On the Computational Consequences of Cost Function Design in Nonlinear Optimal Control

Tyler Westenbroek, Anand Siththaranjan, Mohsin Sarwari, Claire J. Tomlin, Shankar S. Sastry

Abstract—**Optimal control is an essential tool for stabilizing complex nonlinear system. However, despite the extensive impacts of methods such as receding horizon control, dynamic programming and reinforcement learning, the design of cost functions for a particular system often remains a heuristic-driven process of trial and error. In this paper we seek to gain insights into how the choice of cost function interacts with the underlying structure of the control system and impacts the amount of computation required to obtain a stabilizing controller. We treat the cost design problem as a two-step process where the designer specifies outputs for the system that are to be penalized and then modulates the relative weighting of the inputs and the outputs in the cost. We then bound the length of the prediction horizon $T > 0$ that is required for receding horizon control methods to stabilize the system as a concrete way of characterizing the computational difficulty of stabilizing the system using the chosen cost function. Drawing on insights from the 'cheap control' literature, we investigate cases where the chosen outputs lead to minimumphase and non-minimumphase input-output dynamics. When the system is minimumphase, the prediction horizon needed to ensure stability can be made arbitrarily small by making the penalty on the control small enough. This indicates that choices of cost function which implicitly induce minimumphase behavior lead to an optimal control problem from which it is 'easy' to obtain a stabilizing controller. Using these insights, we investigate empirically how the choice of cost function affects the ability of modern reinforcement learning algorithms to learn a stabilizing controller. Taken together, the results in this paper indicate that cost functions which induce non-minimum phase behavior lead to inherent computational difficulties.**

I. Introduction

The stabilization of complex nonlinear systems is one of the most fundamental and important problems in control theory. Approaches based on optimal control [1], [2], [3] form an essential set of tools for solving the stabilization problem and have seen extensive real-world deployment [4], [5]. The primary appeal of optimal control is that it allows the user to implicitly encode potentially complex stabilizing controllers as the feedback solutions to certain infinite horizon optimal control problems which are relatively simple to specify.

However, optimal control is not without limitations. Obtaining an optimal infinite horizon controller requires solving the Hamilton Jacobi Bellman partial differential equation [6]. However, for general nonlinear problems it is rarely possible to solve the equation in closed form. This has lead to the development of dynamic programming methods [2], [6], [7], which construct a sequence of approximations to the optimal controller. However, due to the curse of dimensionality [7], the computational complexity of dynamic programming explodes as the dimension of the system increases. This in turn has lead to the development of methods which approximate the optimal control law through various means, including Receding Horizon Control [8] and Approximate Dynamic Programming [2] (which includes modern reinforcement learning methods [9]). In one way or another, each of these methods trade off the amount of computation used to solve the problem with the quality of the resulting control law. However, despite a steady increase in the availability of computational resources, reliably controlling many high-dimensional systems with optimal control remains infeasible.

Rather than developing new methods for obtaining approximately optimal control laws, in this paper we ask the following: to what extent can we alleviate the computational issues discussed above by designing 'good' cost functions for a particular system? For concreteness, we will study this question through the lens of receding horizon control and investigate how the chosen cost influences the length of prediction horizon needed to stabilize the system. In particular, the critical value of the prediction horizon will serve as our measurement of how computationally challenging it is to obtain a stabilizing control scheme using the chosen cost function. We note that previous work [2], [10] has demonstrated an important correspondence between the prediction horizon that RHC schemes need to stabilize the system and the number of iterations that dynamic programming-based methods need to obtain a stabilizing controller. Guided by these insights, we also empirically investigate how the chosen cost function impacts the ability of modern reinforcement learning algorithms to learn a stabilizing controller.

We will consider two degrees of freedom when addressing the cost-design problem. The first design choice is to choose a set of outputs for the system. The class of costs we consider will have a term which penalizes the $\mathcal{L}_2$ loss of the outputs and a second term which penalizes a weighted $\mathcal{L}_2$ loss of the system inputs. Thus, the second design choice is to pick the parameter which weights the input costs, which will be denoted by $\epsilon > 0$ throughout the paper. We note that cost functions of this variety are used extensively in practice [11], [12], however the cost design process is typically driven by trial and error. Here, we seek to understand how the chosen cost function interacts with the inherent geometry of the control system to affect the computational complexity of obtaining a stabilizing controller.

Our theoretical analysis draws on two distinct bodies of work. The first line of work comes from the receding horizon control literature [13], [14] where bounds on the optimal infinite horizon value function are used to upper bound the size of prediction horizon that RHC schemes need to stabilize the system. Roughly speaking, as the upper bound on the infinite horizon cost becomes smaller the prediction horizon needed to guarantee stability also decreases.

The second line of work we draw on is the so-called 'cheap control' literature [15], [16], [17], [18], which also studies the class of cost functions we consider in this paper. The term 'cheap control' comes from the fact that this body of work studies the asymptotic behavior of the infinite horizon value function as the weighting parameter ϵ is taken to zero. Even though these arguments are asymptotic in nature they still provide qualitative insights into how difficult it is to drive the chosen outputs to zero with a feedback controller which also stabilizes the internal dynamics. The primary result from this literature is a qualitative separation between choices of outputs which lead to minimumphase behavior versus choices of outputs which lead to non-mimimumphase behavior. In particular, for minumumphase systems the performance can be made arbitrarily better, however there exists fundamental performance limitations for non-minumphase systems.

We combine these two lines of work to demonstrate that when the chosen outputs induce minimumphase behavior, receding horizon control schemes can be guaranteed to stabilize the system with arbitrarily small prediction horizons by choosing ϵ to be sufficiently small. We take this qualitative result as an indication that users can design 'good' cost functions for a particular system by considering whether they implicitly induce minimumphase behavior. Indeed, as our empirical results indicate, this can have have major effect on the difficulty of obtaining a stabilizing controller.

II. Problem Formulation and Background

Throughout the paper we will consider control affine systems of the form

$$\begin{aligned} \dot{x} &= f(x) + g(x)u, \qquad x(0) = x_0 \\ y &= h(x) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $x_0 \in \mathbb{R}^n$ is the initial condition, $u \in \mathbb{R}^q$ is the input and $y \in \mathbb{R}^q$ is the output and n and q are positive integers. In line with the majority of the cheap control literature, we assume that the number of inputs and outputs are equal to simplify exposition. We will assume that the maps $f \colon \mathbb{R}^n \to \mathbb{R}$, $g \colon \mathbb{R}^n \to \mathbb{R}^{n \times q}$ and $h \colon \mathbb{R}^n \to \mathbb{R}^q$ are such that $f(0) = 0$ and $h(0) = 0$. As we discussed above, we view the construction of the output map h as a design choice which is made when constructing the cost function discussed below. We will let y_j and $h_j(x)$ denote the j-th entries of y and $h(x)$, respectively. For each $T \in \mathbb{R} \cup \{\infty\}$ we will let U_T denote the set of controls of the form $u \colon [0, T] \to \mathbb{R}^q$ which are measurable and essentially bounded.

A. Nonlinear Geometry: IO Linearization and Normal Forms

Next, for a given choice of output function h we discuss how to construct an associated feedback linearizing controller and introduce the 'normal form' associated to the input-output system (1). This representation is essential for developing a qualitative understanding of how the choice of output interacts with the underlying geometry of the control system and is used extensively in our analysis. Our introduction to these concepts will be brief, as they are covered in many standard references (e.g. [19, Chapter 9]).

