Proceedings of Machine Learning Research vol vvv:1–26, 2026

Online Tracking with Predictions for Nonlinear Systems with Koopman Linear Embedding

Chih-Fan Pai CPAI@UCSD.EDU
Xu Shang X3SHANG@UCSD.EDU
Jiachen Qian JIQ012@UCSD.EDU
Yang Zheng ZHENGY@UCSD.EDU
Department of Electrical and Computer Engineering, University of California San Diego

Abstract

We study the problem of online tracking in unknown nonlinear dynamical systems, where only short-horizon predictions of future target states are available. This setting arises in practical scenarios where full future information and exact system dynamics are unavailable. We focus on a class of nonlinear systems that admit a Koopman linear embedding, enabling the dynamics to evolve linearly in a lifted space. Exploiting this structure, we analyze a model-free predictive tracking algorithm based on Willems' fundamental lemma, which imposes dynamic constraints using only past data within a receding-horizon control framework. We show that, for Koopman-linearizable systems, the cumulative cost and dynamic regret of the nonlinear tracking problem coincide with those of the lifted linear counterpart. Moreover, we prove that the dynamic regret of our algorithm decays exponentially with the prediction horizon, as validated by numerical experiments.
Keywords: Online control, data-driven predictive control, dynamic regret, Koopman lifting

1. Introduction

We study the online tracking problem in unknown nonlinear dynamical systems, where an agent aims to follow a time-varying target trajectory using only short-horizon predictions of future target states. This sequential decision-making problem arises in many real-world applications, such as robotics (Falcone et al., 2007), autonomous systems (Rios-Torres and Malikopoulos, 2016; Kim and Kumar, 2014; Zheng et al., 2016), and adaptive control (Van Trees, 2004; Krstic et al., 1995; Ioannou and Sun, 1996; Annaswamy and Fradkov, 2021), where complete knowledge of future targets or system dynamics is rarely available. At each time step, the agent observes the current system state, receives short-horizon predictions of future target states, and selects a control action to steer the system toward the evolving target. The goal is to minimize the cumulative tracking error as well as its control efforts over a finite horizon.

While online control of linear dynamical systems has received significant attention (Abbasi-Yadkori et al., 2014; Li et al., 2019; Yu et al., 2020; Foster and Simchowitz, 2020; Zhang et al., 2022; Karapetyan et al., 2023; Tsiamis et al., 2024; Agarwal et al., 2019; Hazan and Singh, 2022), its extension to nonlinear systems, especially in the presence of predictions, remains largely unexplored. In this work, we focus on a special class of nonlinear systems that admit a *Koopman linear embedding*: a lifting procedure into a higher-dimensional space in which the system evolution becomes linear (Brunton et al., 2016; Shang et al., 2024). This structure enables us to reformulate the original nonlinear problem as a linear one in the lifted space, allowing us to exploit tools from linear control. However, in practice, the exact Koopman embedding is often unknown. To address this, we study a model-free predictive tracking algorithm inspired by Willems' fundamental lemma (Williams et al., 2015). While the method follows the general template of Model Predictive Control (MPC) (Garcia et al., 1989), a widely used approach for online control with predictions, it departs from classical MPC by imposing nonlinear dynamics constraints purely from offline trajectory data (Shang et al., 2024), without requiring explicit system identification.

The asymptotic properties of MPC have been extensively studied under general assumptions on dynamics and cost functions (Diehl et al., 2010; Grüne and Pirkelmann, 2020; Angeli et al., 2012). However, non-asymptotic performance metrics remain less well understood. *Dynamic regret*, a standard measure in online learning (Jadbabaie et al., 2015) and control (Yu et al., 2020; Goel and Hassibi, 2020; Zhang et al., 2021; Lin et al., 2022), quantifies the performance gap between an online controller and the optimal control sequence in hindsight. This metric is particularly suited for nonstationary environments, such as tracking moving or adversarial targets (Abbasi-Yadkori et al., 2014; Dressel and Kochenderfer, 2019). Recent work has established dynamic regret guarantees for linear MPC, demonstrating near-optimal performance in both stochastic and adversarial disturbance settings (Yu et al., 2020; Zhang et al., 2021; Lin et al., 2021). However, these results rely on known linear dynamics, limiting their applicability to nonlinear or model-free settings. Extending dynamic regret guarantees to such systems remains a significant challenge due to the complexity of nonlinear dynamics and the absence of explicit models. To our knowledge, this work is the first to provide *dynamic regret* guarantees for online tracking in unknown Koopman-linearizable nonlinear systems.

Our contributions. More specifically, for systems that admit exact Koopman embeddings, we establish an equivalence between the nonlinear tracking problem and its linear counterpart in the Koopman space. In the offline setting with full target information, we show that the optimal tracking policy can be characterized by solving the lifted linear problem (Lemma 3.1). Moreover, for any online controller, its dynamic regret can be analyzed by the lifted linear counterpart (Corollary 3.1). Building on this insight, we study a model-free predictive tracking method (Algorithm 1) based on the extended Willems's fundamental lemma (Shang et al., 2024), which enforces the nonlinear dynamics constraint using offline trajectory data and computes control actions by solving a data-driven quadratic program at each time step. In particular, we establish the first dynamic regret bound, showing that the regret grows linearly with full horizon and decays exponentially with the prediction horizon (Theorem 5.1). Our design and analysis depart from standard approaches in two key aspects. First, stability (state boundedness) is ensured by a sufficiently long prediction horizon rather than a carefully designed terminal cost. Second, unlike most regret analyses that rely critically on positive definite stage costs, a condition violated by the Koopman structure, our results hold under the weaker assumptions of positive semidefinite stage costs and detectability.

Related works. Our work intersects with two key areas: Koopman operator theory and online (nonstochastic) control. We here review some closely related results.

(i) **Koopman operator theory**, which enables representing nonlinear dynamical systems as linear systems in a high-dimensional space via lifting functions/observables (Brunton et al., 2016). Standard approaches often approximate the infinite-dimensional Koopman operator using Extended Dynamic Mode Decomposition (Williams et al., 2015; Proctor et al., 2018), which requires a careful selection of lifting functions to avoid modeling bias (Korda and Mezić, 2018; Berberich and Allgöwer, 2020). Recent advances learn embeddings (Shi and Meng, 2022) or bypass explicit lifting (Shang et al., 2024), but these methods typically assume known or partially identified dynamics. In contrast, our work assumes exact Koopman embeddings and employs a model-free predictive control algorithm inspired by a recent Willems' fundamental lemma (Shang et al., 2024). By imposing dynamics constraints through offline trajectory data, we eliminate the need for system identification or lifting approximation, enabling robust tracking of unknown nonlinear systems with theoretical guarantees (Theorem 5.1).

(ii) **Online nonstochastic control,** which adopts a learning-theoretic perspective for decision-making under nonstochastic or adversarial environments. This class of approaches typically reformulates the decision-making problem into online convex optimization, intending to minimize *policy regret*, which compares the performance of the online policy against the best fixed policy in a parameterized class (Agarwal et al., 2019; Simchowitz et al., 2020; Hazan and Singh, 2022). While this framework allows for tractable algorithm design, it is less suited for nonstationary tasks such as target tracking with evolving references. Instead, we focus on *dynamic regret*, which measures the performance gap relative to the best sequence of control actions in hindsight. This stronger benchmark better captures the challenges of online tracking under predictions. Moreover, while most existing results apply to linear systems with known dynamics, our work provides dynamic regret guarantees in a nonlinear, model-free setup.

The rest of the paper is organized as follows. Section 2 presents the online tracking problem for nonlinear systems. Section 3 establishes the equivalence between the original nonlinear problem and its lifted linear counterpart. Section 4 presents a unified predictive tracking algorithm. Section 5 analyzes the dynamic regret. Section 6 provides numerical experiments, and Section 7 concludes the paper. Technical proofs and experimental details are given in the appendix.

Notation. For a matrix M, let $\|M\|$ denote its induced ℓ_2-norm, and let $\rho(M)$ denote its spectral radius, defined as the maximum absolute value of its eigenvalues. For a vector x, $\|x\|$ denotes the standard Euclidean norm, and $\|x\|_P := (x^\top P x)^{1/2}$ denotes the P-weighted norm for any $P \succeq 0$. We write $x_{1:t} = \{x_1, \dots, x_t\}$ for a sequence of vectors and also use it to denote $[x_1^\top, \dots, x_t^\top]^\top$ for notation simplicity. We utilize $[t] = \{1, \dots, t\}$ for the index set.

2. Problem Setup

We study an online control problem where the system initialized with z_1 evolves according to some unknown nonlinear dynamics $z_{t+1} = f(z_t, u_t)$. Here $z_t \in \mathbb{R}^{n_z}$ is the state and $u_t \in \mathbb{R}^{n_u}$ is the control action at time t. The agent aims to track a target trajectory $r_{1:T}$ over a finite horizon of length T. In the classical setting, the entire trajectory is known a priori (Anderson and Moore, 2007; Bertsekas, 2012). However, in many cases, the target may be unknown or even adversarial. We therefore consider the problem of *online tracking with predictions*, where at each time step $t \in [T]$, the agent interacts with the adversary (environment) as follows:

- the agent observes z_t and predictions $r_{t:t+W-1}$ over a prediction horizon W;
- the agent selects and applies control action u_t;
- the adversary selects r_{t+W}, and the agent incurs the stage cost $\|z_t - r_t\|_{Q_z}^2 + \|u_t\|_R^2$;
- the system evolves according to $z_{t+1} = f(z_t, u_t)$.

Let $\pi = \{\pi_t\}_{t=1}^T$ denote the online (causal) control policy followed by the agent. Based on the above interaction protocol, π_t can depend only on the available information up to that point, i.e., $u_t = \pi_t(z_{1:t}, r_{1:t+W-1})$. Let $\mathbf{u} = u_{1:T}$ and $\mathbf{r} = r_{1:T}$. We define the cumulative cost incurred by π as:

$$J_T(\mathbf{u}; \mathbf{r}) = \sum_{t=1}^{T} \ell(z_t, u_t; r_t), \quad \ell(z_t, u_t; r_t) := \|z_t - r_t\|_{Q_z}^2 + \|u_t\|_R^2, \tag{1}$$

where $Q_z \succeq 0$ and $R \succ 0$. For notation simplicity, we omit the dependence of J_T on z_1.

