\infty^T(x) = \min_{\{u_t\}} \sum_{t=0}^{T-1} c(x_t, u_t) + \bar{J}_\infty(x_T), \quad (4)$$

subject to: $x_{t+1} = f(x_t, u_t)$

where $\bar{J}_\infty(\cdot)$ is a terminal cost function that shall be defined below. We shall make the following assumptions for the rest of the paper.

Assumption 1: We assume that the cost function $c(x, u)$ has a global minimum at $(x, u) = (0, 0)$, i.e., $\frac{\partial c}{\partial x}\big|_{x=0, u=0} = 0$ and $\frac{\partial c}{\partial u}\big|_{x=0, u=0} = 0$, $c(0, 0) = 0$, and $c(x, u) > 0$ $\forall \ (x, u) \neq (0, 0)$.

Assumption 2: We assume that given any $x \in \mathcal{X}$, and any $\Omega \subset \mathcal{X}$, such that the origin is in Ω, $\exists$ a control sequence $\{u_t\}_{t=0}^{T(x)}$, that ensures $x_{T(x)} \in \Omega$ for some $T(x) < \infty$, under the dynamics defined above (Eq. (1b)).
Assumption 2 is a controllability assumption that ensures that any state can be controlled into entering the region Ω in finite time.

Assumption 3: We assume that the linearization of the dynamical system Eq. (1b) around $(0, 0)$, is controllable.

Given assumptions 1 and 3, we can define the optimal "linear" infinite-horizon problem:

$$\bar{J}_\infty(x) = \min_{\{u_t\}} \sum_{t=0}^{\infty} (x_t' Q x_t + u_t' R u_t), \quad (5a)$$

subject to: $x_{t+1} = A x_t + B u_t,$ (5b)

where, (Q, R) and (A, B) are obtained by performing a quadratic expansion of $c(x, u)$, and a linear expansion of the dynamics in Eq. (1b) around the origin $(x, u) = (0, 0)$.

Note that owing to the linear controllability assumption 3, $\bar{J}_\infty(\cdot)$ above may be found by solving the stationary algebraic Riccati equation, resulting in $\bar{J}_\infty(x) = x'P_\infty x$, where P_∞ is the solution of the stationary Riccati equation.

Definition 1: The terminal cost $\bar{J}_\infty(\cdot)$ in the finite horizon optimal control problem (Eq. (4)) above is defined as $\bar{J}_\infty(x) = x'P_\infty x$, where P_∞ is the stationary Riccati equation's solution.

A. Heuristic Idea

The heuristic idea behind the solution of the infinite horizon optimal control problem is as follows. Let the optimal trajectory be given by (x_t^*, u_t^*) starting at some state x. For the optimal cost to be well defined, we need that $\sum_{t=0}^{\infty} c(x_t^*, u_t^*) < \infty$. However, for this infinite sum to be well defined, the tail sum $\sum_{t>T}^{\infty} c(x_t^*, u_t^*) \to 0$ as $T \to \infty$. Thus, after some finite time, the cost $c(x_t^*, u_t^*)$ is necessarily small. Given that due to assumption 1 the cost $c(x, u) = 0$ only at the origin, it follows that the system spends a very large time around the origin. Next, given assumption 3 that the linearization around the origin is controllable, it follows that the tail sum of the cost can be well approximated by the optimal linear cost function found by solving the stationary Riccati equation, $\bar{J}_\infty$, as defined above. Thus, the infinite horizon cost may be split into two parts: a "transfer cost" to get the system into the region Ω where the linear approximation holds, and a linear regulation cost for regulation to the origin once within Ω (see Fig. 1). Thus, the basic idea is to turn the infinite horizon problem into a finite horizon problem by approximating the tail sum of the cost by the optimal linear cost function and find the optimal insertion time $T^*(x)$ from state x into the region Ω along with the associated optimal control. As the region Ω gets smaller, we get better approximations of the optimal cost function and obtain the optimum in the limit. However, even for a finite Ω, this procedure results in the construction of a CLF that renders the origin GAS. In fact, we can get an uncountable number of such CLFs based on the choice of our incremental cost function $c(x, u)$.
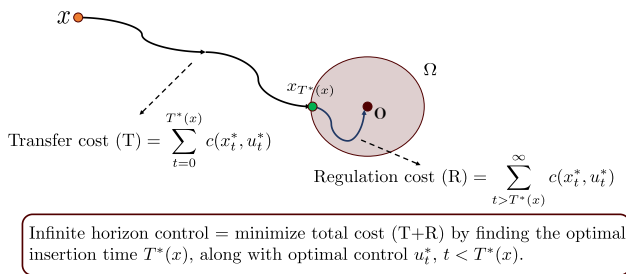