The main idea behind feedback linearization is to differentiate each output until an input appears in one the derivatives, which yields an expression of the form

$$\begin{bmatrix} y_1^{(r_1)} & \dots & y_q^{(r_q)} \end{bmatrix}^T = b(x) + A(x)u, \tag{2}$$

where $y_j^{(k)}$ denotes the k-th time derivative of y_j, $b(x) \in \mathbb{R}^q$ is called the drift term and $A(x) \in \mathbb{R}^{q \times q}$ is the decoupling matrix, and the r_i are positive integers. If $A(x)$ is bounded away from singularity for each $x \in \mathbb{R}^n$ then the control law

$$u = A^{-1}(x)[b(x) + v] \tag{3}$$

yields the decoupled linear relationship

$$\begin{bmatrix} y_1^{(r_1)} & \dots & y_q^{r_q} \end{bmatrix}^T = \begin{bmatrix} v_1, \dots, v_q \end{bmatrix} \tag{4}$$

where v_J is the j-th entry of v. In this case we say that the system has a well-defined (vector) relative degree $(r_1, r_2, \dots, r_q)$ on D, and we denote $\bar{r} = r_1 + \dots + r_q$. Under this assumption we can construct a partial change of coordinates using the outputs and their derivatives, namely, for each $j \in \{1, \dots, q\}$ we define the coordinates

$$\xi_j^1 = h_j(x), \quad \xi_j^2 = L_f h_j(x), \quad \dots, \quad \xi_j^{r_j} = L_f^{r_j - 1} h_j(x), \tag{5}$$

and the Lie derivatives, defined successively by $L_f h_j(x) = \frac{d}{dx} h_j(x) \cdot f(x)$ and $L_f^{k+1} h_j(x) = \frac{d}{dx}(L_f h_j(x)) \dot{f}(x)$, are such that $y_j^{(k)} = L_f^k h_j(x)$. Letting $\xi \in \mathbb{R}^{|\bar{r}|}$ collect the ξ_j^i coordinates, we can construct and additional set of coordinates $\eta \in \mathbb{R}^{n - \bar{r}}$ so that $x \to (\xi, \eta)$ is a valid coordinate transformation and in the new coordinates the dynamic are of the form

$$\begin{aligned} \dot{\xi} &= F\xi + G\Big[b(\xi, \eta) + A(\xi, \eta)u\Big] \\ \dot{\eta} &= q(\xi, \eta) + P(\xi, \eta)u \\ y &= H\xi \end{aligned} \tag{6}$$

where (F, G) is controllable, (F, H) is observable, and we have permitted a slight abuse of notation when defining $b(\xi, \eta)$ and $A(\xi, \eta)$. The preceding representation of the dynamics is known as the *normal form* for the input output system (1).

Note that if $\xi = 0$ then the control $u = A^{-1}(0, \eta)b(0, \eta)$ will keep the outputs zeroed. This leads to the zero dynamics

$$\dot{\eta} = q(0, \eta) + P(0, \eta)A^{-1}(0, \eta)b(0, \eta), \tag{7}$$

which represents the remaining internal dynamics of the system. We say that the system is *minimumphase* if the zero dynamics are asymptotically stable and *exponentially minimumphase* if the zero dynamics are exponentially stable.

We will say that a system is *non-minimumphase* if it is not minimimpase. In the case where $\bar{r} = n$ there are no zero dynamics and we refer to the system as *full state linearizable*. In this case the system is trivially both minumumphase.

B. Strict Feedback Systems

In this section we introduce the special case of dynamics which we will study throughout the paper. A full justification for these assumptions will be given when discussing the cheap control problem below. We begin with the following assumption:

Assumption 1. *There exists $r \in \mathbb{N}$ such that the vector relative degree of the system* (1) *is* $(r, r, \ldots, r)$.

Under assumption 1, through the rest of the paper we will arrange the ξ coordinates in the normal form (6) as $\xi = (\xi_1, \ldots, \xi_r)$ where

$$\xi_1 = h(x), \quad \xi_2 = L_f h(x), \quad \ldots \quad, \xi_r = L_f^{r-1} h(x) \tag{8}$$

as the coordinates for outputs and their derivatives. Note that in these coordinates the F and G matrices in (6) are of the following form:

$$F = \begin{bmatrix} 0 & I & 0 & \ldots & 0 \\ 0 & 0 & I & \ldots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \ldots & I \end{bmatrix} \; G = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ I \end{bmatrix}, H = \begin{bmatrix} I & 0 & \ldots & 0 \end{bmatrix} \tag{9}$$

Next we restrict the form of the interconnection between the ξ and η subsystems. We say that the input output system (1) can be put into *strict feedback form* if the η coordinates can be chosen so that the normal form of the dynamics takes the following form:

$$\begin{aligned} \dot{\xi} &= F\xi + G[b(\xi, \eta) + A(\xi, \eta)u] \\ \dot{\eta} &= f_0(\eta) + g_0(\eta)\xi_1 \end{aligned} \tag{10}$$

where $f_0 \colon \mathbb{R}^{(n-\bar{r})} \to \mathbb{R}^{(n-\bar{r})}$ and $g_0 \colon \mathbb{R}^{(n-\bar{r})} \to \mathbb{R}^{(n-\bar{r}) \times q}$.

Assumption 2. *The input-output system* (1) *can be put into the strict feedback form* (10).

In particular this assumption forbids the input and the derivatives of the outputs to appear directly in the dynamics of the zeros. We note that in the special case of linear dynamics the system can always be put into strict feedback form when Assumption 1 is satisfied (see e.g. [18]).

C. Cost Function Design and the Cheap Control Problem

Given that that the choice of output in (1) satisfies Assumptions 1, for each parameter $\epsilon > 0$ we will study the infinite horizon optimal control problem

$$\inf_{u(\cdot) \in U_\infty} J_\infty^\epsilon(u(\cdot); x_0) := \int_0^\infty \|h(x(t))\|_2^2 + \epsilon^{2r}\|u(t)\|_2^2, \tag{11}$$

and also for each prediction horizon $T > 0$ the following finite-horizon approximation:

$$\inf_{u(\cdot) \in U_\infty} J_T^\epsilon(u(\cdot); x_0) := \int_0^T \|h(x(t))\|_2^2 + \epsilon^{2r}\|u(t)\|_2^2. \tag{12}$$

The parameter ϵ controls the relative importance of the inputs and outputs in the cost, and the reason for raising it to the power of $2r$ will be made clear momentarily. To ease exposition we will assume for that each $\epsilon > 0$ and $x \in \mathbb{R}^n$ there is are unique controls $u_\infty(\cdot; x_0) \in U_\infty$ and $U_T(\cdot; x_0) \in U_T$ which achieve the minimum values on the right hand side of (11) and (12).

To these problems we associate the following value functions:

$$V_\infty^\epsilon(x) = \inf_{u \in U_\infty} J_\infty^\epsilon(u(\cdot); x).$$
$$V_T^\epsilon(x) = \inf_{u \in U_T} J_T^\epsilon(u(\cdot); x).$$

We will assume that for each $\epsilon > 0$ the value function is positive definite and continuously differentiable. Under this condition it is well-known (see e.g. [6, Chapter 3.2]) that the value function satisfies the Hamilton-Jacobi-Bellman (HJB) partial differential equation, nameley,

$$\nabla V^\epsilon(x) f(x) + \|h(x)\|_2^2 - \frac{1}{\epsilon^{2r}} \nabla V^\epsilon(x) g(x) g(x)^T \nabla V^\epsilon(x) = 0 \tag{13}$$

and that an optimal control law which asymptotically stabilizes the system is given by

$$u^\epsilon(x) = -\frac{1}{\epsilon^{2r}} g(x)^T \nabla V(x). \tag{14}$$

The focus of the cheap control literature has been to characterize the optimal value function $\hat{V}_\infty^\epsilon$ and structure of the corresponding optimal controller u^ϵ for small values of ϵ. In particular, the limiting value $\lim_{\epsilon \to 0} V_\infty^\epsilon(x)$ provides qualitative insight into how difficult it is to drive the chosen outputs to zero from the state $x \in \mathbb{R}^n$ while also stabilizing the internal dynamics. These results are primarily obtained via asymptotic analysis techniques. In particular, singular perturbation techniques and and asymptotic series expansions are used to characterize the behavior of V^ϵ and u^ϵ around $\epsilon = 0$. While these method necessarily lead to coarse bounds, as with all asymptotic methods, it is understood that the qualitative insights these techniques enable also apply to the cases where $\epsilon > 0$ takes on moderate values (i.e. $\epsilon \approx 1$).

The essential result from the literature is a qualitative separation between the performance limitations of minimumphase and non-minimumphase systems. While the majority of the literature has focused on the case where the dynamics are linear [17], [18], [20], the seminal line of work in in [15] and [21] extends these results to nonlinear strict-feedback systems of the form (10). An integral part of the analysis for strict feedback systems is the 'minimum energy problem' which is formulated using the normal form (10):

$$\hat{V}_0(\eta_0) = \inf_{\xi_1(\cdot)} J_0(\xi_1(\cdot); \eta_0) = \int_0^\infty \|\xi_1(t)\|_2^2 \tag{15}$$

where $\dot{\eta} = q_0(\eta) + p_0(\eta)\xi_1$ where the output $\xi_1(\cdot)$ is viewed as an input to the zero subsystem and the infimum in (15) is understood to be over $\xi_1(\cdot)$ which drive $\eta(t) \to 0$ asymptotically. Thus, $\hat{V}_0(\eta)$ can be interpreted as the minimum 'energy' of the outputs (in an $\mathcal{L}_2$ sense) that must be accrued by a feedback controller which stabilizes the internal dynamics.