Dynamic regret. We measure the performance of the agent's policy π via the notion of dynamic regret, which compares the cumulative cost of π to that of the optimal noncausal policy π^* in hindsight (with full knowledge of the target trajectory $\mathbf{r}$ and system dynamics f):

$$\mathrm{Reg}_T(\pi) = J_T(\mathbf{u}; \mathbf{r}) - J_T^*, \tag{2}$$

where $J_T^* := \min_{\mathbf{u}} \left\{ J_T(\mathbf{u}; \mathbf{r}) \mid z_{t+1} = f(z_t, u_t) \right\}$. Dynamic regret reflects how well the online controller adapts to a nonstationary, possibly adversarial environment (Goel and Hassibi, 2020). This contrasts with policy regret (Agarwal et al., 2019), which compares against the best fixed policy from a parameterized class. We make no assumptions on the target trajectory $\mathbf{r}$ beyond boundedness; without loss of generality, we assume $\|r_t\| \leq 1$ for all $t \in [T]$. In this paper, we focus on a special class of nonlinear systems (Shang et al., 2024; Brunton et al., 2016).

Koopman-linearizable dynamics. A (nonlinear) system $z_{t+1} = f(z_t, u_t)$ is said to be *Koopman-linearizable* if there exist a lifting function $\psi : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ and matrices $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, and $C \in \mathbb{R}^{n_z \times n_x}$ such that, for all $z_t \in \mathbb{R}^{n_z}$, the lifted state $x_t = \psi(z_t)$ evolves linearly $x_{t+1} = Ax_t + Bu_t$ and the original state satisfies $z_t = Cx_t = C\psi(z_t)$. We refer to the tuple (A, B, C, ψ) as the Koopman embedding. This class of nonlinear dynamics strictly goes beyond standard linear systems and can be used as an alternative to classical local linearization techniques.

Assumption 2.1 *The nonlinear system f is Koopman-linearizable. In addition, there exists a uniform bound $D_\psi < \infty$ such that $\|\psi(r)\| \leq D_\psi$ for all $\|r\| \leq 1$.*

Example 2.1 ((Brunton et al., 2016)) *Consider the nonlinear system (3). With the lifted coordinates $x = \psi(z) := [z_1,\ z_2,\ z_1^2]^\top$, it admits an exact linear embedding (4).* □

The lifting function ψ is generally unknown even if it exists. In this work, we develop a model-free predictive controller that exploits the Koopman structure without requiring explicit knowledge of f or ψ, and provides a dynamic regret guarantee.

$$\begin{bmatrix} z_1^+ \\ z_2^+ \end{bmatrix} = \begin{bmatrix} 0.99 z_1 \\ z_2 + z_1^2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u. \tag{3}$$

$$\begin{bmatrix} x_1^+ \\ x_2^+ \\ x_3^+ \end{bmatrix} = \begin{bmatrix} 0.99 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0.99^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u. \tag{4}$$

3. From Nonlinear to Linear via Koopman Embedding

We here establish a formal equivalence between the original nonlinear tracking problem and its lifted linear counterpart, then use linear control theory to characterize the optimal offline noncausal policy.

3.1. Equivalence in the lifted space

Consider the nonlinear dynamics f initialized at z_1 with Koopman embedding (A, B, C, ψ). Let $x_1 = \psi(z_1)$. Define $Q := C^\top Q_z C$ and the lifted cumulative cost

$$J_T^{\mathrm{lft}}(\mathbf{u}; \mathbf{r}) := \sum_{t=1}^{T} \ell_{\mathrm{lft}}(x_t, u_t; r_t), \quad \ell_{\mathrm{lft}}(x_t, u_t; r_t) := \|x_t - \psi(r_t)\|_Q^2 + \|u_t\|_R^2.$$

We now present a lemma capturing the cost equivalence under Koopman lifting.

Lemma 3.1 *Let $x_1 = \psi(z_1)$. The following statements hold:*

(i) $J_T(\mathbf{u}; \mathbf{r}) = J_T^{\mathrm{lft}}(\mathbf{u}; \mathbf{r})$ *for any control sequence* $\mathbf{u}$*;*

(ii) $J_T^* = (J_T^{\mathrm{lft}})^*$, *where* $(J_T^{\mathrm{lft}})^* := \min_{\mathbf{u}} \{J_T^{\mathrm{lft}}(\mathbf{u}; \mathbf{r}) : x_{t+1} = Ax_t + Bu_t\}$.

Proof The proof is straightforward. Since $z_t = Cx_t$ and $r_t = C\psi(r_t)$, the stage cost satisfies $\|z_t - r_t\|_{Q_z}^2 = \|Cx_t - C\psi(r_t)\|_{Q_z}^2 = \|x_t - \psi(r_t)\|_Q^2$, and thus (i) follows. For (ii), because J_T and J_T^{lft} agree for all control sequences $\mathbf{u}$, their optimal values are identical. ■

We next define the lifted dynamic regret $\text{Reg}_T^{\text{lft}}(\pi) := J_T^{\text{lft}}(\mathbf{u}; \mathbf{r}) - (J_T^{\text{lft}})^*$. By Lemma 3.1, the following result follows immediately.

Corollary 3.1 (Equivalence of dynamic regrets) *For any online policy π, $\text{Reg}_T(\pi) = \text{Reg}_T^{\text{lft}}(\pi)$.*

This result allows us to analyze the dynamic regret (2) in the linear Koopman space, while drawing conclusions for the original nonlinear system. We impose the following assumption.

Assumption 3.1 *(A, B) is stabilizable, (Q, A) is detectable, $Q_z \succeq 0$, $Q \succeq 0$, and $R \succ 0$.*

The assumption $Q \succeq 0$ is nontrivial. Since the recovery matrix C is not full column rank, the lifted cost matrix $Q = C^\mathsf{T} Q_z C$ cannot be positive definite, even if $Q_z \succ 0$.

3.2. Optimal offline policy

We now analyze the optimal (noncausal) tracking policy. The key idea is to exploit an exact Koopman embedding to transform the nonlinear tracking problem into an equivalent linear-quadratic tracking problem in the lifted space. The two formulations are presented below:

$$J_T^* := \min_{\mathbf{u}} \textstyle\sum_{t=1}^T \ell(z_t, u_t; r_t) \qquad \text{s.t. } z_{t+1} = f(z_t, u_t),\ z_1 \text{ given}; \tag{N}$$

$$(J_T^{\text{lft}})^* := \min_{\mathbf{u}} \textstyle\sum_{t=1}^T \ell_{\text{lft}}(x_t, u_t; r_t) \qquad \text{s.t. } x_{t+1} = Ax_t + Bu_t,\ x_1 = \psi(z_1). \tag{L}$$

To solve the lifted problem (L), we first introduce the standard Riccati recursion.

Definition 3.1 (Riccati recursion) *Let $P_T = Q$ and define the backward Riccati recursion:*

$$P_t = Q + A^\mathsf{T} P_{t+1} A - A^\mathsf{T} P_{t+1} B \Sigma_t^{-1} B^\mathsf{T} P_{t+1} A, \quad K_t = \Sigma_t^{-1} B^\mathsf{T} P_{t+1} A$$

for all $t \in [T-1]$ where we denote $\Sigma_t = R + B^\mathsf{T} P_{t+1} B$.

We also introduce the following notation for brevity. Let the closed-loop matrix at time $t \in [T]$ be $A_{\text{cl},t} := A - BK_t$, and define the state transition matrix for any $t_1 < t_2 < T$ as $A_{\text{cl},t_1 \to t_2} := A_{\text{cl},t_2} A_{\text{cl},t_2-1} \cdots A_{\text{cl},t_1+1}$ with the convention $A_{\text{cl},t_1 \to t_1} := I$.

Now we are ready to characterize the optimal noncausal policy for (L), which follows from (Foster and Simchowitz, 2020; Zhang et al., 2021; Goel and Hassibi, 2022). For completeness, we give a self-contained derivation in Appendix A.

Theorem 3.1 *Let $u_t^* = \pi_t^*(x_t; \mathbf{r})$ denote the optimal policy for (L). This policy admits the following closed-form expression for all $t \in [T]$:*

$$u_t^* = \pi_t^*(x_t; \mathbf{r}) = -K_t(x_t - \psi(r_t)) - \textstyle\sum_{i=t}^{T-1} K_{t\to i} \left(A\psi(r_i) - \psi(r_{i+1})\right), \tag{5}$$

where K_t denotes the feedback gain and $K_{t\to i} := \Sigma_t^{-1} B^\mathsf{T} A_{\text{cl},t\to i}^\mathsf{T} P_{i+1}$ the feedforward gain, both computed via the Riccati recursion in Definition 3.1.

By the cost equivalence established in Lemma 3.1, the following result follows directly.

Corollary 3.2 *The optimal control sequence from (5) achieves the minimum cost J_T^* for (N).*

While (5) gives a closed-form solution to the nonlinear problem (N), this policy is not implementable as it requires full knowledge of the target trajectory and the exact Koopman embedding. Instead, it serves as an offline benchmark for evaluating the dynamic regret of our proposed online controller introduced in the next section.

Algorithm 1 Predictive Tracking

1: **for** $t \in [T-W]$ **do**
2: Observe current state z_t and receive target predictions $r_{t:t+W-1}$.
3: Solve (N-MPC), (L-MPC), or (DDPC), and apply the control action $u_t^{\text{MPC}} := u_{1|t}^*$.
4: **end for**
5: At time step $t = T-W+1$, observe current state z_t and receive target predictions $r_{t:T}$.
6: Solve (N-MPC), (L-MPC), or (DDPC), and apply the control actions $u_{t:T}^{\text{MPC}} := u_{1:W|t}^*$.