Fig. 1: Schematic illustrating the strategy to solve the infinite horizon optimal control problem

B. Optimality of the Finite Horizon Optimal Control Problem

In the following, we shall establish the optimality of the FH-OCP (4), in particular, that $J_\infty^T(x) \to J_\infty^*(x)$ for all $x \in \mathcal{X}$.

To this end, define the set:

$\Omega_\varepsilon = \{x : |\bar{J}_\infty(x) - J_\infty^*(x)| < \varepsilon$, and $\bar{J}_\infty(x) = x'P_\infty x \leq M_\varepsilon\}$, i.e., M_ϵ is chosen such that in the level set Ω_ε, the error between the linear optimal cost function and the true optimal cost function is less than the tolerance ϵ. The set above defines a sub-level set of the optimal linear cost function in which the cost function is $\varepsilon-$close to the true (unknown) optimum $J_\infty^*(x)$.

For simplicity, we will consider the cost $c(x, u)$ to be quadratic in the following. Due to paucity of space, the proofs for the results below are given in the extended version [28].

Lemma 1: There exists a finite time $T_\varepsilon < \infty$, such that the solution to the FH-OCP (4) for $T = T_\varepsilon$, denoted by $(x_t^\varepsilon, u_t^\varepsilon)$ is such that $\bar{J}_\infty(x_T) = x_T^{\varepsilon'} P_\infty x_T^\varepsilon \leq M_\varepsilon$ for the first time, i.e., for any $T < T_\varepsilon$, $\bar{J}_\infty(x_T) = x_T^{\varepsilon'} P_\infty x_T^\varepsilon > M_\varepsilon$.

The basic idea in the proof for the result above is that if the optimal trajectory never enters Ω_ε, $J^T \to \infty$ starting from x, however, since there is a feasible trajectory with $\bar{J} < \infty$, we can always find a suitable finite time T where the cost of the optimal control is above $\bar{J}$ thereby contradicting optimality at that time T.

Lemma 2: Given the time T_ε that the solution to Eq. (4) first hits the set Ω_ε, the optimal solution T_ε satisfies: $|J_\infty^{T_\varepsilon}(x) - J_\infty^*(x)| \leq \varepsilon$.

Corollary 2: Let $T' < T$, then the solution to Eq. (4) for T, T' follow: $J_\infty^{T'}(x) \geq J_\infty^T(x)$. The proof follows from the last part of the proof to lemma 2 above.

Theorem 1: Given any $x \in \mathcal{X}$, $\lim_{T \to \infty} J_\infty^T(x) \to J_\infty^*(x)$.

Proof: We just showed in lemma 2 that given any $\varepsilon > 0$, we may find $T_\varepsilon(x) < \infty$ such that: $|J_\infty^{T_\varepsilon}(x) - J_\infty^*(x)| \leq \varepsilon$. Further, owing to corollary 2, $J_\infty^T(x) \leq J_\infty^{T_\varepsilon}(x)$ for any $T > T_\varepsilon$. Noting that $J_\infty^{T_\varepsilon}(x) > J_\infty^*(x)$, it follows that $|J_\infty^T(x) - J_\infty^*(x)| \leq \varepsilon \ \forall \ T > T_\varepsilon(x)$, this establishes that $J_\infty^T(x) \to J_\infty^*(x)$ for any $x \in \mathcal{X}$. $\blacksquare$

C. An Alternative Construction

In the following, we define a sub-optimal construction that nonetheless furnishes a Control Lyapunov Function (CLF) that renders the origin globally asymptotically stable for the dynamical system Eq. (1b).