Crucially, one may observe that if the system is minimumphase then $\hat{V}_0(\cdot) \equiv 0$ since no 'energy' must be expended to stabilize the zeros. As is shown in [21], in this case as $\epsilon \to 0$ the optimal control law for the overall problem (14) is a high-gain feedback controller which directly drives the outputs and their derivatives to zero. On the other hand, when the system is non-minimumphase then we must have $\hat{V}_0(\cdot) \not\equiv 0$ and the outputs must be 'steered' to stabilize the zero. In both cases the performance limitation for the system is given by $\lim_{\epsilon \to 0} \hat{V}^{\epsilon}(\xi, \eta) = \hat{V}_0(\eta)$, where $\hat{V}^{\epsilon}_{\infty}(\xi, \eta)$ is the representation of the value function in the normal coordinates. Thus, for systems which can be put into strict feedback form, there is a qualitative separation between the performance limitations of minimumphase and non-minimumphase systems.

However, for more general nonlinear systems of the form (6) it is not clear that even if the system is asymptotically or even exponentially minimumphase that we will have $\lim_{\epsilon \to 0} V^{\epsilon}(x) = 0$ in the case where $r > 1$ due to the well documented *peaking phenomena* (see. [22] for a comprehensive discussion). Essentially, this property dictates that as we increase the gain of a feedback controller which drives $y(t) \to 0$ both the derivatives of the outputs and the input will 'peak' to larger and larger transient magnitudes, potentially perturbing the zeros to a greater and greater extent. Thus, as examples demonstrate [22], without additional structural assumptions it may be impossible to rule out that a high-gain controller which rapidly drives $y \to 0$ (which is a necessity for $\lim_{\epsilon \to 0} V^{\epsilon}(x)$) does not cause the zeros to escape to infinity.

D. Fast-Slow Representations

As mentioned above, singular perturbation techniques play a crucial role in obtaining the aforementioned results and will play an essential role in our analysis. In particular, for each $\epsilon > 0$ consider the rescaled variables $\xi \to \tilde{\xi}$ and $u \to \tilde{u}$ defined for each $\epsilon > 0$ by

$$\tilde{\xi} = S(\epsilon)\xi \qquad \text{and} \quad \tilde{u} = \epsilon^r u, \tag{16}$$

where

$$S(\epsilon) = \mathrm{diag}(1, \epsilon, ..., \epsilon^{r-1}). \tag{17}$$

In the new coordinates strict feedback systems take the form

$$\begin{aligned} \epsilon\dot{\tilde{\xi}} &= F\tilde{\xi} + G\big[\epsilon^r \tilde{b}(\tilde{\xi}, \eta) + \tilde{A}(\tilde{\xi}, \eta)\tilde{u}\big] \\ \dot{\eta} &= \tilde{f}_0(\eta) + g_0(\eta)\tilde{\xi}_1, \end{aligned} \tag{18}$$

where $\tilde{b}(\tilde{\xi}, \eta) = b(S(\epsilon)^{-1}\tilde{\xi}, \eta)$ and $\tilde{A}(\tilde{\xi}, \eta) = A(S(\epsilon)^{-1}, \eta)$ and we have suppressed the dependence of these terms on ϵ. In the rescaled coordinates the infinite and finite horizon problems take the forms:

$$\inf_{\tilde{u}(\cdot) \in \mathcal{U}_{\infty}} \tilde{J}_{\infty}(\tilde{u}(\cdot); \tilde{\xi}_0, \eta_0) = \int_0^{\infty} \|\tilde{\xi}_1(t)\|_2^2 + \|\tilde{u}(t)\|_2^2 dt \tag{19}$$

$$\inf_{\tilde{u}(\cdot) \in U_T} \tilde{J}_T(\tilde{u}(\cdot); \tilde{\xi}_0, \eta_0) = \int_0^{T} \|\tilde{\xi}_1(t)\|_2^2 + \|\tilde{u}(t)\|_2^2 dt \tag{20}$$

The corresponding value functions are given by:

$$\tilde{V}^{\epsilon}_{\infty}(\tilde{\xi}_0, \eta_0) = \inf_{\tilde{u} \in U_{\infty}} J^{\epsilon}_{\infty}(\tilde{u}(\cdot); \tilde{\xi}_0, \eta_0)$$

$$\tilde{V}^{\epsilon}_{T}(\tilde{\xi}_0, \eta_0) = \inf_{u \in U_T} J^{\epsilon}_{T}(u(\cdot); x),$$

and $\tilde{\xi}_{\infty}(\cdot; \tilde{\xi}_0, \tilde{\eta}) \in U_{\infty}$ and $\tilde{\xi}_{T}(\cdot; \tilde{\xi}_0, \tilde{\eta}) \in U_{\infty}$ denote the controls which achieve the preceding minimizations.

The form of the rescaled cost function and and dynamics clearly evokes the intuition that we should expect a fast transient respons from the outputs for small values of epsilon. Together, Assumption 1 and the ϵ^{2r} scaling of the control in (11) enable this convenient fast-slow representation of the dynamics. In the case where there are outputs with different relative degrees the cheap control problem induces a fast-slow problem with multiple fast time scales (see e.g. [17]). This significantly complicates the statement and analysis of theoretical results without ultimately affecting the key conceptual insights that can be gleaned from the cheap control problem. Thus, we work with Assumption 1 throughout the paper primarily to ease exposition.

E. Stability of Receding Horizon Control

Before proceeding to our main results we briefly review receding horizon control and introduce the specific stability result we will employ in our analysis. For this section we turn way from the costs (11)-(12) and study more general problems:

$$\inf_{u(\cdot) \in U_{\infty}} J_{\infty}(u(\cdot); x_0) = \int_0^{\infty} \ell(x(t), u(t)) dt \tag{21}$$

$$\inf_{u(\cdot)} J_T(u(\cdot); x_0) = \int_0^{T} \ell(x(t), u(t)) dt, \tag{22}$$

where $\ell \colon \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$ is a non-negative function with $\ell(0, 0) = 0$. We associate to this problem the values

$$V_{\infty}(x) = \inf_{u(\cdot) \in U_{\infty}} J_{\infty}(u(\cdot); x_0).$$

$$V_T(x) = \inf_{u(\cdot) \in U_T} J_T(u(\cdot); x_0),$$

and again let $u_{\infty}(\cdot; x_0) \in U_{\infty}$ and $u_{\infty}(\cdot; x_0) \in U_T$ denote the optimal controls for the above problems, and we will also let $x_T(\cdot; x_0)$ denote the solution to (1) under $u_T(\cdot; x_0)$ with $x_T(0; x_0) = x_0$. We assume that V_{∞} positive definite and continuously differentiable to that it admits and optimal feedback controller u_{∞} which asymptotically stabilizes the system.

For each prediction horizon $T > 0$ and replanning interval $\Delta t > 0$ RHC schemes comprise sampled-data feedback law of the form $u(t) = u_T(t - k\Delta t, x(k\Delta t))$ for each

$t \in [k\Delta t, (k+1)\Delta t]$ and $k \in \mathbb{N}$, where $(x(\cdot), u(\cdot))$ are the trajectory and input for the physical system (1). The interpretation of this control scheme is that at each time $k\Delta t$ the open loop control $u_T(\cdot, x(k\Delta t))$ is obtained by solving $J_T(\cdot, x(k\Delta x))$ and the parameter Δt is understood to be the time between replanning instances. Note that the state at successive planning instances is given by $x((k+1)\Delta t) = x_T(\Delta t; x(k\Delta t))$.

At their core, stability results from the literature [23], [1] are founded on the notion that as T increase the RHC scheme more closely approximates the infinite horizon continuous-time feedback controller u_∞. However, increasing T comes at the cost of additional computational complexity when solving (22).Throughout the paper we will employ the stability result described below which bounds the length of the time horizon needed to stabilize the system using, in part, an bound on the growth of the infinite horizon value function V_∞. This result is essentially a sampled-data adaptation of the main result from [1] (with several specializations for our setting). In the following, the map $\sigma(\cdot)\colon \mathbb{R}^n \to \mathbb{R}$ is a positive definite function which is used to measure the distance of the state to the origin. The stability result employs the following two assumptions:

Assumption 3. *There exists $\bar{\alpha}_V > 0$ such that:*

$$V_\infty(x) \leq \bar{\alpha}\sigma(x) \quad \forall x \in \mathbb{R}^n. \tag{23}$$

Assumption 4. *There exists a continuously differentiable function $W\colon \mathbb{R}^n \to \mathbb{R}^n$ and $\bar{\alpha}_W, \underline{\alpha}_W > 0$ and $K_W > 0$ such that for each $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$:*

$$\underline{\alpha}_W \sigma(x) \leq W(x) \leq \bar{\alpha}_W \sigma(x) \tag{24}$$

$$\frac{d}{dx} W(x)[f(x) + g(x)u] \leq -K_W \sigma(x) + \ell(x, u). \tag{25}$$

Note that because the running cost ℓ is non-negative we must have $J_\infty(x) > J_T(x) > 0$ for each $T > 0$. Thus, the condition in Assumption 3 also bounds the growth of $V_T(x)$. The existence of the map W in Assumption 4 ensures that the state measure $\sigma(\cdot)$ is detectable with respect to the loss function ℓ. In the statement of the following result we will let $x_T(\cdot)$

Theorem 1. *Let Assumptions 3 and 4 hold. Then for each $T \geq \Delta t \geq 0$ we have*

$$Y_T(x_T(\Delta t; x)) \leq \left(M(\Delta t) + \bar{\alpha} \frac{\bar{\alpha} + \bar{\alpha}_W}{K \underline{\alpha}_W T} \right) Y_T(x) \tag{26}$$

where $Y_T = W + V_T$ and $M(\Delta t) = \exp(-\frac{\underline{\alpha}_W \Delta t}{\bar{\alpha} + \bar{\alpha}_W})] < 1$ and $x_T(\cdot; x)$ is defined as above.