4. Predictive Tracking for Koopman-linearizable Systems

MPC is a widely used online control strategy that repeatedly solves a finite-horizon optimal control problem using short-term predictions. At each step, only the first control action from the optimized sequence is executed, and the procedure repeats. Below, we introduce a unified predictive tracking algorithm that encompasses three MPC variants, differing in the level of model knowledge:

- **Nonlinear MPC**, which optimizes directly over the true nonlinear dynamics;
- **Lifted Linear MPC**, which assumes access to a known Koopman embedding;
- **Data-driven MPC**, which uses a data-driven dynamics constraint without model knowledge.

Note that Algorithm 1 follows the standard MPC procedure up to time step $T-W$. After that, a final open-loop control sequence is applied over the remaining horizon.

4.1. Predictive control with known dynamics

We start with the case where the nonlinear dynamics f and its Koopman embedding (A, B, C, ψ) are known. In this case, Algorithm 1 can be instantiated using either the nonlinear formulation (N-MPC) or the lifted linear formulation (L-MPC), as detailed below:

$$\min_{u_{1:W|t}} \sum_{i=1}^{W} \ell(z_{i|t}, u_{i|t}; r_{t+i-1}) \quad \text{s.t. } z_{i+1|t} = f(z_{i|t}, u_{i|t}),\ z_{1|t} = z_t; \tag{N-MPC}$$

$$\min_{u_{1:W|t}} \sum_{i=1}^{W} \ell_{\text{lft}}(x_{i|t}, u_{i|t}; r_{t+i-1}) \quad \text{s.t. } x_{i+1|t} = Ax_{i|t} + Bu_{i|t},\ x_{1|t} = x_t = \psi(z_t). \tag{L-MPC}$$

This lifted formulation (L-MPC) allows the use of linear control techniques. A key observation is that Theorem 3.1 can be applied to characterize this online MPC policy.

Theorem 4.1 *Let $u_t^{\text{MPC}} = \pi_t^{\text{MPC}}(x_t; \mathbf{r})$ denote the online MPC policy induced by solving (L-MPC) in Algorithm (1). This policy admits the following closed-form expression:*

(i) *For all $t \in [T-W]$, the policy is given by*

$$u_t^{\text{MPC}} := u_{1|t}^* = -\bar{K}_1(x_t - \psi(r_t)) - \sum_{i=t}^{t+W-2} \bar{K}_{1\to i-t+1} \left(A\psi(r_i) - \psi(r_{i+1})\right), \tag{6}$$

where $\bar{K}_1 := K_{T-W}$ denotes the feedback gain and $\bar{K}_{1\to k} := K_{T-W\to T-W+k}$ for $k \in [W]$ the feedforward gain, all defined via the Riccati recursion in Definition 3.1 and Theorem 3.1.

(ii) *For $T-W < t \leq T$, the policy coincides with the optimal offline policy from Theorem 3.1, i.e.,*

$$\pi_t^{\text{MPC}}(x_t; \mathbf{r}) = \pi_t^*(x_t; \mathbf{r}).$$

For $t \in [T-W]$, the MPC gains $\bar{K}_1$ and $\{\bar{K}_{1\to k}\}_{k=1}^{W}$ are time-invariant, unlike the time-varying offline optimal gains K_t and $\{K_{t\to i}\}_{i=t}^{T-1}$ in Theorem 3.1. This invariance arises from the constant weight matrices Q and R. If the weights were time-varying, the resulting MPC gains would also vary with time (Zhang et al., 2021). By Lemma 3.1, the following result is immediate.

Corollary 4.1 *Let* $u_t^{\text{MPC}} = \pi_t^{\text{MPC}}(z_t; \mathbf{r})$ *denote the online MPC policy induced by solving* (N-MPC) *in Algorithm 1. Then* $\pi_t^{\text{MPC}}(z_t; \mathbf{r}) = \pi_t^{\text{MPC}}(x_t; \mathbf{r})$ *for all* $t \in [T]$.

4.2. Data-driven predictive control with unknown dynamics

We now consider the case where both the dynamics and Koopman embedding are unknown, but a sufficiently long input-state trajectory, denoted $\mathbf{u}_\text{d} := \bar{u}_{1:n_\text{d}}$ and $\mathbf{z}_\text{d} := \bar{z}_{1:n_\text{d}}$, is available. We then design a predictive controller using only the collected trajectory data. This approach is referred to as *data-driven predictive control* (DDPC).

Online DDPC for tracking. We instantiate Algorithm 1 by solving:

$$\min_{u_{1:W|t}, z_{1:W|t}, g} \sum_{i=1}^{W} \|z_{i|t} - r_{t+i-1}\|_{Q_z}^2 + \|u_{i|t}\|_R^2 \quad \text{s.t. } H_\text{d} g = [\mathbf{u}_{\text{ini},t}^\mathsf{T}, u_{1:W|t}^\mathsf{T}, \mathbf{z}_{\text{ini},t}^\mathsf{T}, z_{1:W|t}^\mathsf{T}]^\mathsf{T}, \tag{DDPC}$$

where H_d is a data matrix constructed from the offline trajectory $(\mathbf{u}_\text{d}, \mathbf{z}_\text{d})$ as detailed below, and $\mathbf{u}_{\text{ini},t} := u_{t-T_\text{ini}:t-1} \in \mathbb{R}^{n_u T_\text{ini}}$ and $\mathbf{z}_{\text{ini},t} := z_{t-T_\text{ini}:t-1} \in \mathbb{R}^{n_z T_\text{ini}}$ denote the most recent length-T_ini input-state trajectory, which serves as the initial condition for the predictive control. With a rich enough data library H_d and a sufficiently long initial trajectory $(\mathbf{u}_{\text{ini},t}, \mathbf{z}_{\text{ini},t})$, the equality constraint provides a model-free representation of the nonlinear system, as guaranteed by the extended Willem's fundamental lemma (Shang et al., 2024). The (DDPC) formulation offers two key advantages: (i) it bypasses explicit system identification and the need for choosing lifting functions, and (ii) it transforms the nonlinear constraint into a linear one. As we will see below, the control actions from (DDPC) coincide with those from (N-MPC) and (L-MPC). Below, we detail the data-driven constraint and the associated data requirement.

Construction of data library. We first describe the construction of H_d. It is obtained by partitioning the single length-n_d trajectory $(\mathbf{u}_\text{d}, \mathbf{z}_\text{d})$ into l shorter overlapping segments of length $L := T_\text{ini} + W$, where $l = n_\text{d} - L + 1$. Specifically, we form $H_\text{d} \in \mathbb{R}^{(n_u+n_z)L \times l}$ as below:

$$H_\text{d} := \begin{bmatrix} \mathbf{u}_\text{d}^1 & \mathbf{u}_\text{d}^2 & \cdots & \mathbf{u}_\text{d}^l \\ \mathbf{z}_\text{d}^1 & \mathbf{z}_\text{d}^2 & \cdots & \mathbf{z}_\text{d}^l \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(\mathbf{u}_\text{d}) \\ \mathcal{H}_L(\mathbf{z}_\text{d}) \end{bmatrix}, \quad \mathcal{H}_L(w_{1:n_\text{d}}) := \begin{bmatrix} w_1 & w_2 & \cdots & w_{n_\text{d}-L+1} \\ w_2 & w_3 & \cdots & w_{n_\text{d}-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ w_L & w_{L+1} & \cdots & w_{n_\text{d}} \end{bmatrix},$$

where $(\mathbf{u}_\text{d}^i = \bar{u}_{i:i+L-1}, \mathbf{z}_\text{d}^i = \bar{z}_{i:i+L-1})$ for $i \in [\ell]$ are the shorter input-state trajectories of length L. To ensure that the collected data is sufficiently informative to represent the behavior of the nonlinear system, we impose the following condition (Shang et al., 2024).

Definition 4.1 (Lifted excitation) *We say that* H_d *provides* lifted excitation of order L *if the matrix* $\begin{bmatrix} \mathbf{u}_\text{d}^1 & \cdots & \mathbf{u}_\text{d}^l \\ \psi(z_1^1) & \cdots & \psi(z_1^l) \end{bmatrix} \in \mathbb{R}^{(n_u L + n_x) \times l}$ *has full row rank, where* $z_1^i \in \mathbb{R}^{n_z}$ *denotes the first element of* $\mathbf{z}_\text{d}^i$.

Although the Koopman embedding is unknown, its dimension n_x can be identified from a sufficiently long trajectory (Markovsky and Dörfler, 2022), and in practice, an estimated upper bound often suffices. Moreover, for any Koopman-linearizable systems, a data matrix H_d satisfying lifted excitation is guaranteed to exist and can be constructed by designing $\mathbf{u}_\text{d}$ for the controllable Koopman embedding (Shang et al., 2024). Given lifted excitation and an initial trajectory of length larger than n_x, we are ready to present the model-free linear representation of the nonlinear system.

Theorem 4.2 (Model-free representation (Shang et al., 2024)) *Consider a Koopman-linearizable nonlinear system. Suppose* H_d *satisfies lifted excitation of order* L*. At time* t*, we collect the most recent input-state sequence* $\mathbf{u}_{\text{ini},t}, \mathbf{z}_{\text{ini},t}$ *of length* $T_\text{ini} \geq n_x$*. Then, the concatenated sequence*

$[\mathbf{u}_{\text{ini}}^\top, \mathbf{u}_{\text{F}}^\top, \mathbf{z}_{\text{ini}}^\top, \mathbf{z}_{\text{F}}^\top]^\top$, *where* $\mathbf{u}_{\text{F}} \in \mathbb{R}^{n_u W}, \mathbf{z}_{\text{F}} \in \mathbb{R}^{n_z W}$, *is a valid length-$L$ trajectory if and only if there exists* $g \in \mathbb{R}^l$ *such that* $H_{\text{d}} g = [\mathbf{u}_{\text{ini}}^\top, \mathbf{u}_{\text{F}}^\top, \mathbf{z}_{\text{ini}}^\top, \mathbf{z}_{\text{F}}^\top]^\top$.

To apply the above representation in (DDPC), we set $\mathbf{u}_{\text{F}}$, $\mathbf{z}_{\text{F}}$ as decision variables $u_{1:W|t}, z_{1:W|t}$. Theorem 4.2 shows that the feasible region of $u_{1:W|t}, z_{1:W|t}$ includes all future input-state trajectories consistent with the initial condition $\mathbf{u}_{\text{ini}}, \mathbf{z}_{\text{ini}}$, thereby fully capturing the system's behavior.