Let the region of the attraction (ROA) of the linear controller corresponding to the terminal cost function $\bar{J}_\infty$ be $\mathcal{D}$. Suppose that we define a set $\Omega \subset \mathcal{D}$ as follows: $\Omega = \{x : \bar{J}_\infty(x) = x'P_\infty x \leq M\}$, which is a level set of $\bar{J}_\infty(\cdot)$ that is constrained in the ROA $\mathcal{D}$. Fix $M < \infty$, and consider the problem:

$$J_\infty(x) = \min_{u_t, T} \sum_{t=0}^{T-1} c(x_t, u_t) + \max\left(\bar{J}_\infty(x_T), M\right), \ x \notin \Omega,$$

$$J_\infty(x) = \bar{J}_\infty(x) = x'P_\infty x, \ x \in \Omega. \tag{6}$$

Please note that the above optimization problem has a free final time T that needs to be optimized over in conjunction with the control action u_t. The rationale behind posing the above problem is that the optimal control transfers the system from any state x to the boundary $\partial\Omega$ of the set Ω, from where the linear controller takes over to regulate to the origin. Since Ω is within the region of attraction $\mathcal{D}$, the linear controller can always regulate to the origin regardless of the insertion point on the boundary.

In the next result, we show that the optimal time for the problem posed above is the first time that the solution to the problem posed above hits the boundary of the set Ω.

Lemma 3: The optimal time $T^*(\Omega)$ for the optimal control problem (Eq. (6)) is finite and is the first time T that the solution to Eq. (6) hits the set Ω.

Proof: Note that Eq. (6) is the exact same as Eq. (4) as long as $x_T \notin \Omega$. Thus, lemma 1 implies that the first hitting time of Ω, say $T^*(\Omega)$, is finite. Furthermore, corollary 2 implies that $J_\infty^T(x) < J_\infty^{T(\Omega)}(x)$ for any $T > T(\Omega)$, where:

$$J_\infty^T(x) = \min_{u_t} \sum_{t=0}^{T-1} c(x_t, u_t) + \bar{J}_\infty(x_T). \tag{7}$$

Let $T > T(\Omega)$ and let the optimal trajectory be $\{\tilde{x}_t, \tilde{u}_t\}$ for the time T according to the OCP in Eq. (6). Then:

$$J_\infty^T(x) = \sum_{t=0}^{T-1} c(\tilde{x}_t, \tilde{u}_t) + \max\left(\bar{J}_\infty(\tilde{x}_T), M\right). \tag{8}$$

Since $T > T(\Omega)$, $\bar{J}_\infty(\tilde{x}_T) < M$, and thus, $\max(\bar{J}_\infty(\tilde{x}_T), M) = M$. Noting that $T(\Omega)$ is the first time the solution to Eq. (6) hits $\partial\Omega$, $J_\infty^T(x) > J_\infty^{T(\Omega)}(x)$, and the result follows. ∎

Next, we show that the sub-optimal cost function thus constructed satisfies the Bellman equation.

Lemma 4: The optimal cost $J_\infty(x)$ corresponding to OCP Eq. (6) satisfies the Bellman equation for any $x \notin \Omega$:

$$J_\infty(x) = \min_u[c(x, u) + J_\infty(f(x, u))].$$

Proof: Let us denote the terminal cost function in the optimal control problem Eq. (6) as $\Phi(x_T) = \max\left(\bar{J}_\infty(x_T), M\right)$. Then for any $x \notin \Omega$:

$$J_\infty(x) = \min_{u_t, T}\left[\sum_{t=0}^{T-1} c(x_t, u_t) + \Phi(x_T)\right], \tag{9}$$

$$= \min_{u_0}\min_{\{u_t\}_{t=1}^{T-1}, T}\left[c(x, u_0) + \sum_{t=1}^{T-1} c(x_t, u_t) + \Phi(x_T)\right],$$

$$= \min_{u_0}\left[c(x, u_0) + \min_{\{u_t\}_{t=1}^{T-1}, T}\sum_{t=1}^{T-1}[c(x_t, u_t) + \Phi(x_T)]\right].$$