Theorem 1 suggests that for a fixed $\Delta t > 0$ we use Y_T for $T > 0$ sufficiently large to certify the asymptotic stability of the corresponding RHC scheme. In particular, the bound in (26) indicates that Y_T decays between replanning instances if we pick $T > \frac{\bar{\alpha}(\bar{\alpha} + \bar{\alpha}_W)}{K \underline{\alpha}_W (1 - M(\Delta t))}$. Note in particular how this bound depends on the performance of the infinite horizon cost as prescribed by Assumption 3. As $\bar{\alpha}$ decreases we can ensure stability of the closed-loop system by using RHC schemes with smaller and smaller prediction horizons.

III. Overcoming Computational Limitations in Nonlinear Optimal Control

We are now ready to present our main results which demonstrate that choices of outputs which lead to minimum-phase dynamics allow us to ensure the stability of RHC schemes with arbitrarily small prediction horizons by choosing ϵ to be sufficiently small. Theorem 2 covers the special case where the input output system is full-state linearizable, while Theorem 3 covers the more general case where the zero dynamics are exponentially minimumphase. For each $\epsilon > 0$ and $T \geq \Delta t > 0$ we will let $u^\epsilon_{T,\Delta t}(\cdot; x) = u^\epsilon_T(\cdot; x)|_{0,\Delta t}$ denote the corresponding sampled-data receding horizon controller, and for each initial condition $x_0 \in \mathbb{R}^n$ we will let $x_{T,\Delta t}(\cdot; x_0)$ denote the corresponding trajectory of the closed loop system starting from $x_{T,\Delta t}(\cdot; x_0) = x_0$.

Our theoretical results will require the following growth conditions on the dynamics. In the following assumption we use $\sigma_{min}(\cdot)$ to denote the minimum singular value of a matrix.

Assumption 5. *There exists constants $L, C, \gamma > 0$ such that the following conditions hold for each $(\xi, \eta) \in \mathbb{R}^n$:*

$$\|b(\xi, \eta)\|_2 \leq L(\|\xi\|_2 + \|\eta\|_2)$$
$$\|A(\xi, \eta)\|_2, \|g_0(\eta)\|_2 \leq C$$
$$\lambda_{min}(A(\xi, \eta)) \geq \gamma.$$

A. Full State Linearizable Systems

We state our result for systems which are full-state linearizable:

Theorem 2. *Let Assumptions 1 and 5 hold. Then for every $T \geq \Delta t > 0$ there exists $\epsilon_0 > 0$ such that for each $\epsilon \in (0, \epsilon_0]$ the receding horizon controller $u_{T,\Delta t}(\cdot, x)$ renders the closed-loop system globally exponentially stable.*

The proof of the Theorem can be found in the appendix, and consists of two main parts. First, using the $(\tilde{\xi}, \eta)$ coordinates, for each $\epsilon > 0$ we apply a sub-optimal feedback linearizing controller to drive the system to zero and bound the resulting cost. By doing so we are able to obtain an upper bound on $V^\epsilon(x)$ of the form $V_\alpha(\tilde{\xi}) \leq C\epsilon(\|\xi\|_2^2 + \|\eta\|_2^2)$. Then, inspired by the proof of Corollary 4 from [13], we construct a W which satisfies Assumption 2. Then the desired result follows from an application of Theorem 1 by taking ϵ to be sufficiently small so that the bound on V^ϵ is small enough to ensure stability. These two steps are encapsulated as a special case of Lemmas 1 and 2 which actually demonstrate these steps for the more general case where the system is exponentially minumumphase. The needed result for the full-state linearizable case is obtained by ignoring the presence of the zeros in these constructions.

B. Exponentially Minimumphase Systems

Next we state our result for systems which are exponentially minimumphase but not necessarily fullstate linearizable.

Theorem 3. *Let Assumptions 1, 2 and 5 hold. Further assume that the zero dynamics* $\dot{\eta} = f_0(\eta)$ *are globally exponentially stable. Then for each* $R > \delta > 0$ *and* $T \geq \Delta t > 0$ *there exists* $\epsilon_0 > 0$ *such that for each* $\epsilon \in (0, \epsilon_0]$ *and each* $x_0 \in \mathbb{R}^n$ *such that* $\|x_0\|_2 < r$ *we have that*

$$\|x_{T,\Delta t}(t, x_0)\|_2 \leq \beta(\|x_0\|, t) + \delta. \tag{27}$$

where β *is of class* $\mathcal{KL}$*.*

The proof can again be found in the appendix. The proof again uses the performance bound in Lemma 1 and candidate W function in Lemma 2. However, due to additional coupling terms which arise between the dynamics of the output subsystem and the zeros, in this case we are only able to guarantee that we can drive the system to a ball around the origin of prescribed radius (as specified by $\delta > 0$) starting in any desired operating region (as specified by $R > 0$).

C. Non-minimumphase Systems

Due to the performance limitations inherent to non-minumphase systems there is a lower bound to how small we can make the prediction horizon $T > 0$ while ensuring stability of the system using the preceding analysis. In particular, since the value function will be lower bounded by the solution to the minimum-energy problem for the zero stabilization problem (15), we can only decrease $\bar{\alpha}$ so much when trying to improve the bound in (26) (in terms of the required $T > 0$ needed for stability). This provides an indication that choices of outputs which lead to non-minimumphase dynamics will lead to optimal control problems from which it is more computationally challenging to obtain a stabilizing controller. This is corroborated by the experiments below, and it is an important matter for future work to characterize if there are *lower bounds* on how small we can make $T > 0$ and stabilize the system. This would provide a sharper characterization of the fundamental hardness in controlling non-minumumphase system using optimal control.

D. Input Constraints

It has also been noted [24] that, even if the system (1) is minimumphase, constraints on the inputs of the form $\|u\| < k$ will also lead to performance limitations for the system. In particular, these constraints will limit how quickly we can drive the outputs to zero and will again lead to lower-bounds on the infinite-horizon cost. Thus, in light of the preceding discussion, we should expect that it is computationally more difficult to obtain stabilizing controllers for systems with tight constraints on the inputs. This matches practical experiences and is demonstrated empirically in our experiments.

IV. Numerical Experiments with Reinforcement Learning

Guided by our theoretical analysis and the connections between RHC and dynamic programming-based methods that have been documented in the literature [25], we now investigate how the choice of cost function affects the ability of modern reinforcement learning algorithms to learn a stabilizing controller. In particular, we conjecture that choices of outputs which lead to non-minimumphase dynamics will make it more difficult for these algorithms to learn stabilizing controllers. We also investigate how constraints on the inputs lead to similar difficulties. In particular, for each of the following experiments we use the Soft Actor-Critic algorithm [3]. In each of the plots, different colors correspond to training with different values of ϵ (except in Figure 6 where they denote different magnitudes of input bounds). Each of the reward plots indicate the average reward per epoch at a given number of training samples. The plots with various states on the y-axis depict trajectories from the trained controllers for different values of ϵ. Finally, the 'action' plots depict the evolution of the input over time.

Inverted Pendulum: We first consider the dynamics of an inverted pendulum. The states are $(x_1, x_2) = (\theta, \dot{\theta})$, where θ is the angle of the arm from vertical. Units have been normalized so that the model is of the form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ \sin(x_2) + u \end{bmatrix},$$

Note that the system is fully-state linearizable with the output $y = x_1$.