5. Performance Guarantees for Predictive Tracking

We here demonstrate the effectiveness of Algorithm 1, especially with (DDPC). We establish a dynamic regret bound and outline the main analysis strategy.

5.1. Dynamic regret bound

Before presenting the regret bound, we introduce key system-dependent quantities used in the analysis. Under Assumption 3.1, there exists a unique $P_\infty \succeq 0$ satisfying the discrete algebraic Riccati equation (DARE): $P_\infty = Q + A^\top P_\infty A - A^\top P_\infty B(R + B^\top P_\infty B)^{-1} B^\top P_\infty A$, with the optimal gain $K_\infty := (R + B^\top P_\infty B)^{-1} B^\top P_\infty A$ and closed-loop matrix $A_{\text{cl},\infty} := A - BK_\infty$ (Bertsekas, 2012; Zhou et al., 1996). The first quantity $\rho_\infty \in (0,1)$ denotes the exponential rate at which the Riccati recursion converges to P_∞ (Hager and Horowitz, 1976). Second, the factor $\gamma_\infty := \frac{1}{2}\left(1 + \rho(A_{\text{cl},\infty})\right) \in (0,1)$ captures the closed-loop stability of state transition matrices $A_{\text{cl},t_1 \to t_2}$. Finally, the stabilizing window $\Delta_{\text{stab}} := \mathcal{O}(\log\left(1 - \rho(A_{\text{cl},\infty})\right)^{-1})$ represents the minimal horizon length for the system to exhibit stability. All these quantities depend only on the lifted system parameters (A, B, Q, R) and are independent of the full horizon length T. Throughout, the $\mathcal{O}(\cdot)$ notation hides constants that depend only on (A, B, Q, R, D_ψ) but not on T.

Theorem 5.1 (Dynamic regret) *Let* $\lambda_\infty := \max\{\rho_\infty, \gamma_\infty\}$ *and suppose that the prediction horizon satisfies* $W \geq \Delta_{\text{stab}}$. *Under Assumptions 2.1 and 3.1, the dynamic regret of Algorithm 1 satisfies*

$$\text{Reg}_T(\text{MPC}) = \mathcal{O}(W^2 \lambda_\infty^{2W} T)^{1}.$$

To the best of our knowledge, Theorem 5.1 provides the first dynamic regret bound for online control of nonlinear systems that admit an exact Koopman embedding. The result shows that the regret grows linearly with the time horizon T, but decays exponentially with the prediction horizon W. In particular, a prediction horizon of length $W = \Theta(\log T)$ suffices to achieve constant $\mathcal{O}(1)$ regret, paralleling the results in (Yu et al., 2020; Zhang et al., 2021; Lin et al., 2021) for linear systems. Notably, this performance guarantee is obtained without requiring a terminal cost or knowledge of the dynamics, relying instead on a sufficiently long prediction horizon to ensure stability. This feature is unique to Koopman-linearizable systems due to the lifting process.

5.2. Dynamic regret analysis

We here provide a proof sketch of Theorem 5.1, with details deferred to Appendix C. Our analysis builds on dynamic regret results for linear MPC with predictions but differs in two crucial aspects due to the Koopman lifting structure: (i) the lifted stage cost matrix Q is only positive semidefinite, and (ii) our MPC formulation lacks a terminal cost.

The first key step uses Corollary 3.1 to reduce the analysis to the lifted linear system:

$$\text{Reg}_T(\text{MPC}) = \text{Reg}_T^{\text{lft}}(\text{MPC}) = J_T^{\text{lft}}(\mathbf{u}^{\text{MPC}}; \mathbf{r}) - (J_T^{\text{lft}})^*,$$

1. Hidden constants in $\mathcal{O}(\cdot)$ also do not depend on W throughout entire analysis. Moreover, this bound can be tightened to $\mathcal{O}(\rho_\infty^{2W} T)$ if $\max\{\rho_\infty, \gamma_\infty\} = \rho_\infty$; see Appendix C.7.

where $J_T^{\mathrm{lft}}(\mathbf{u}^{\mathrm{MPC}};\mathbf{r})$ is the cost incurred by u_t^{MPC} from Algorithm 1. By the cost difference lemma (Kakade, 2003), the dynamic regret can be expressed as a cumulative sum of control deviations.

Lemma 5.1 *The lifted dynamic regret can be expressed as* $\mathrm{Reg}_T^{\mathrm{lft}}(\mathrm{MPC}) = \sum_{t=1}^{T} \|u_t^{\mathrm{MPC}} - u_t^*\|_{\Sigma_t}^2$, *where* $u_t^{\mathrm{MPC}} = \pi_t^{\mathrm{MPC}}(x_t;\mathbf{r})$ *and* $u_t^* = \pi_t^*(x_t;\mathbf{r})$ *denote the control actions generated by the MPC and the optimal non-causal policy, respectively, both evaluated on the lifted MPC trajectory* x_t.

Let $w_t := A\psi(r_i) - \psi(r_{i+1})$. From Theorems 3.1 and 4.1, it is clear that the control deviation vanishes for $T - W < t \le T$, and admits the following decomposition for $t \in [T-W]$:

$$u_t^{\mathrm{MPC}} - u_t^* = (K_t - \bar{K}_1)\left(x_t - \psi(r_t)\right) + \sum_{i=t}^{t+W-2}(K_{t\to i} - \bar{K}_{1\to i-t+1})w_i + \sum_{i=t+W-1}^{T-1} K_{t\to i} w_i.$$

With the above expression and Lemma 5.1, we obtain the following regret bound decomposition

$$\begin{aligned}\mathrm{Reg}_T^{\mathrm{lft}}(\mathrm{MPC}) \le &\underbrace{\left(\sum_{t=1}^{T-W}\left\|\sum_{i=t+W-1}^{T-1} K_{t\to i} w_i\right\|_{\Sigma_t}^2\right)}_{\text{truncation deviation}} + \underbrace{\left(\sum_{t=1}^{T-W}\left\|(K_t - \bar{K}_1)\left(x_t - \psi(r_t)\right)\right\|_{\Sigma_t}^2\right)}_{\text{feedback deviation}} \\ &+ \underbrace{\left(\sum_{t=1}^{T-W}\left\|\sum_{i=t}^{t+W-2}(K_{t\to i} - \bar{K}_{1\to i-t+1})w_i\right\|_{\Sigma_t}^2\right)}_{\text{feedforward deviation}}. \end{aligned} \tag{7}$$

We now bound the regret term-by-term, with all technical details deferred to Appendices B and C. All the results below are stated with the same settings in Theorem 5.1. We first establish the exponential decay behavior of some key quantities.

Lemma 5.2 (Exponential decay of state transition matrices and feedforward gains) *For all* $t_1 < t_2 < T$, *we have* $\|A_{\mathrm{cl},t_1\to t_2}\| = \mathcal{O}(\gamma_\infty^{t_2-t_1})$ *and* $\|K_{t_1\to t_2}\| = \mathcal{O}(\gamma_\infty^{t_2-t_1})$.

The proof leverages the strong stability (Cohen et al., 2018) of the closed-loop matrix $A_{\mathrm{cl},\infty}$ together with the exponential convergence of the Riccati recursion; see details in Appendix C.2. This result directly yields a bound on the truncation deviation term in the regret decomposition.

Proposition 5.1 *The truncation deviation term in* (7) *satisfies* (*truncation deviation*) $= \mathcal{O}(\gamma_\infty^{2W} T)$.

Next, to handle the remaining two terms in the regret decomposition, we establish exponential decay bounds for the gain deviations and ensure boundedness of the lifted MPC state trajectory.

Lemma 5.3 (Exponential decay of feedback and feedforward deviation) *For* $t \in [T-W]$ *and* $t \le i \le t+W-2$, *we have* $\|K_t - \bar{K}_1\| = \mathcal{O}(\rho_\infty^W)$ *and* $\|K_{t\to i} - \bar{K}_{1\to i-t+1}\| = \mathcal{O}(\gamma_\infty^{i-t}\rho_\infty^{W-i+t})$.

The proof of bounding gain deviations is more delicate and involved, but builds on similar techniques used in the proof of Lemma 5.2; see Appendices C.4 and C.5 for details.

The bound below ensures the lifted MPC state remains bounded; see Section C.6 for proof.

Lemma 5.4 (Uniform bound on lifted MPC states) *Suppose* $W \ge \Delta_{\mathrm{stab}}$. *There exists a constant* $D_x > 0$, *independent of* T, *such that for all* $t \in [T-W]$, *we have* $\|x_t - \psi(r_t)\| \le D_x$.

The key idea in Lemma 5.4 is to leverage a sufficiently long prediction horizon. This contrasts with standard MPC, which often relies on a specially designed terminal cost to guarantee stability (Yu et al., 2020; Zhang et al., 2021). Relying on Lemmas 5.3 and 5.4, we obtain the following bounds.

Proposition 5.2 *Suppose* $W \ge \Delta_{\mathrm{stab}}$. *In* (7), *we have (feedback gain deviation)* $= \mathcal{O}(\rho_\infty^{2W} T)$ *and (feedforward gain deviation)* $= \mathcal{O}(W^2\lambda_\infty^{2W} T)$.

It is now clear that combining Propositions 5.1 and 5.2 yields the desired bound in Theorem 5.1.

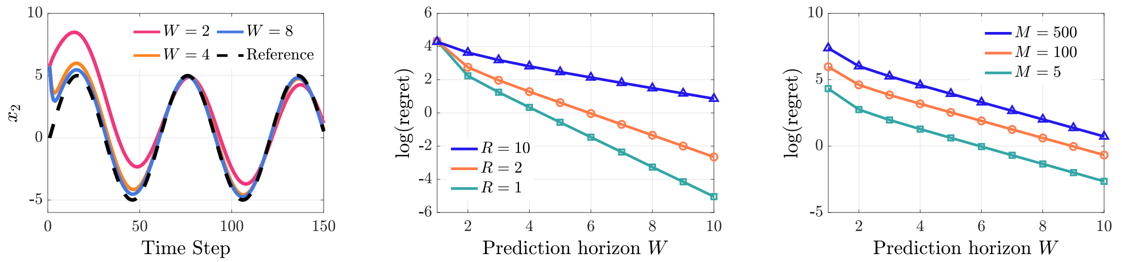

Figure 1: Performance of Algorithm 1 with (DDPC) under varying prediction horizon W: tracking and reference trajectories (left), dynamic regret for different R (middle) and different M (right).