Noting that $x_1 = f(x, u_0)$ and

$$J_\infty(f(x, u_0)) = \min_{\{u_t\}_{t=1}^{T}, T}\sum_{t=1}^{T-1}[c(x_t, u_t) + \Phi(x_T)],$$

it follows that: $J_\infty(x) = \min_u[c(x, u) + J_\infty(f(x, u))]$, establishing the result. ∎

Note that the reason $J_\infty(\cdot)$ satisfies the stationary Bellman equation is due to the free final time T in the problem formulation as otherwise the sum would be time varying. Then, owing to the fact that $J_\infty(\cdot)$ satisfies the Bellman equation, one has:

Corollary 3: The optimal cost function $J_\infty(\cdot)$ of the OCP Eq. (6) is a control Lyapunov function that renders the origin of system Eq. (1b) global asymptotically stable.

Proof: It is straightforward that for any $x \notin \Omega$, since $J_\infty(.)$ satisfies the Bellman equation outside of Ω, a trajectory following the optimal control has to hit Ω. Now, since Ω is within the domain of attraction $\mathcal{D}$ of the linear controller, $\bar{J}_\infty$ is a CLF for the system and the state $x_t \to 0$, as $t \to \infty$. Thus, the optimal cost function $J_\infty(.)$ is a CLF for the system. ∎

We note here that the CLF $J_\infty(.)$ as defined is continuous but not necessarily smooth, in particular, there is a continuous but non-smooth transition at the boundary of Ω for any finite M.

Furthermore, one can also establish the following convergence result. Given $M < \infty$, define:

$$J_\infty^M(x) = \min_{u_t, T}\left[\sum_{t=0}^{T-1} c(x_t, u_t) + \max\left(\bar{J}_\infty(x_T), M\right)\right], \tag{10}$$

where recall that $\Omega = \{x : \bar{J}_\infty(x) = x'P_\infty x \leq M\}$.

Theorem 2: Given the optimal control problem (Eq. (10)), the solution satisfies: $\lim_{M \to 0} J_\infty^M(x) = J_\infty^*(x)$ for all $x \in \mathcal{X}$.

Proof: The proof is essentially identical to that of Lemma 2 and Theorem 1. ∎

Finally, note that $J_\infty^M(\cdot) \neq J_\infty^*(\cdot)$ for any finite $M < \infty$. However, $J_\infty^M(\cdot)$ still satisfies Bellman's equation, and this is a non-trivial case where satisfying the BE is not sufficient for optimality.

D. Discussion

In the following, we discuss some of the salient points of the proposed regularization approach.

Computational efficiency: The approach proposed converges to the optimal cost function as the time horizon T becomes large, or equivalently, the size of the terminal set characterized by M becomes small. A valid question is how is this different from posing a problem without any terminal cost function and letting the time become large, where in the limit, the optimal cost function of the finite horizon problem converges to the optimal infinite horizon cost? To see the advantage of the proposed approach, let the optimal transfer time be $T_M^*(x)$ from a state x into the set $\Omega_M = \{\bar{J}_\infty \leq M\}$. Further, denote the effective regulation time required within the set Ω_M as $T_M^R(x)$. Then, it is easy to see that $T_M^*(x) << T_M^R(x)$, for any finite M. Therefore, the optimization problem for our approach has much shorter horizon when compared to a finite horizon problem without any terminal cost. This aspect is clearly borne out by our experiments in the next Section where the effective horizon

of the problem we need to solve is a small fraction of that required for the finite horizon problem without any terminal cost.

Free final time and global asymptotic stability: Another critical aspect of our approach is the free final time of the optimization problem from a given state $x \in \mathcal{X}$. In fact, we have already noted that the free final time is the reason that the approximate cost function $J_\infty^M(.)$ satisfies the stationary Bellman equation, as opposed to a time varying Bellman equation, thereby establishing global asymptotic stability. If we fixed the time horizon, then we would be back to the situation in MPC with the stabilizing ingredient of a terminal cost/ set constraint, and could only conclude asymptotic stability locally as in [21]. Physically, this makes sense since the optimal transfer time from different points in the state space will be different, and thus, the finite horizon in the optimal control problem needs to be variable. The free final time models precisely this aspect of the problem leading to the GAS closed loop system. The free final time problem can be solved efficiently by taking a "large enough" horizon rather than sweeping through time. Due to the argument above, "large enough" in our case is much smaller than the "large enough" time required in the problem without a terminal constraint.