Flexible Link Manipulator: Next we consider a model of a flexible link manipulator. The state is $(x_1, x_2, x_3, x_4) = (\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2)$, where θ_1 is the angle of the arm from vertical and θ_2 is the internal angle of the motor. The dynamics are

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ sin(x_1) + K(x_3 - x_1) - \beta_1 x_2 \\ x_4 \\ K(x_1 - x_3) - \beta_2 x_4 + u, \end{bmatrix}$$

where $K > 1$ is a spring coefficient used to model the flexibility of the joint and $\beta_1, \beta_2 \geq 0$ are friction coefficients. One may observe that if the output $y = x_1$ is chosen then the system is full state lineraizable. However when the output $y = x_3$ is chosen the system has a relative degree of two and the zeros are also two dimensional. In this case a Jacobian linearization at the origin reveals that when the model is friction-less ($\beta_1 = \beta_2 = 0$) the system is non-minimumphase but when damping is present ($\beta_1, \beta_2 > 0$) the system is minimumphase.

A. Flexible Manipulator Without Friction

For our first experiment we consider the flexible link manipulator without friction. We run experiments for $y = x_1$ in Figure 1 and $y = x_3$ in Figure 2. Recall that the system is full state linearizable for $y = x_1$ and non-minimumphase

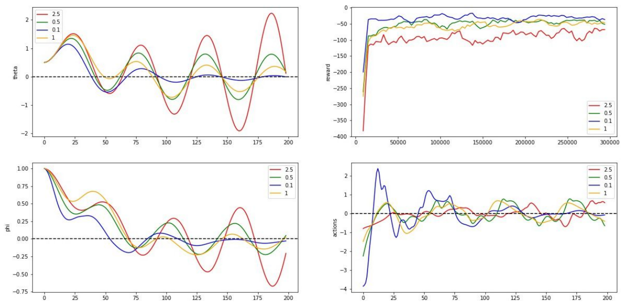

Fig. 1. Flexible link manipulator without friction when $y = x_1$.

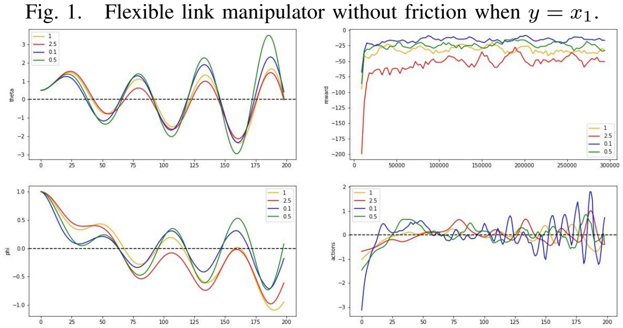

Fig. 2. Flexible link manipulator without friction when $y = x_3$.

for $y = x_3$. As the figures clearly show, the reinforcement learning algorithm struggles to learn a stabilizing controller when $y = x_3$. However when $y = x_1$ the algorithm is able to rapidly learn a stabilizing controller for small values of ϵ but again struggles when the parameter is large. Thus, we conclude that the flat output $y = x_1$ is a 'better' choice of output.

B. Flexible Link Manipulator With Friction

Next we consider the flexible link manipulator with friction with $y = x_1$ in Figure 3 and $y = x_3$ in Figure 4. In both cases the algorithm is able reliably learn a stabilizing controller as the dynamics are minimumphase. We again observe that the convergence of the learning algorithm is generally faster for small values of ϵ. When compared to the previous experiments, we observe that the added passivity from the friction terms generally makes it easier to learn stabilizing controllers for the system, reguardless of the output that is chosen since both choices now yield minimumphase behavior.

C. Inverted Pendulum With and Without Input Constraints

Next we consider the inverted pendulum without input constraints in Figure 5 and with input constraints in Figure 6 (where the different colors correspond to different input bounds of the form $|u| \leq k$). In both cases we choose $y_1 = x_1$. For the unconstrained case we see that as ϵ decreases the algorithm is able to rapidly learn a stabilizing controller. For the constrained case, where $\epsilon = 0.1$, we see that as the input constraints are decreased the algorithm takes longer to learn a stabilizing controller. Interestingly, we observed that as we increased ϵ and decreased the bounds on the inputs, the learned controllers display 'swing-up' behavior where the arm pumps multiple times before swinging up.

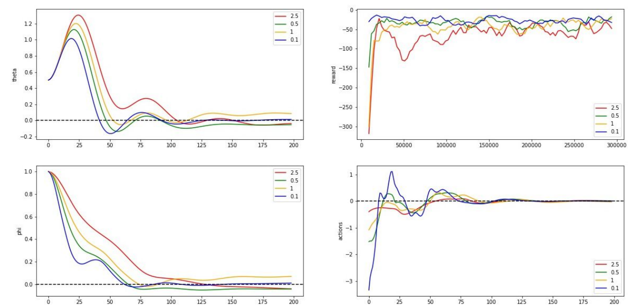

Fig. 3. Flexible link manipulator with friction with $y = x_1$.

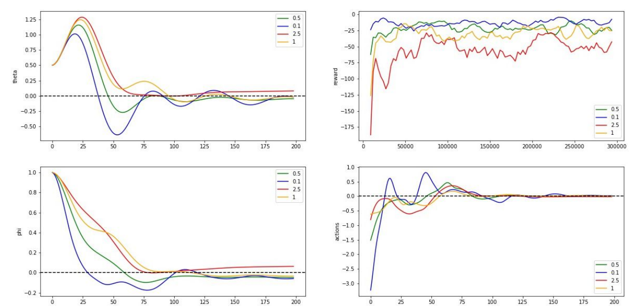

Fig. 4. Flexible link manipulator with friction with $y = x_3$.

V. Conclusion

In this paper, we studied how the geometry of a control system introduces computational limitations when practically solving for optimal controllers. Through the lenses of receding horizon control and cheap control, we identify a separation in qualitative behaviour between minimumphase and non-minimumphase systems. In particular, we show the existence of performance limitations between these systems by highlighting how for the former class of systems, we can use shorter horizons for RHC to compute a stabilizing controller. As similar results generally don't hold for the latter class, this suggests how the choice of cost functions that determine the geometric properties of the control system can greatly affect the practical performance of controllers synthesized with limited computational resources. Further we experimentally verified this intuitions by testing a RL algorithm to solve various stabilization problems. For future work, we hope to develop lower bounds for non-minimumphase systems to show that a RHC prediciton horizon can't be made arbitrarily small while ensuring stability of the internal dynamics. Additionally, we plan to study these results in the context of more general reinforcement learning algorithms to explore how their theoretical limits and practical performance varies with different classes of nonlinear systems.

References

[1] G. Grimm, M. J. Messina, S. E. Tuna, and A. R. Teel, "Model predictive control: for want of a local control lyapunov function, all is not lost," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 546–558, 2005.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-dynamic programming*. Athena Scientific, 1996.

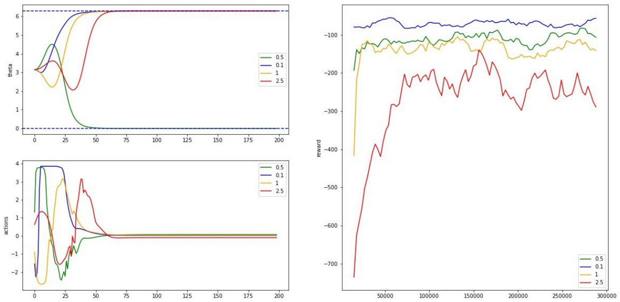

Fig. 5. Inverted pendulum without input constraints and $y = x_1$.

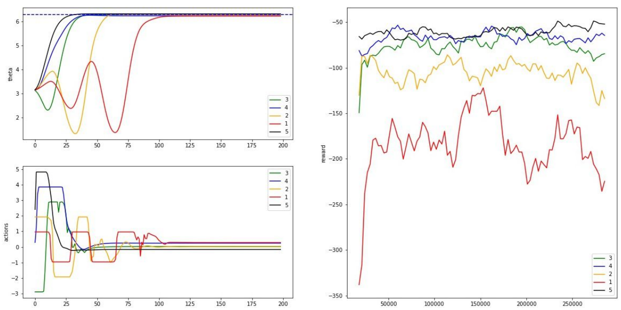

Fig. 6. Inverted pendulum with input constraints and $y = x_1$.

[3] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. PMLR, 2018, pp. 1861–1870.

[4] J. W. Grizzle, C. Chevallereau, R. W. Sinnet, and A. D. Ames, "Models, feedback control, and open problems of 3d bipedal robotic walking," *Automatica*, vol. 50, no. 8, pp. 1955–1988, 2014.

[5] J. B. Rawlings, D. Q. Mayne, and M. Diehl, *Model predictive control: theory, computation, and design*. Nob Hill Publishing Madison, 2017, vol. 2.

[6] M. Bardi, I. C. Dolcetta *et al.*, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer, 1997, vol. 12.