6. Numerical Experiments

We validate Theorem 5.1 by evaluating the tracking performance of (DDPC) on the nonlinear system (8), which similar to Example 2.1 is Koopman-linearizable with the lifted state $x := [z_1, z_2, z_1^2, z_1^3, z_1^4]^\top$. Here, the control objective is to let the second state z_2 track a sinusoidal reference $r_{2,t} = M\sin(\pi t/30)$. In the experiment, we set the horizon $T = 200$, state cost matrix $Q_z = \mathrm{diag}(0, 1)$, and initial trajectory length $T_{\text{ini}} = 10$. For construction of the data library H_{d}, a single trajectory of length $2W + 24$ is collected for each prediction horizon W (the minimum length to satisfy the lifted excitation), where the control input is uniformly sampled from $[-1, 1]$.

$$\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} = \begin{bmatrix} 0.99 z_{1,t} \\ 0.9 z_{2,t} + z_{1,t}^2 + z_{1,t}^3 + z_{1,t}^4 + u_t \end{bmatrix} \quad (8)$$

We show in Figure 1 the tracking performance and dynamic regret. In the left subfigure, we observe that a longer prediction horizon W yields smaller tracking error and faster convergence to the reference trajectory, as indicated by the quicker correction in the blue curve. The middle and right subfigures show that the dynamic regret decreases exponentially with increasing prediction horizon W, which aligns with our bound in Theorem 5.1. Moreover, we observe that the decay rate depends on the control cost matrix R and reference magnitude M.

Finally, to illustrate how the proposed data-driven predictive control method extends to systems that are not Koopman-linearizable, we conduct an additional experiment on path tracking of two-wheeled robots. Due to the page limit, we put these additional experimental results in Appendix D.

7. Conclusion

This paper studied the online performance of a data-driven predictive tracking method for a class of Koopman-linearizable nonlinear systems. The approach relies only on sufficiently rich offline data and requires no explicit model identification or knowledge of Koopman lifting functions. Leveraging the lifted representation, we established a dynamic regret bound that quantifies the tracking performance relative to the optimal offline control sequence and reveals the impact of the prediction horizon on regret. Numerical experiments supported our theoretical findings. Future directions include extending the data-driven method and regret analysis to settings with disturbances or approximate Koopman embeddings, relaxing data richness conditions, incorporating imperfect predictions, and exploring alternative regret measures.

Acknowledgments

This work is supported by NSF CMMI Award 2320697 and NSF CAREER Award 2340713.

References

Yasin Abbasi-Yadkori, Peter Bartlett, and Varun Kanade. Tracking adversarial targets. In *International Conference on Machine Learning*, pages 369–377. PMLR, 2014.

Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119. PMLR, 2019.

Brian DO Anderson and John B Moore. *Optimal control: linear quadratic methods*. Courier Corporation, 2007.

David Angeli, Rishi Amrit, and James B. Rawlings. On average performance and stability of economic model predictive control. *IEEE Transactions on Automatic Control*, 57(7):1615–1626, 2012. doi: 10.1109/TAC.2011.2179349.

Anuradha M Annaswamy and Alexander L Fradkov. A historical perspective of adaptive control and learning. *Annual Reviews in Control*, 52:18–41, 2021.

Julian Berberich and Frank Allgöwer. A trajectory-based framework for data-driven system analysis and control. In *2020 European Control Conference (ECC)*, pages 1365–1370. IEEE, 2020.

Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 4. Athena scientific, 2012.

Steven L Brunton, Bingni W Brunton, Joshua L Proctor, and J Nathan Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PloS one*, 11(2):e0150171, 2016.

Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1029–1038. PMLR, 2018.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. *Advances in Neural Information Processing Systems*, 31, 2018.

Moritz Diehl, Rishi Amrit, and James B Rawlings. A lyapunov function for economic optimizing model predictive control. *IEEE Transactions on Automatic Control*, 56(3):703–707, 2010.

Florian Dörfler, Jeremy Coulson, and Ivan Markovsky. Bridging direct and indirect data-driven control formulations via regularizations and relaxations. *IEEE Transactions on Automatic Control*, 68(2):883–897, 2022.

Louis Dressel and Mykel J Kochenderfer. Hunting drones with other drones: Tracking a moving radio target. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 1905–1912. IEEE, 2019.

Paolo Falcone, Francesco Borrelli, Jahan Asgari, Hongtei Eric Tseng, and Davor Hrovat. Predictive active steering control for autonomous vehicle systems. *IEEE Transactions on control systems technology*, 15(3):566–580, 2007.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International conference on machine learning*, pages 1467–1476. PMLR, 2018.

Dylan Foster and Max Simchowitz. Logarithmic regret for adversarial online control. In *International Conference on Machine Learning*, pages 3211–3221. PMLR, 2020.

Carlos E Garcia, David M Prett, and Manfred Morari. Model predictive control: Theory and practice—a survey. *Automatica*, 25(3):335–348, 1989.

Gautam Goel and Babak Hassibi. Regret-optimal control in dynamic environments. *arXiv preprint arXiv:2010.10473*, 2020.

Gautam Goel and Babak Hassibi. The power of linear controllers in lqr control. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 6652–6657. IEEE, 2022.

Lars Grüne and Simon Pirkelmann. Economic model predictive control for time-varying system: Performance and stability results. *Optimal Control Applications and Methods*, 41(1):42–64, 2020.

William W Hager and Larry L Horowitz. Convergence and stability properties of the discrete riccati operator equation and the associated optimal control and filtering problems. *SIAM Journal on Control and Optimization*, 14(2):295–312, 1976.

Elad Hazan and Karan Singh. Introduction to online nonstochastic control. *arXiv preprint arXiv:2211.09619*, 2022.

Petros A Ioannou and Jing Sun. *Robust adaptive control*, volume 1. PTR Prentice-Hall Upper Saddle River, NJ, 1996.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pages 398–406. PMLR, 2015.

Sham Machandranath Kakade. *On the sample complexity of reinforcement learning*. University of London, University College London (United Kingdom), 2003.

Aren Karapetyan, Diego Bolliger, Anastasios Tsiamis, Efe C Balta, and John Lygeros. Online linear quadratic tracking with regret guarantees. *IEEE Control Systems Letters*, 2023.

Kyoung-Dae Kim and Panganamala Ramana Kumar. An mpc-based approach to provable system-wide safety and liveness of autonomous ground traffic. *IEEE Transactions on Automatic Control*, 59(12):3341–3356, 2014.

Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.

Miroslav Krstic, Petar V Kokotovic, and Ioannis Kanellakopoulos. *Nonlinear and adaptive control design*. John Wiley & Sons, Inc., 1995.

Yingying Li, Xin Chen, and Na Li. Online optimal control with linear dynamics and predictions: Algorithms and regret analysis. *Advances in Neural Information Processing Systems*, 32, 2019.

Yiheng Lin, Yang Hu, Guanya Shi, Haoyuan Sun, Guannan Qu, and Adam Wierman. Perturbation-based regret analysis of predictive control in linear time varying systems. *Advances in Neural Information Processing Systems*, 34:5174–5185, 2021.

Yiheng Lin, Yang Hu, Guannan Qu, Tongxin Li, and Adam Wierman. Bounded-regret mpc via perturbation analysis: Prediction error, constraints, and nonlinearity. *Advances in Neural Information Processing Systems*, 35:36174–36187, 2022.

Ivan Markovsky and Florian Dörfler. Identifiability in the behavioral setting. *IEEE Transactions on Automatic Control*, 68(3):1667–1677, 2022.

Joshua L Proctor, Steven L Brunton, and J Nathan Kutz. Generalizing koopman theory to allow for inputs and control. *SIAM Journal on Applied Dynamical Systems*, 17(1):909–930, 2018.

Jackeline Rios-Torres and Andreas A Malikopoulos. A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps. *IEEE Transactions on Intelligent Transportation Systems*, 18(5):1066–1077, 2016.

Xu Shang, Jorge Cortés, and Yang Zheng. Willems' fundamental lemma for nonlinear systems with koopman linear embedding. *IEEE Control Systems Letters*, 2024.

Haojie Shi and Max Q-H Meng. Deep koopman operator with control for nonlinear systems. *IEEE Robotics and Automation Letters*, 7(3):7700–7707, 2022.

Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. In *Conference on Learning Theory*, pages 3320–3436. PMLR, 2020.

Anastasios Tsiamis, Aren Karapetyan, Yueshan Li, Efe C Balta, and John Lygeros. Predictive linear online tracking for unknown targets. *arXiv preprint arXiv:2402.10036*, 2024.

Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25:1307–1346, 2015.

Chenkai Yu, Guanya Shi, Soon-Jo Chung, Yisong Yue, and Adam Wierman. The power of predictions in online control. *Advances in Neural Information Processing Systems*, 33:1994–2004, 2020.

Runyu Zhang, Yingying Li, and Na Li. On the regret analysis of online lqr control with predictions. In *2021 American Control Conference (ACC)*, pages 697–703. IEEE, 2021.

Zhiyu Zhang, Ashok Cutkosky, and Ioannis Paschalidis. Adversarial tracking control via strongly adaptive online learning with memory. In *International Conference on Artificial Intelligence and Statistics*, pages 8458–8492. PMLR, 2022.

Yang Zheng, Shengbo Eben Li, Keqiang Li, Francesco Borrelli, and J Karl Hedrick. Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies. *IEEE Transactions on Control Systems Technology*, 25(3):899–910, 2016.

Kemin Zhou, John C. Doyle, and Keith Glover. *Robust and Optimal Control*. Prentice Hall, 1996.