Optimal control solution: Our approach requires the solution of an optimal control problem, and in general, satisfying the Bellman equation requires that we find the global minimum of the nonlinear optimal control problem. In our recent work [29], we have shown that under the conditions of affine in control dynamics, and quadratic in control cost (which can be relaxed to a convex in control cost), satisfying the minimum principle [30] is also sufficient to assure us of a global minimum. Albeit the problem is nominally non-convex, using the Method of Characteristics, we may show that the trajectories satisfying the minimum principle are guaranteed to be unique. Further, we have shown that the iterated LQR (ILQR) algorithm can be guaranteed to satisfy the minimum principle under relatively mild conditions, thereby assuring us that we can find the global minimum [20]. Hence, we use the ILQR approach to solve the optimal control problem in our experiments in Section IV.

IV. EMPIRICAL RESULTS

The proposed theory is tested with simulations on two non-linear systems: cart-pole and pendulum. Due to the paucity of space, the pendulum experiment and some additional simulations are shown in the extended version [28]. The simulations were done on MATLAB, using ode45 to simulate the nonlinear dynamics of the two systems. The list of experiments performed is shown in Table I. The simulations are designed to take the system upright from different initial conditions and balance the pole at the top. The simulations are run for a total of 150 time-steps with a time discretization of $0.1s$. We take the cost function to be quadratic: $c(x,u) = x'Qx + u'Ru$.

Our objective in these experiments is to show that the cost in our formulation of the FH-OCP converges to the infinite

Exp.#	System	Initial state	Terminal state
1.	Cart-pole $(x, \theta, \dot{x}, \dot{\theta})$	$(0, 0, 0, 0)$	$(0, \pi, 0, 0)$
2.	Cart-pole $(x, \theta, \dot{x}, \dot{\theta})$	$(0, 3\pi/4, 0, 0)$	$(0, \pi, 0, 0)$

TABLE I: Experiments performed

(a) Total cost (Exp.1)

(d) Total cost (Exp.2)

(b) Terminal regulation cost (Exp.1)

(e) Terminal regulation cost (Exp.2)

(c) Error at the end of finite horizon transfer (Exp.1)

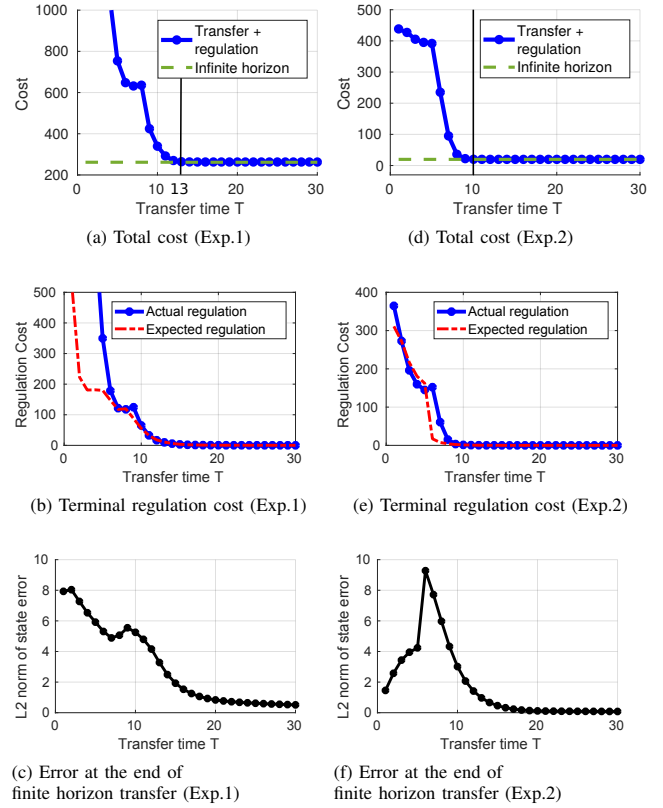(f) Error at the end of finite horizon transfer (Exp.2)