[7] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.

[8] A. Jadbabaie, J. Yu, and J. Hauser, "Receding horizon control of the caltech ducted fan: A control lyapunov function approach," in *Proceedings of the 1999 IEEE International Conference on Control Applications (Cat. No. 99CH36328)*, vol. 1. IEEE, 1999, pp. 51–56.

[9] T. Haarnoja, S. Ha, A. Zhou, J. Tan, G. Tucker, and S. Levine, "Learning to walk via deep reinforcement learning," *arXiv preprint arXiv:1812.11103*, 2018.

[10] M. Granzotto, R. Postoyan, D. Nešić, L. Buşoniu, and J. Daafouz, "When to stop value iteration: stability and near-optimality versus computation," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 412–424.

[11] Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for robust parameterized locomotion control of bipedal robots," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 2811–2817.

[12] J. Siekmann, Y. Godse, A. Fern, and J. Hurst, "Sim-to-real learning of all common bipedal gaits via periodic reward composition," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 7309–7315.

[13] G. Grimm, M. J. Messina, S. E. Tuna, and A. R. Teel, "Examples when nonlinear model predictive control is nonrobust," *Automatica*, vol. 40, no. 10, pp. 1729–1738, 2004.

[14] V. Gaitsgory, L. Grüne, and N. Thatcher, "Stabilization with discounted optimal control," *Systems & Control Letters*, vol. 82, pp. 91–98, 2015.

[15] M. M. Seron, J. H. Braslavsky, P. V. Kokotovic, and D. Q. Mayne, "Feedback limitations in nonlinear systems: From bode integrals to cheap control," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 829–833, 1999.

[16] A. P. Aguiar, J. P. Hespanha, and P. V. Kokotović, "Performance limitations in reference tracking and path following for nonlinear systems," *Automatica*, vol. 44, no. 3, pp. 598–610, 2008.

[17] A. Saberi and P. Sannuti, "Cheap and singular controls for linear quadratic regulators," *IEEE Transactions on Automatic Control*, vol. 32, no. 3, pp. 208–219, 1987.

[18] P. Sannuti, "Direct singular perturbation analysis of high-gain and cheap control problems," *Automatica*, vol. 19, no. 1, pp. 41–51, 1983.

[19] S. Sastry, *Nonlinear systems: analysis, stability, and control*. Springer Science & Business Media, 1999, vol. 10.

[20] B. Francis, "The optimal linear-quadratic time-invariant regulator with cheap control," *IEEE Transactions on Automatic Control*, vol. 24, no. 4, pp. 616–621, 1979.

[21] J. H. Braslavsky, M. M. Seron, and P. V. Kokotović, "Near-optimal cheap control of nonlinear systems," *IFAC Proceedings Volumes*, vol. 31, no. 17, pp. 107–112, 1998.

[22] R. Sepulchre, M. Jankovic, and P. V. Kokotovic, *Constructive nonlinear control*. Springer Science & Business Media, 2012.

[23] A. Jadbabaie and J. Hauser, "On the stability of receding horizon control with a general terminal cost," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 674–678, 2005.

[24] T. Pérez, G. C. Goodwin, and M. M. Serón, "Cheap control performance limitations of input constrained linear systems," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 503–508, 2002.

[25] M. Granzotto, R. Postoyan, L. Buşoniu, D. Nešić, and J. Daafouz, "Finite-horizon discounted optimal control: stability and performance," *IEEE Transactions on Automatic Control*, vol. 66, no. 2, pp. 550–565, 2020.

APPENDIX

A. Auxilary Lemmas

Lemma 1. *Let Assumptions 1,2, and 5 hold. Also assume that $\dot{\eta} = f_0(\eta)$ is globally exponentially stable. Then there exist $K_1 > 0$ such that for each $0 < \epsilon \le 1$ the infinite horizon performance satisfies $V^{\epsilon}_{\infty}(x) \le K_1\big(\epsilon\|\tilde{\xi}_2^2\| + \epsilon^{2(r+1)}\|\eta\|_2^2\big)$.*

Proof: Let $\tilde{u} = \tilde{A}^{-1}(\tilde{\xi},\eta)[-\tilde{b}(\tilde{\xi},\eta) + K\tilde{\xi}]$. where K is chosen such that for some $M > 0$, $\|\tilde{\xi}(t)\|_2 \le Me^{-\frac{t}{\epsilon}}\|\tilde{\xi}(0)\|_2$ for all $t \ge 0$. Applying this generally suboptimal controller gives us an upper bound on $V^{\epsilon}_{\infty}(x)$:

$$
\begin{aligned}
V^{\epsilon}_{\infty}(x) &\le \int_0^{\infty} (\|\tilde{\xi}_1(t)\|_2^2 + \|\tilde{A}^{-1}(\tilde{\xi},\eta)[-\tilde{b}(\tilde{\xi},\eta) + K\tilde{\xi}]\|_2^2) \cdot dt \\
&\le \int_0^{\infty} (\|\tilde{\xi}_1(t)\|_2^2 + \|\tilde{A}^{-1}(\tilde{\xi},\eta)\|_2^2\|\tilde{b}(\tilde{\xi},\eta)\|_2^2 \\
&\quad + \|\tilde{A}^{-1}(\tilde{\xi},\eta)\|_2^2\|K\|_2^2\|\tilde{\xi}(t)\|_2^2) \cdot dt
\end{aligned}
$$

Note that, by assumption, $\|\tilde{A}^{-1}(\tilde{\xi},\eta)\| \le \frac{1}{\gamma^2}$. As well, $\|\tilde{\xi}_1(t)\|_2 \le \|\tilde{\xi}(t)\|_2$ for all $t \ge 0$. Now consider the following bound on $\|b(\xi,\eta)\|_2^2$ based on the growth assumption, we can see that $\|b(\xi,\eta)\|_2^2 \le 3L^2(\|\xi\|_2^2 + \|\eta\|_2^2)$ using the AM-GM inequality on $\|\xi\|_2\|\eta\|_2$. Hence we can find a bound on $\|\tilde{b}(\tilde{\xi},\eta)\|_2^2 = \epsilon^{2r}\|b(S^{-1}(\epsilon)\tilde{\xi},\eta)\|_2^2 \le 3L^2\epsilon^{2r}(\|S^{-1}(\epsilon)\tilde{\xi}\|_2^2 + \|\eta\|_2^2) \le 3L^2\epsilon^{2r}(\max\{1, \frac{1}{\epsilon^{2(r-1)}}\}\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2)$. We now try to find a bound on $\|\eta(t)\|_2^2$. As $\dot{\eta} = f_0(\eta)$ is exponential minimum phase, by the converse Lyapunov theorem $\exists c_1, c_2, c_3, c_4 > 0$ and $V(\eta)$ a Lyapunov function s.t $\forall\eta$, $c_1\|\eta\|_2^2 \le V(\eta) \le c_2\|\eta\|_2^2$, $\frac{dV(\eta)}{d\eta}f_0(\eta) \le -c_3\|\eta\|_2^2$, and $\|\frac{dV(\eta)}{d\eta}\|_2 \le c_4\|\eta\|_2$. Now consider the time derivative of

$V(\eta)$:

$$\begin{aligned}\dot{V}(\eta) &= \frac{dV(\eta)}{d\eta}(f_0(\eta)+g_0(\eta)\tilde{\xi}_1)\\ &\le -c_2\|\eta\|_2^2 + \|\frac{dV(\eta)}{d\eta}\|_2\|g_0(\eta)\|_2\|\tilde{\xi}_1\|_2\\ &\le -c_2\|\eta\|_2^2 + c_4\|g_0(\eta)\|_2\|\eta\|_2\|\tilde{\xi}\|_2\\ &\le (c_4^2Ck - c_2)\|\eta\|_2^2 + \frac{1}{k}\|\tilde{\xi}\|_2^2\\ &\le (\frac{c_4^2Ck}{c_2} - 1)V(\eta) + \frac{1}{k}\|\tilde{\xi}\|_2^2 = -\tilde{c}V(\eta) + \frac{1}{k}\|\tilde{\xi}\|_2^2\end{aligned}$$

where k s.t $\tilde{c} = 1 - \frac{c_4^2Ck}{c_2} > 0$. As this is a linear ODE in $V(\eta)$, then we can find the following bound:

$$\begin{aligned}V(\eta(t)) &\le e^{-\tilde{c}t}V(\eta(0)) + \frac{1}{k}\int_0^t e^{-\tilde{c}(t-\tau)}\|\tilde{\xi}(\tau)\|\cdot d\tau\\ &\le e^{-\tilde{c}t}V(\eta(0)) + \frac{M^2\|\tilde{\xi}(0)\|_2^2 e^{-\tilde{c}t}}{k}\int_0^t e^{(\tilde{c}-\frac{1}{\epsilon})\tau}\cdot d\tau\\ \|\eta(t)\|_2^2 &\le \frac{c_2}{c_1}e^{-\tilde{c}t}\|\eta(0)\|_2^2 + \epsilon\frac{M^2\|\tilde{\xi}(0)\|_2^2}{kc_1(\tilde{c}-\frac{1}{\epsilon})}e^{-\tilde{c}t}[e^{(\tilde{c}-\frac{1}{\epsilon})t} - 1]\end{aligned}$$