Appendix A. The Optimal Noncausal Policy for Linear Quadratic Tracking (LQT)

This section derives the optimal offline (noncausal) policy given in Theorem 3.1 and characterizes its value function (cost-to-go). Recall $\ell_{\text{lft}}(x_t, u_t; r_t) := \|x_t - \psi(r_t)\|_Q^2 + \|u_t\|_R^2$ for $t \in [T]$ and the LQT problem (L) given full knowledge of the target trajectory $\mathbf{r} = r_{1:T}$

$$\begin{aligned} \min_{u_{1:T}} \quad & \textstyle\sum_{t=1}^{T} \ell_{\text{lft}}(x_t, u_t; r_t) \\ \text{s.t.} \quad & x_{t+1} = Ax_t + Bu_t, \ \ x_1 \text{ given.} \end{aligned}$$

To formalize the optimal control policy under reference tracking, we introduce notions of the optimal value function V_t^*, Q-function Q_t^*, and optimal policy π_t^*, defined as follows.

Definition A.1 (LQT optimal value function, Q-function, policy) *Define the terminal quantities:*

$$Q_T^*(x, u; \mathbf{r}) = \ell_{\text{lft}}(x, u; r_T), \quad \pi_T^*(x; \mathbf{r}) = \min_u Q_T^*(x, u; \mathbf{r}) = 0, \quad V_T^*(x; \mathbf{r}) = \ell_{\text{lft}}(x, 0; r_T).$$

For each $t \in [T-1]$, define recursively:

$$\begin{aligned} Q_t^*(x, u; \mathbf{r}) &= \ell_{\text{lft}}(x, u; r_t) + V_{t+1}^*(Ax + Bu; \mathbf{r}), \\ \pi_t^*(x; \mathbf{r}) &= \operatorname*{argmin}_u Q_t^*(x, u; \mathbf{r}), \\ V_t^*(x; \mathbf{r}) &= \min_u Q_t^*(x, u; \mathbf{r}) = Q_t^*(x, \pi_t^*(x; \mathbf{r}); \mathbf{r}). \end{aligned}$$

A.1. Equivalence between LQT and LQR with disturbance

We first establish the equivalence between LQT and linear quadratic regulator (LQR) with disturbances, which serves as a key step in the subsequent proof. By a change of variables, we define the lifted tracking error state $\tilde{x}_t$ and the disturbance sequence w_t as follows

$$\tilde{x}_t := x_t - \psi(r_t), \quad w_t := A\psi(r_t) - \psi(r_{t+1}), \tag{9}$$

we can readily verify that the LQT problem (L) is equivalent to LQR with disturbance $\mathbf{w} = w_{1:T-1}$

$$\begin{aligned} \min_{u_{1:T}} \quad & \textstyle\sum_{t=1}^{T} \tilde{x}_t^\mathsf{T} Q \tilde{x}_t + u_t^\mathsf{T} R u_t \\ \text{s.t.} \quad & \tilde{x}_{t+1} = A\tilde{x}_t + Bu_t + w_t, \ \ \tilde{x}_1 \text{ given.} \end{aligned} \tag{10}$$

To formalize the optimal control policy under this equivalent LQR setting, we define the corresponding value functions and policies in terms of $\tilde{x}_t$ and w_t, mirroring the structure in Definition A.1. Under this change of variables (9), we note that $\ell_{\text{lft}}(x, u; r_t) = \ell_{\text{lft}}(\tilde{x}, u; 0) = \tilde{x}^\mathsf{T} Q \tilde{x} + u^\mathsf{T} R u$.

Definition A.2 (LQR optimal value function, Q-function, policy) *Define the terminal quantities:*

$$\tilde{Q}_T^*(\tilde{x}, u; \mathbf{w}) := \ell_{\text{lft}}(\tilde{x}, u; 0), \quad \tilde{\pi}_T^*(\tilde{x}; \mathbf{w}) = \min_u \tilde{Q}_T^*(\tilde{x}, u; \mathbf{w}) = 0 \quad \tilde{V}_T^*(\tilde{x}; \mathbf{w}) := \ell_{\text{lft}}(\tilde{x}, 0; 0).$$

For each $t \in [T-1]$, define recursively:

$$\begin{aligned} \tilde{Q}_t^*(\tilde{x}, u; \mathbf{w}) &:= \ell_{\text{lft}}(\tilde{x}, u; 0) + \tilde{V}_{t+1}^*(A\tilde{x} + Bu + w_t; \mathbf{w}), \\ \tilde{\pi}_t^*(\tilde{x}; \mathbf{w}) &:= \operatorname*{argmin}_u \tilde{Q}_t^*(\tilde{x}, u; \mathbf{w}), \\ \tilde{V}_t^*(\tilde{x}; \mathbf{w}) &:= \min_u \tilde{Q}_t^*(\tilde{x}, u; \mathbf{w}) = \tilde{Q}_t^*(\tilde{x}, \tilde{\pi}_t^*(\tilde{x}; \mathbf{w}); \mathbf{w}). \end{aligned}$$

A.2. Proof of Theorem 3.1

We restate Theorem 3.1 in terms of the lifted error state $\tilde{x}$ and the disturbance sequence $\mathbf{w}$ as follows, and provide its proof.

Theorem A.1 *The optimal (noncausal) policy for (10), given full knowledge of* $\mathbf{w}$*, is given by*

$$u_t^* = \pi_t^*(\tilde{x}_t; \mathbf{w}) = -K_t \tilde{x}_t - \textstyle\sum_{i=t}^{T-1} K_{t \to i} w_i,$$

where the feedback gain K_t *and the feedforward gain* $K_{t\to i} = \Sigma_t^{-1} B^\top (A_{\mathrm{cl},i} \cdots A_{\mathrm{cl},t+1})^\top P_{i+1}$ *are defined via the Riccati recursion in Definition 3.1. Moreover, the corresponding optimal value function (cost-to-go) at time* t *is given by* $\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = \tilde{x}^\top P_t \tilde{x} + v_t^\top \tilde{x} + q_t$*, where the recursive relation of* (P_t, v_t, q_t) *is given in (12). Equivalently, in terms of the original lifted state* x *and* $\mathbf{r}$*, the value function admits the form* $V_t^*(x; \mathbf{r}) = \big(x - \psi(r_t)\big)^\top P_t \left(x - \psi(r_t)\right) + v_t^\top \left(x - \psi(r_t)\right) + q_t$*.*

Proof The proof proceeds via dynamic programming, recursively computing the optimal control actions. For $t \in [T-1]$, consider the optimal value function

$$\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = \min_u \left(\tilde{x}^\top Q \tilde{x} + u^\top R u + V_{t+1}^*(A\tilde{x} + Bu + w_t) \right),$$

with the terminal condition $\tilde{V}_T^*(\tilde{x}; \mathbf{w}) = \tilde{x}^\top Q \tilde{x}$. We aim to show that for all $t \in [T]$, the value function admits the quadratic form $\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = \tilde{x}^\top P_t \tilde{x} + v_t^\top \tilde{x} + q_t$. At $t = T$, the claim holds trivially with $(P_T, v_T, q_T) = (Q, 0, 0)$. We now proceed by backward induction. Suppose that $\tilde{V}_{t+1}^*(\tilde{x}) = \tilde{x}^\top P_{t+1} \tilde{x} + v_{t+1}^\top \tilde{x} + q_{t+1}$ holds for some $t \in [T-1]$ as the induction hypothesis. Then the value function at time t is

$$\begin{aligned}\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = \min_u \Big(&\tilde{x}^\top Q \tilde{x} + u^\top R u + (A\tilde{x} + Bu + w_t)^\top P_{t+1} (A\tilde{x} + Bu + w_t) \\ &+ v_{t+1}^\top (A\tilde{x} + Bu + w_t) + q_{t+1} \Big).\end{aligned}$$

The above is a quadratic optimization problem in u, rewritten as follows:

$$\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = \min_u \left(\begin{bmatrix} u \\ \tilde{x} \\ w_t \\ v_{t+1} \end{bmatrix}^\top \begin{bmatrix} R + B^\top P_{t+1} B & B^\top P_{t+1} A & B^\top P_{t+1} & \frac{1}{2} B^\top \\ A^\top P_T B & Q + A^\top P_{t+1} A & A^\top P_{t+1} & \frac{1}{2} A^\top \\ P_{t+1} B & P_{t+1} A & P_{t+1} & \frac{1}{2} I \\ \frac{1}{2} B & \frac{1}{2} A & \frac{1}{2} I & 0 \end{bmatrix} \begin{bmatrix} u \\ \tilde{x} \\ w_t \\ v_{t+1} \end{bmatrix} + q_{t+1} \right),$$

where the minimizer (optimal control action) is given by

$$u_t^* = \tilde{\pi}_t^*(\tilde{x}; \mathbf{w}) = -(R + B^\top P_{t+1} B)^{-1} B^\top \left(P_{t+1} A \tilde{x} + P_{t+1} w_t + \frac{1}{2} v_{t+1} \right). \tag{11}$$

Substituting (11) back into the value function yields

$$\begin{aligned}\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = &\begin{bmatrix} \tilde{x} \\ w_t \\ v_{t+1} \end{bmatrix}^\top \begin{bmatrix} Q + A^\top P_{t+1} A & A^\top P_{t+1} & \frac{1}{2} A^\top \\ P_{t+1} A & P_{t+1} & \frac{1}{2} I \\ \frac{1}{2} A & \frac{1}{2} I & 0 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ w_t \\ v_{t+1} \end{bmatrix} \\ &- \begin{bmatrix} \tilde{x} \\ w_t \\ v_{t+1} \end{bmatrix}^\top \begin{bmatrix} A^\top P_{t+1} B \\ P_{t+1} B \\ \frac{1}{2} B \end{bmatrix} (R + B^\top P_{t+1} B)^{-1} \begin{bmatrix} A^\top P_{t+1} B \\ P_{t+1} B \\ \frac{1}{2} B \end{bmatrix}^\top \begin{bmatrix} \tilde{x} \\ w_t \\ v_{t+1} \end{bmatrix} + q_{t+1},\end{aligned}$$

which is in the quadratic form $\tilde{V}_t^*(\tilde{x}; \mathbf{w}) = \tilde{x}^\top P_t \tilde{x} + v_t^\top \tilde{x} + q_t$ with the recursion given by

$$
\begin{aligned}
P_t &= Q + A^\top P_{t+1} A - A^\top P_{t+1} B (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A, \\
v_t^\top &= 2 \begin{bmatrix} w_t \\ v_{t+1} \end{bmatrix}^\top \left(\begin{bmatrix} P_{t+1} A \\ \frac{1}{2} A \end{bmatrix} - \begin{bmatrix} P_{t+1} B \\ \frac{1}{2} B \end{bmatrix} \Big(R + B^\top P_{t+1} B \Big)^{-1} B^\top P_{t+1} A \right), \\
q_t &= \begin{bmatrix} w_t \\ v_{t+1} \end{bmatrix}^\top \left(\begin{bmatrix} P_{t+1} & \frac{1}{2} I \\ \frac{1}{2} I & 0 \end{bmatrix} - \begin{bmatrix} P_{t+1} B \\ \frac{1}{2} B \end{bmatrix} \Big(R + B^\top P_{t+1} B \Big)^{-1} \begin{bmatrix} P_{t+1} B \\ \frac{1}{2} B \end{bmatrix}^\top \right) \begin{bmatrix} w_t \\ v_{t+1} \end{bmatrix} + q_{t+1}.
\end{aligned}
\tag{12}
$$