Fig. 2: Cart-pole swing-up. Note plots labeled (a) - (c) are Exp.1 and (d) - (f) are Exp.2 in Table I. The results in plots (a) and (d) show that the finite horizon transfer + regulation costs converge to the infinite horizon cost after some time T^* (marked with a black vertical line). One can also note that the value of T^* in plots (a) and (d) are different since the system starts from different initial conditions.

horizon optimal cost: $J_\infty^T(x) = \min_{u_t} \sum_{t=0}^{T-1} c(x_t, u_t) + \bar{J}_\infty(x_T)$, subject to the nonlinear dynamics, converges to the the optimal cost $J_\infty^*(x)$ as T is increased, i.e. For $T > T^*$, $|J_\infty^T(x) - J_\infty^*| < \varepsilon$. The experiments will also show that the time T^* also depends on the initial condition of the system.

To find $\bar{J}_\infty(x)$, we linearize the system around the origin, and since the cost is quadratic, we can calculate P_∞ at the origin using the stationary Riccati equation. Hence, $\bar{J}_\infty(x) = x'P_\infty x$. Using $\bar{J}_\infty(x)$ as the terminal cost for the FH-OCP, we solve it for different values of T using the iterative Linear Quadratic Regulator (iLQR) algorithm for optimization [31]. It has been shown that one can converge to the unique global optimum for smooth nonlinear problems with control affine dynamics and quadratic control cost by satisfying the necessary conditions for optimal control [29], which iLQR does. So, the problem of local optimum is avoided in our experiments. After the finite horizon or 'transfer time' T, we switch the control to the LQR regulator designed at the origin till $t = 150$. Since the true infinite horizon problem cannot be solved, we solve the FH-OCP for $T = 150$ without

any terminal cost, as a surrogate for the infinite horizon cost which is labeled as "Infinite horizon" in the plots. We note that solving for smaller horizons does not stabilize the systems to the upright position.

In our plots, the cost mentioned as "Transfer + regulation" is the total cost of the finite horizon controller and the regulator. In all four experiments, we will observe that the total cost converges after some time $T = T^*$. It can also be seen that the converged cost matches the infinite horizon cost. The experiments also show that the value T^* is dependent on the initial conditions chosen. The expected regulation cost is calculated using $x'_T P_\infty x_T$, where x_T is the terminal state of the FH-OCP. The actual regulation is calculated from the trajectory of the terminal regulator. The mismatch in expected and actual regulation cost implies that the linear terminal regulator is not optimal for the state x_T.

In practice, one would typically solve the infinite horizon OCP by formulating it as a finite horizon OCP with a very large T, which would be computationally very expensive. In this case, it is $T = 150$. But by regularizing the finite horizon OCP using the terminal cost $\bar{J}_\infty(\cdot)$ as defined in section III, one can solve it for a much smaller horizon T^*. Note that in our experiments, $T^* < 15 << 150 = T$ for both the cases considered, and we achieve near-optimality as shown in our results. Solving for a smaller horizon is computationally much cheaper, and also opens up avenues where one has to replan frequently in a stochastic setting. Moreover, since the cost-to-go from our formulation is a control Lyapunov function, it guarantees global asymptotic stability for the origin.

V. CONCLUSION

In this paper, we have developed a tractable approach to the approximate solution of nonlinear infinite horizon optimal control problems that is globally asymptotically stabilizing and converges to the true optimal solution in the limit of a vanishing terminal set. Empirical results show that the practical convergence occurs in a very short time compared to the effective horizon required for a solution of the infinite horizon without the free final time and terminal set regularization employed by the approach. Future work will involve the incorporation of state and control constraints and the testing of the approach on a suite of nonlinear problems with varying degrees of complexity which will require a data-based generalization leveraging our prior work [20]. We shall also consider the extension of the approach to the problem of optimal nonlinear output feedback control along with a suitable data-based generalization.