Putting these bounds together gives the following:

$$\begin{aligned}V_\infty^\epsilon(x) &\le \int_0^\infty (\|\tilde{\xi}\|_2^2 + \frac{3L^2\epsilon^{2r}}{\gamma^2}(\max\{1, \frac{1}{\epsilon^{2(r-1)}}\}\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2)\\ &+ \frac{\|K\|_2^2}{\gamma^2}\|\tilde{\xi}\|_2^2)\cdot dt\\ &\le \int_0^\infty [(1 + \frac{\|K\|_2^2}{\gamma^2} + \frac{3L^2\epsilon^{2r}}{\gamma^2}\max\{1, \frac{1}{\epsilon^{2(r-1)}}\})\|\tilde{\xi}\|_2^2\\ &+ \frac{3L^2\epsilon^{2r}}{\gamma^2}(\frac{c_2}{c_1}e^{-\tilde{c}t}\|\eta(0)\|_2^2\\ &+ \epsilon\frac{M^2\|\tilde{\xi}(0)\|_2^2}{kc_1(\tilde{c}-\frac{1}{\epsilon})}e^{-\tilde{c}t}[e^{(\tilde{c}-\frac{1}{\epsilon})t} - 1])]\cdot dt\\ &\le \int_0^\infty [(k_1(\epsilon)e^{-\frac{t}{\epsilon}} + k_2(\epsilon)e^{-\tilde{c}t}[e^{(\tilde{c}-\frac{1}{\epsilon})t} - 1])\|\tilde{\xi}(0)\|_2^2\\ &+ k_3(\epsilon)e^{-\tilde{c}t}\|\eta(0)\|_2^2]\cdot dt\\ &= \mathcal{O}(\epsilon)\|\tilde{\xi}(0)\|_2^2 + \mathcal{O}(\epsilon^{2(r+1)})\|\eta(0)\|_2^2\end{aligned}$$

where $k_1(\epsilon) = M^2(1 + \frac{\|K\|_2^2}{\gamma^2} + \frac{3L^2\epsilon^{2r}}{\gamma^2}\max\{1, \frac{1}{\epsilon^{2(r-1)}}\})$, $k_2(\epsilon) = \frac{3L^2\epsilon^{2r}}{\gamma^2}$ and $k_3(\epsilon) = \frac{3L^2\epsilon^{2r+1}c_2}{\gamma^2 c_1}\frac{M^2}{kc_1(\tilde{c}-\frac{1}{\epsilon})}$. ■

Lemma 2. *Let the assumptions in Lemma 1 hold. Then there exists constants $C_1, C_2, C_3 > 0$ and such that for each $\epsilon > 0$ sufficiently small there exists a continuously differentiable function $W^\epsilon\colon \mathbb{R}^n \to \mathbb{R}^n$ which satisfies for each $(\tilde{\xi}, \eta) \in \mathbb{R}^n$ and $u \in \mathbb{R}^q$:*

$$C_1\big(\epsilon\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2\big) \le W^\epsilon(\tilde{\xi}, \eta) \le C_2\big(\epsilon\|\tilde{\xi}\| + \|\eta\|_2^2\big) \quad (28)$$

$$\dot{W}(\tilde{\xi}, \eta, \tilde{u}) \le -C\big(\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2\big) + \|\tilde{\xi}_1\|_2^2 + \|\tilde{u}\|_2^2 \quad (29)$$

Proof: First, choose the matrix L such that the matrix $F + LH$ is Hurwitz with F, H as in (9) (note that this is possible because (F, H) is observable). Then select P such that $(F + LH)^TP + P(F + LH) = -2I$. Next, let the function $V(\eta)$ be obtained from the converse exponential stability theorem as in the proof of Lemma 1. Without loss we assume that $c_3 > 2$ (if this is not the case we can re-scale V to αV below for $\alpha > 1$ sufficiently large). We then define for each $\epsilon > 0$, $W^\epsilon(\tilde{\xi}, \eta) = \epsilon\tilde{\xi}^TP\tilde{\xi} + V(\eta)$. The time derivative of this function satisfies:

$$\begin{aligned}\dot{W}^\epsilon(\tilde{\xi}, \eta, \tilde{u}) &= \tilde{\xi}^T(F^TP + PF)\xi + 2\tilde{\xi}^TPG\epsilon^r\tilde{b}(\tilde{\xi}, \eta)\\ &+ 2\tilde{\xi}^TPG\tilde{A}(\tilde{\xi}, \eta)u + \frac{d}{d\eta}V(\eta)[f_0(\eta) + g_0(\eta)\tilde{\xi}_1]\end{aligned} \quad (30)$$

Using the definition of P above and $\frac{d}{d\eta}V(\eta)f_0(\eta) < -c_3\|\eta\|_2^2 < -2\|\eta\|_2^2$ and $\|\frac{d}{d\eta}V(\eta)\|_2 \le c_4$ as in the proof of Lemma **??**, and the growth conditions in Assumption 5, in particular, $\epsilon^r\|\tilde{b}(\tilde{\xi}, \eta)\|_2 \le L(\epsilon\|\tilde{\xi}\|_2 + \epsilon^r\|\eta\|_2)$, from the preceding equation we obtain:

$$\begin{aligned}&\dot{W}^\epsilon(\tilde{\xi}, \eta, \tilde{u}) \le -2(\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2) + 2\epsilon L\|PG\|_2\|\tilde{\xi}\|_2(\|\tilde{\xi}\|_2 + \|\eta\|_2)\\ &+ C\|PG\|_2\|\tilde{\xi}\|_2\|u\|_2 + c_4C\|\eta\|_2\|\tilde{\xi}\|_2 + \tilde{\xi}(H^TL^TP + PLH)\tilde{\xi}\\ &\le -2(\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2) + 2\epsilon L\|PG\|_2(\|\tilde{\xi}\|_2^2 + \|\eta\|_2\|\tilde{\xi}\|_2)\\ &+ C\|PG\|_2\|\tilde{\xi}\|_2\|u\|_2 + c_4C\|\eta\|_2\|\tilde{\xi}_1\|_2 + 2\|PL\|_2\|\tilde{\xi}\|_2\|\tilde{\xi}_1\|_2\\ &\le -2(\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2) + (\|\tilde{\xi}\|_2^2 + \|\eta\|_2^2) + 3\epsilon L\|PG\|_2(\|\tilde{\xi}\|_2^2\\ &+ \|\eta\|_2^2) + \frac{1}{2}C\|PG\|_2\|u\|_2^2 + (\frac{1}{2}c_4C + \|PL\|_2\|)\|\tilde{\xi}_1\|_2^2,\end{aligned}$$

where in the final step we have repeatedly used the AM-GM inequality. The final expression demonstrates that αW^ϵ has the desired properties for $\alpha < \min\{\frac{1}{2}C\|PG\|_2, \frac{1}{2}c_4C + \|PL\|_2\|\}$ and sufficiently small ϵ. ■

B. Proof of Theorem 1

Let u be such that $J_T(x_0, u) = V_T(x_0)$. Denote $\phi(\tau) = \phi(\tau, x_0, u)$ as the evolution of $x(t)$ for τ time units under control input u starting from x_0. Consider $j \in [0, T - \Delta t]$. Then

$$\begin{aligned}&V_T(x(\Delta t)) - V_T(x(0)) = V(\phi(\Delta t)) - \int_0^T l(\tau)\cdot d\tau\\ &\le -\int_0^T l(\tau)\cdot dt + \int_{\Delta t}^{T-j} l(\tau)\cdot d\tau\\ &+ \min_{\tilde{u}}\int_{T-j}^{T+\Delta t} l(\phi(\tau, \phi(T-j), \tilde{u}), \tilde{u}(\tau))\cdot d\tau\\ &\le -\int_0^{\Delta t} l(\tau)\cdot d\tau + V_j(\phi(T-j))\\ &\le \bar{\alpha}(\sigma(\phi(T-j))) - \int_0^{\Delta t} l(\tau)\cdot d\tau\end{aligned}$$