We then proceed to derive an explicit expression for u_t^* in (11) in terms of $\mathbf{w} = w_{1:T}$. For brevity, we use the notations $\Sigma_t := R + B^\top P_{t+1} B$, $\Sigma_t := R + B^\top P_{t+1} B$, $A_{\text{cl},t} := A - BK_t$ and

$$S_t := I - P_{t+1} B \Sigma_t^{-1} B^\top.$$

Hence, u_t^* in (11) is rewritten as

$$u_t^* = -K_t \tilde{x} - \Sigma_t^{-1} B^\top \left(P_{t+1} w_t + \frac{1}{2} v_{t+1} \right). \tag{13}$$

We start with the recursive expression of v_t

$$
\begin{aligned}
\frac{1}{2} v_t =& \left(\begin{bmatrix} P_{t+1} A \\ \frac{1}{2} A \end{bmatrix}^\top - A^\top P_{t+1} B \Sigma_t^{-1} \begin{bmatrix} P_{t+1} B \\ \frac{1}{2} B \end{bmatrix}^\top \right) \begin{bmatrix} w_t \\ v_{t+1} \end{bmatrix} \\
=& A^\top (P_{t+1} - P_{t+1} B \Sigma_t^{-1} B^\top P_{t+1}) w_t + A^\top (I + P_{t+1} B \Sigma_t^{-1} B^\top) v_{t+1} \\
=& A^\top S_t \left(P_{t+1} w_t + \frac{1}{2} v_{t+1} \right).
\end{aligned}
\tag{14}
$$

Substituting (14) into (13) gives

$$
\begin{aligned}
u_t^* =& -K_t \tilde{x} - \Sigma_t^{-1} B^\top \left(P_{t+1} w_t + A^\top S_{t+1} \left(P_{t+2} w_{t+1} + \frac{1}{2} v_{t+2} \right) \right) \\
=& -K_t \tilde{x} - \Sigma_t^{-1} B^\top \Big(P_{t+1} w_t + A^\top S_{t+1} P_{t+2} w_{t+1} \\
& \qquad + A^\top S_{t+1} A^\top S_{t+2} P_{t+3} w_{t+2} + \frac{1}{2} A^\top S_{t+1} A^\top S_{t+2} v_{t+2} \Big) \\
=& -K_t \tilde{x} - \Sigma_t^{-1} B^\top \Big(P_{t+1} w_t + A_{\text{cl},t+1}^\top P_{t+2} w_{t+1} \\
& \qquad + A_{\text{cl},t+1}^\top A_{\text{cl},t+2}^\top P_{t+3} w_{t+2} + \frac{1}{2} A^\top S_{t+1} A^\top S_{t+2} v_{t+2} \Big),
\end{aligned}
$$

where the last equality follows from the identity $A^\top S_t = A_{\text{cl},t}^\top$. By iterative substitution and noting that $v_T = 0$, we eventually obtain

$$
\begin{aligned}
u_t^* &= -K_t \tilde{x} - \Sigma_t^{-1} B^\top \Big(P_{t+1} w_t + A_{\text{cl},t+1}^\top P_{t+2} w_{t+1} \\
&\qquad\qquad + (A_{\text{cl},t+2} A_{\text{cl},t+1})^\top P_{t+3} w_{t+2} + \cdots + (A_{\text{cl},T-1} \cdots A_{\text{cl},t+1})^\top P_T w_{T-1} \Big) \\
&= -K_t \tilde{x} - \sum_{i=t}^{T-1} K_{t \to i} w_i, \qquad K_{t\to i} := \Sigma_t^{-1} B^\top (A_{\text{cl},i} \cdots A_{\text{cl},t+1})^\top P_{i+1}.
\end{aligned}
$$

Lastly, the optimal noncausal policy in Theorem 3.1 follows directly from the change of variables. ■

Appendix B. Structural Properties in Linear Quadratic Control

B.1. Noiseless linear quadratic regulator

We briefly review the optimal controllers for the noiseless LQR problem in both the finite- and infinite-horizon settings, which are key components for our subsequent analysis. See Zhou et al. (1996); Bertsekas (2012) for more details. We begin with the finite-horizon problem:

$$
\begin{aligned}
\min_{u_{1:T}} \quad & \textstyle\sum_{t=1}^T x_t^\top Q x_t + u_t^\top R u_t \\
\text{s.t.} \quad & x_{t+1} = A x_t + B u_t, \;\; x_1 \text{ given.}
\end{aligned}
\tag{15}
$$

Recall the Riccati recursion in Definition 3.1. The solution to (15) is given below.

Proposition B.1 *The optimal control policy for (15) is given by $u_t = -K_t x_t$ for all $t \in [T]$. Moreover, the optimal cost-to-go is quadratic in the state, i.e., for any $t \in [T]$ and x_t,*

$$x_t^\top P_t x_t = \min_{u_{t:T}} \textstyle\sum_{i=t}^T x_i^\top Q x_i + u_i^\top R u_i.$$

We now turn to the infinite-horizon case, where the optimal solution is characterized by the discrete algebraic Riccati equation (DARE).

Proposition B.2 *There exists a unique positive semidefinite matrix $P_\infty \succeq 0$ that solves the DARE:*

$$P_\infty = Q + A^\top P_\infty A - A^\top P_\infty B (R + B^\top P_\infty B)^{-1} B^\top P_\infty A. \tag{16}$$

Given the solution P_∞, we define the following associated steady-state quantities:

$$\Sigma_\infty := R + B^\top P_\infty B, \quad K_\infty := \Sigma_\infty^{-1} B^\top P_\infty A, \quad A_{\text{cl},\infty} := A - BK_\infty.$$

Note that these are intrinsic control-theoretic quantities determined solely by (A, B, Q, R).

The following result describes the infinite-horizon optimal policy and its stability.

Proposition B.3 *As $T \to \infty$, the optimal control policy for (15) converges to the stationary policy $u_t = -K_\infty x_t$. Moreover, the corresponding closed-loop matrix $A_{cl,\infty}$ is stable, i.e., $\rho(A_{cl,\infty}) < 1$.*

B.2. Strong stability and regularity of Riccati-based quantities

We establish some structural properties of key control-theoretic quantities intrinsic to (A, B, Q, R) and the Riccati recursion, which underpin the analysis of the lifted-space optimal offline and online MPC controllers in Theorems 3.1 and 4.1. We start with a quantitative notion of stability.

Definition B.1 (Strong Stability (Cohen et al., 2018)) *A matrix $A-BK$ is said to be (κ, γ)* strongly stable *for some $\kappa > 0$ and $\gamma \in [0, 1)$ if there exist matrices H and L such that*

$$A - BK = HLH^{-1}, \quad \|H\| \cdot \|H^{-1}\| \le \kappa, \quad \|L\| \le \gamma.$$

We may also refer to a gain matrix K as (strongly) stable, meaning that its corresponding closed-loop matrix $A - BK$ is (strongly) stable. Moreover, any stable matrix is strongly stable for some κ and γ.

Lemma B.1 *(Cohen et al., 2018, Lemma B.1.) Suppose that $A-BK$ is stable, i.e., $\rho(A-BK) < 1$. Then there exist constants $\kappa > 0$ and $\gamma \in [0, 1)$ such that $A - BK$ is (κ, γ) strongly stable.*

Below we establish the strong stability of the closed-loop matrix $A_{\mathrm{cl},\infty} := A - BK_\infty$.

Corollary B.1 *There exists some $\kappa_\infty > 0$ and $\tilde{\gamma}_\infty \in [0, 1)$ such that $A_{\mathrm{cl},\infty}$ is $(\kappa_\infty, \tilde{\gamma}_\infty)$ strongly stable. That is, there exist matrices $H_{\mathrm{cl},\infty}$ and $L_{\mathrm{cl},\infty}$ such that*

$$A_{\mathrm{cl},\infty} = H_{\mathrm{cl},\infty} L_{\mathrm{cl},\infty} H_{\mathrm{cl},\infty}^{-1}, \quad \|H_{\mathrm{cl},\infty}\| \cdot \|H_{\mathrm{cl},\infty}^{-1}\| \le \kappa_\infty, \quad \|L_{\mathrm{cl},\infty}\| \le \tilde{\gamma}_\infty.$$

Proof Since $A_{\mathrm{cl},\infty}$ is stable by Proposition B.3, Lemma B.1 implies that it is also strongly stable. ■

In fact, from the proof of Lemma B.1, we can choose $\tilde{\gamma}_\infty := \rho(A_{\mathrm{cl},\infty}) \in [0, 1)$.