REFERENCES

[1] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. 2nd. Vol. I. Athena Scientific, 2000.
[2] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.
[3] D. S. Bernstein. "Nonquadratic cost and nonlinear feedback control". *International Journal of Robust and Nonlinear Control* 3 (1993).
[4] C.-J. Wan and D. S. Bernstein. "A family of optimal nonlinear feedback controllers that globally stabilize angular velocity". *IEEE Conference on Decision and Control*. 1992.
[5] C.-J. Wan and D. S. Bernstein. "Nonlinear feedback control with global stabilization". *Dynamics and Control* 5 (1995).
[6] J. Si, A. G. Barto, W. B. Powell, and D. Wunsch. *Handbook of learning and approximate dynamic programming*. Vol. 2. John Wiley & Sons, 2004.
[7] F. L. Lewis and D. Liu. *Reinforcement learning and approximate dynamic programming for feedback control*. Vol. 17. John Wiley & Sons, 2013.
[8] D. P. Bertsekas and D. A. Castanon. "Adaptive aggregation methods for infinite horizon dynamic programming". *IEEE Transactions on Automatic Control* 34 (1989).
[9] W. B. Haskell, R. Jain, and D. Kalathil. "Empirical dynamic programming". *Mathematics of Operations Research* 41 (2016).
[10] A. Geramifard, T. J. Walsh, S. Tellex, G. Chowdhary, N. Roy, and J. P. How. "A tutorial on linear function approximators for dynamic programming and reinforcement learning". *Foundations and Trends® in Machine Learning* 6 (2013).
[11] R. Munos and C. Szepesvári. "Finite-time bounds for fitted value iteration". *Journal of Machine Learning Research* 9 (2008).
[12] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
[13] C. J. Watkins and P. Dayan. "Q-learning". *Machine learning* 8 (1992).
[14] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour. "Policy gradient methods for reinforcement learning with function approximation". *Advances in neural information processing systems*. 2000.
[15] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, and M. Lanctot. "Mastering the game of Go with deep neural networks and tree search". *Nature* 529 (2016).
[16] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. "Trust region policy optimization". *International Conference on Machine Learning*. 2015.
[17] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor". *International Conference on Machine Learning*. 2018.
[18] S. Fujimoto, H. van Hoof, and D. Meger. "Addressing Function Approximation Error in Actor-Critic Methods". *ArXiv* abs/1802.09477 (2018).
[19] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. "Deep reinforcement learning that matters". *AAAI Conference on Artificial Intelligence*. 2018.
[20] R. Wang, K. S. Parunandi, A. Sharma, R. Goyal, and S. Chakravorty. *On the Search for Feedback in Reinforcement Learning*. 2022. arXiv: 2002.09478.
[21] D. Q. Mayne. "Model predictive control: Recent developments and future promise". *Automatica* 50 (2014).
[22] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert. "Constrained model predictive control: Stability and optimality". *Automatica* 36 (2000).
[23] D. Limon, T. Alamo, and E. Camacho. "Enlarging the domain of attraction of MPC controllers". *Automatica* 41 (2005).
[24] P. Falugi and D. Q. Mayne. "Model predictive control for tracking random references". *European Control Conference (ECC)*. 2013.
[25] L. Fagiano and A. R. Teel. "Generalized terminal state constraint for model predictive control". *Automatica* 49 (Sept. 2013).
[26] L. Grüne and J. Pannek. *Nonlinear model predictive control*. Springer, 2017.
[27] J. Köhler and F. Allgöwer. "Stability and performance in MPC using a finite-tail cost". *IFAC-PapersOnLine* 54 (2021).
[28] M. N. G. Mohamed, R. Goyal, and S. Chakravorty. *An Optimal Solution to Infinite Horizon Nonlinear Control Problems*. 2023. arXiv: 2304.00375.
[29] M. N. G. Mohamed, S. Chakravorty, R. Goyal, and R. Wang. *On the Optimal Feedback Law in Stochastic Optimal Nonlinear Control*. 2022. arXiv: 2004.01041.
[30] A. Bryson and H. Y.-C. *Applied Optimal Control: Optimization, Estimation and Control*. Washington: Hemisphere Pub. Corp., 1975.
[31] Y. Tassa, T. Erez, and E. Todorov. "Synthesis and stabilization of complex behaviors through online trajectory optimization". *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2012.