As $V_T(\phi(0)) \ge \int_0^T l(t)dt$:

$$\begin{aligned}W(\phi(T)) - W(\phi(0)) &\le -k\int_0^T \|x(t)\|_2^2 dt + \int_0^T l(t)dt\\ &= V_T(\phi(0)) - k\int_0^T \|x(t)\|_2^2 dt\end{aligned}$$

Noting that $0 \le W(x) \le \bar{\alpha}_W(\sigma(x))$ and $V_T(x) \le \bar{\alpha}(\sigma(x))$, we can thus rearrange and bound terms to show the following:

$$k\int_0^T \|x(t)\|_2^2 dt \le (\bar{\alpha}_W + \bar{\alpha}) \circ \sigma(\phi(0))$$

Now consider $t^* \in [0, T]$ such that

$$t^* = \arg\min_{t\in[0,T]} \|x(t)\|_2^2$$

which exists by continuity of $x(t)$. Then we can note that

$$\|x(t^*)\|_2^2 \le \frac{(\bar{\alpha}_W + \bar{\alpha}) \circ \sigma(\phi(0))}{kT}$$

Taking $j = T - t^*$ we have that

$$\|x(T-j)\|_2^2 \le \frac{(\bar{\alpha}_W + \bar{\alpha}) \circ \sigma(\phi(0))}{kT}$$

We can combine this with the previous result on $V_T(x(\Delta t)) - V_T(x(0))$ to get the following:

$$\begin{aligned} V_T(x(\Delta t)) - V_T(x(0)) \le &- \int_0^{\Delta t} l(\tau)\cdot d\tau \\ &+ \bar{\alpha}\left(\frac{(\bar{\alpha}_W + \bar{\alpha}) \circ \sigma(\phi(0))}{kT}\right) \end{aligned}$$

where we leverage the fact that $\bar{\alpha}$ is non-decreasing.

Now note that

$$V_T(x(0)) = \int_0^{\Delta t} l(t)dt + V_{T-\Delta t}(x(\Delta t))$$

Hence $\exists \bar{T} \ge 0$ s.t $\forall T \ge \bar{T}$

$$\begin{aligned} V_T(x(\Delta t)) - V_{T-\Delta t}(x(\Delta t)) &= V_T(x(\Delta t)) - V_T(x_0) \\ &\quad + \int_0^{\Delta t} l(\tau)d\tau \\ &\le \bar{\alpha}\left(\frac{(\bar{\alpha}_W + \bar{\alpha}) \circ \sigma(\phi(0))}{kT}\right) \end{aligned}$$

Also note that by assumption there exists k_1, k_2 s.t

$$\begin{aligned} W(x) + V_{T-\Delta t}(x) &\le W(x) + V_\infty(x) \\ &\le (k_1 + k_2)\|x\|^2 \end{aligned}$$

Hence we can show the following

$$\begin{aligned} &\frac{d}{dt}(W(x) + V_{T-t}(x)) \le -k\|x\|^2 \\ \implies &\frac{d}{dt}(W(x) + V_{T-t}(x)) \le -\bar{k}(W(x) + V_{T-t}(x)) \\ \implies &W(x) + V_{T-t}(x) \le e^{-\bar{k}t}(W(x(0)) + V_T(x(0))) \end{aligned}$$

where $\bar{k} = \frac{k}{k_1+k_2}$. Thus we have the following:

$$\begin{aligned} W(x(\Delta t)) + V_T(x(\Delta t)) &= W(x(\Delta t)) + V_{T-\Delta t}(x(\Delta t)) \\ &\quad + V_T(x(\Delta t)) - V_{T-\Delta t}(x(\Delta t)) \\ &\le e^{-\bar{k}\Delta t}(W(x(0)) + V_T(x(0))) \\ &\quad + \bar{\alpha}\left(\frac{(\bar{\alpha}_W + \bar{\alpha}) \circ \sigma(\phi(0))}{kT}\right) \end{aligned}$$

Defining $Y = W + V_T$ and $\alpha = (\bar{\alpha}_w + \bar{\alpha})$ completes the proof. Then, use the bound $\underline{\alpha}_W \sigma(x) \le W(x) \le Y_T(x)$ we have

$$Y_T(\Phi(\Delta t)) \le \left(e^{-\bar{k}\Delta t} + \bar{\alpha}\left(\frac{(\bar{\alpha}_W + \bar{\alpha})}{\underline{\alpha}_W kT}\right)\right) Y_T(\phi(0)) \quad (31)$$

C. Proof of Theorem 3

To show global exponential stability, we note that it is sufficient to show that infinite horizon cost is bounded in the initial state and there is a Lyapunov function for the discrete time system that geometrically converges. On the first point, observe that we can use the conclusion of Lemma 1, which applies to a more general class of systems. In particular, we can just ignore references to the zero dynamics η. We can apply similar reasoning to leverage the results of Theorem 1, which gives us Y_T. Thus we have that $Y_T(\Phi(\Delta t)) - Y_T(\phi(0)) \le (e^{-\bar{k}\Delta t} + \bar{\alpha}(\frac{(\bar{\alpha}_W + \bar{\alpha})}{\underline{\alpha}_W kT}) - 1) Y_T(\phi(0))$. Note that $\bar{\alpha} = \mathcal{O}(\epsilon)$ and $e^{-\bar{k}\Delta t} < 1$, so this can be made negative using sufficiently small ϵ for a given Δt and T. This gives global exponential stability. □

D. Proof of Theorem 2

Consider $V_1(\tilde{\xi}) = \epsilon \tilde{\xi}^\top P \tilde{\xi}$ with the same P, L, H as in Lemma 2 and $V_2(\eta)$ from the converse Lyapunov theorem with constants as defined in Lemma 1. Using an intermediate result from the aforementioned lemma, our bound on V_∞^ϵ, and noting that $\int_0^t \|\tilde{\xi}(\tau)\| \cdot d\tau \le V_\infty^\epsilon$, we can say $\exists k_1 > 0$:

$$\begin{aligned} V_2(\eta(t)) &\le e^{-\tilde{c}t} V_2(\eta(0)) + \frac{1}{k}\int_0^t e^{-\tilde{c}(t-\tau)} \|\tilde{\xi}(\tau)\| \cdot d\tau \\ &\le e^{-\tilde{c}t} V_2(\eta(0)) + \mathcal{O}(\epsilon)\frac{\|\tilde{\xi}(0)\|_2^2 + \|\eta(0)\|_2^2}{k} \end{aligned}$$

$$V_2(\eta(t)) - V_2(\eta(0)) \le (e^{-\tilde{c}t} - 1)V_2(\eta(0)) + \frac{R^2}{k}k_1\epsilon$$

By a similar approach in Lemma 2, we have that there exists a scaling α and $k_2 > 0$ s.t

$$\begin{aligned} \dot{\tilde{V}}_1(\tilde{\xi}) = \alpha \dot{V}_1(\tilde{\xi}) &\le -C\|\tilde{\xi}\|_2^2 + \|u\|_2^2 \\ &\le -\frac{b'}{\epsilon}\tilde{V}_1(\tilde{\xi}) + \|u\|_2^2 \\ \implies \tilde{V}_1(\tilde{\xi}) &\le e^{-\frac{b'}{\epsilon}t}\tilde{V}_1(\tilde{\xi}(0)) + \int_0^t e^{-\frac{b'}{\epsilon}(t-\tau)}\|u(\tau)\|_2^2 \cdot d\tau \\ &\le e^{-\frac{b'}{\epsilon}t}\tilde{V}_1(\tilde{\xi}(0)) + e^{-\frac{b'}{\epsilon}t}\int_0^t e^{\frac{b'}{\epsilon}\tau}\|u(\tau)\|_2^2 \cdot d\tau \\ &\le e^{-\frac{b'}{\epsilon}t}\tilde{V}_1(\tilde{\xi}(0)) + O(\epsilon)R^2 e^{-\frac{b'}{\epsilon}t}\frac{\mathcal{O}(\epsilon)}{b'}[e^{\frac{b'}{\epsilon}t} - 1] \\ &\le e^{-\frac{b'}{\epsilon}t}\tilde{V}_1(\tilde{\xi}(0)) + k_2\epsilon^2 R^2 \end{aligned}$$

where $b_1, b_2 > 0$ s.t $\epsilon b_1 \|\tilde{\xi}\|_2^2 \le V_1(\tilde{\xi}) \le \epsilon b_2 \|\tilde{\xi}\|_2^2$, and $b' = Cb_2$. Choosing $\epsilon \le \min\{\frac{\delta^2 k c_1}{k_1 R^2}, \frac{\delta^2 b_1 \alpha}{k_2 R^2}\}$ allows us to have exponential convergence of $(\tilde{\xi}, \eta)$ to the ball of radius δ.