Lemma B.2 *For any $i \ge 0$, we have $\|A_{\mathrm{cl},\infty}^i\| \le \kappa_\infty \tilde{\gamma}_\infty^i$.*

Proof By Corollary B.1, it follows that $\|A_{\mathrm{cl},\infty}^i\| = \|H_{\mathrm{cl},\infty} L_{\mathrm{cl},\infty}^i H_{\mathrm{cl},\infty}^{-1}\| \le \kappa_\infty \tilde{\gamma}_\infty^i$. ■

We next recall a classical result from Hager and Horowitz (1976), which guarantees exponential convergence of the Riccati recursion to the DARE solution P_∞. This result requires only $Q \succeq 0$ and detectability of (A, Q), unlike more recent analyses (e.g., Dean et al. (2018); Foster and Simchowitz (2020); Zhang et al. (2021)) that assume $Q \succ 0$ to obtain explicit rates. Since the Koopman lifting yields only $Q \succeq 0$, we rely on this more general guarantee.

Proposition B.4 (Hager and Horowitz (1976)) *Let P_t and K_t be defined as in Definition 3.1. Then there exists a constant $\rho_\infty \in (0, 1)$, independent of the horizon length T, such that for all $t \in [T]$,*

$$\|P_t - P_\infty\| = \mathcal{O}(\rho_\infty^{T-t}) \quad \textit{and} \quad \|K_t - K_\infty\| = \mathcal{O}(\rho_\infty^{T-t}),$$

where the hidden constant in $\mathcal{O}(\cdot)$ depends only on (A, B, Q, R) and not on T.

To facilitate the analysis of time-varying matrices $A_{\mathrm{cl},t}$, we also define the transformed versions $L_{\mathrm{cl},t} := H_{\mathrm{cl},\infty}^{-1} A_{\mathrm{cl},t} H_{\mathrm{cl},\infty}$. We next show that $L_{\mathrm{cl},t}$ remain uniformly bounded by 1 for all but the last Δ_{stab} steps of the horizon T, where Δ_{stab} is independent of T.

Lemma B.3 *Let $\gamma_\infty := \frac{1}{2}(1 + \tilde{\gamma}_\infty)$ and $\Delta_{\mathrm{stab}} := \mathcal{O}(\log(1 - \tilde{\gamma}_\infty)^{-1})$, where the hidden constant is independent of T. Then it holds that $\|L_{\mathrm{cl},t}\| \le \gamma_\infty < 1$ for all $t \le T_{\mathrm{stab}} := T - \Delta_{\mathrm{stab}}$.*

Proof For any $t \in [T]$, we have $\|L_{\text{cl},t}\| \le \|L_{\text{cl},\infty}\| + \|L_{\text{cl},t} - L_{\text{cl},\infty}\| \le \tilde{\gamma}_\infty + \kappa_\infty \|A_{\text{cl},t} - A_{\text{cl},\infty}\| \le \tilde{\gamma}_\infty + \kappa_\infty \|B\| \cdot \|K_t - K_\infty\|$. By Proposition B.4, $\|K_t - K_\infty\| = \mathcal{O}(\rho_\infty^{T-t})$. Thus, there exists a constant $\Delta_{\text{stab}} = \mathcal{O}(\log(1-\tilde{\gamma}_\infty)^{-1})$, independent of T, such that the desired result holds. ■

To facilitate the MPC analysis, we introduce the following Riccati recursion.

Definition B.2 (Riccati recursion for MPC) *Define the backward Riccati recursion as:*

$$\bar{P}_W := Q, \quad \bar{P}_t := Q + A^\top \bar{P}_{t+1} A - A^\top \bar{P}_{t+1} B \bar{\Sigma}_t^{-1} B^\top \bar{P}_{t+1} A, \quad \bar{K}_t := \bar{\Sigma}_t^{-1} B^\top \bar{P}_{t+1} A$$

for $t \in [W-1]$ where we define $\bar{\Sigma}_t := R + B^\top \bar{P}_{t+1} B$.

Note that the above time indices for MPC correspond to those in Definition 3.1 via $t \mapsto T - W + t - 1$. That is, $\bar{K}_t = K_{T-W+t-1}$, $\bar{P}_{t+1} = P_{T-W+t}$, and $\bar{\Sigma}_t = \Sigma_{T-W+t-1}$ for all $t \in [W]$. We also define

$$\bar{A}_{\text{cl},t} := A - B\bar{K}_t, \quad \bar{A}_{\text{cl},1 \to i-t+1} := \bar{A}_{\text{cl},i-t+1} \cdots \bar{A}_{\text{cl},2}$$

for $t \le i$ with $t \in [W]$. Finally, we have the following bound on $\bar{A}_{\text{cl},1} = A_{\text{cl},T-W}$.

Lemma B.4 *For any $W \ge \Delta_{\text{stab}}$, we have $\|\bar{A}_{\text{cl},1}^i\| \le \kappa_\infty \gamma_\infty^i$.*

Proof From Lemma B.3, it follows that for $W \ge \Delta_{\text{stab}}$ we have $\|L_{\text{cl},T-W}\| \le \gamma_\infty$. Therefore, $\|A_{\text{cl},T-W}^i\| = \|H_{\text{cl},\infty} L_{\text{cl},T-W}^i H_{\text{cl},\infty}^{-1}\| \le \kappa_\infty \gamma_\infty^i$. ■

Lemma B.4 establishes a stability property of the MPC closed-loop dynamics and plays a key role in proving the boundedness of the MPC state trajectory in Lemma 5.4.

Appendix C. Proofs of Section 5

C.1. Proof of Lemma 5.1

The proof uses the cost difference lemma (Kakade, 2003; Fazel et al., 2018; Zhang et al., 2021).

Lemma C.1 *For two policies π_1, π_2, the difference of their cost-to-go V_1 and V_2 is given by*

$$V_1^{\pi_2}(x) - V_1^{\pi_1}(x) = \textstyle\sum_{t=1}^T Q_t^{\pi_1}(x_t^{\pi_2}, u_t^{\pi_2}) - V_t^{\pi_1}(x_t^{\pi_2}) \tag{25}$$

where $\{x_t^{\pi_2}, u_t^{\pi_2}\}$ are the trajectory generated by starting at initial state x and imposing policy π_2.

Proof By Lemma C.1, the relative cost between the MPC and offline optimal policy is equal to

$$J_T^{\text{lft}}(\mathbf{u}^{\text{MPC}}; \mathbf{r}) - J_T^{\text{lft}}(\mathbf{u}^*; \mathbf{r}) = \textstyle\sum_{t=1}^T Q_t^*(x_t^{\text{MPC}}, u_t^{\text{MPC}}; \mathbf{r}) - Q_t^*(x_t^{\text{MPC}}, u_t^*; \mathbf{r}).$$

By Definition A.1 and Theorem A.1, $Q_t^*(x_t^{\text{MPC}}, \cdot; \mathbf{r})$ is strongly convex quadratic and u_t^* is its minimizer. Therefore, the first-odrer optimality condition implies that

$$Q_t^*(x_t^{\text{MPC}}, u_t^{\text{MPC}}; \mathbf{r}) - Q_t^*(x_t^{\text{MPC}}, u_t^*; \mathbf{r}) = \|u_t^{\text{MPC}} - u_t^*\|^2_{\nabla_u^2 Q_t^*(x_t^{\text{MPC}}, u_t^*; \mathbf{r})}.$$

Finally, the result follows from $\nabla_u^2 Q_t^*(x_t^{\text{MPC}}, u_t^*; \mathbf{r}) = \Sigma_t$. ■

C.2. Proof of Lemma 5.2

We highlight the key proof idea from Lemma B.3: matrices $L_{\mathrm{cl},t}$ are uniformly bounded by 1 for all but the final Δ_{stab} steps, where Δ_{stab} is independent of the horizon T. Therefore, any bounded product involving at most Δ_{stab} terms near the end of the horizon can be absorbed into the constant in $\mathcal{O}(\cdot)$, ensuring that the overall product still decays exponentially in $t_2 - t_1$.

Proof Since $A_{\mathrm{cl},t} = H_{\mathrm{cl},\infty} L_{\mathrm{cl},t} H_{\mathrm{cl},\infty}^{-1}$, we have $A_{\mathrm{cl},t_1\to t_2} = H_{\mathrm{cl},\infty} L_{\mathrm{cl},t_1\to t_2} H_{\mathrm{cl},\infty}^{-1}$ and therefore

$$\|A_{\mathrm{cl},t_1\to t_2}\| \le \|H_{\mathrm{cl},\infty}\| \cdot \|H_{\mathrm{cl},\infty}^{-1}\| \cdot \|L_{\mathrm{cl},t_1\to t_2}\| \le \kappa_\infty \|L_{\mathrm{cl},t_1\to t_2}\|.$$

It remains to bound $\|L_{\mathrm{cl},t_1\to t_2}\|$ in the following three cases.

Case 1: $t_2 \le T_{\mathrm{stab}}$**.** By Lemma B.3, we have $\|L_{\mathrm{cl},t}\| \le \gamma_\infty < 1$ for all $t \in [t_1+1, t_2]$. Thus

$$\|L_{\mathrm{cl},t_1\to t_2}\| \le \gamma_\infty^{t_2-t_1}.$$

Case 2: $t_1 \le T_{\mathrm{stab}} < t_2$**.** In this case, decompose the product at T_{stab}

$$\begin{aligned}
\|L_{\mathrm{cl},t_1\to t_2}\| &\le \|L_{\mathrm{cl},t_1\to T_{\mathrm{stab}}}\| \cdot \|L_{\mathrm{cl},T_{\mathrm{stab}}\to t_2}\| \\
&\overset{(a)}{\le} \gamma_\infty^{T_{\mathrm{stab}}-t_1} \|L_{\mathrm{cl},T_{\mathrm{stab}}\to t_2}\| \\
&\overset{(b)}{\le} \gamma_\infty^{T_{\mathrm{stab}}-t_1} \gamma_\infty^{t_2-T_{\mathrm{stab}}} (\gamma_\infty^{-1})^{\Delta_{\mathrm{stab}}} \|L_{\mathrm{cl},T_{\mathrm{stab}}\to t_2}\| \\
&\overset{(c)}{=} \mathcal{O}(\gamma_\infty^{t_2-t_1}),
\end{aligned}$$

where (a) follows from Lemma B.3, (b) uses the fact that $t_2 - T_{\mathrm{stab}} \le \Delta_{\mathrm{stab}}$, and (c) absorbs $(\gamma_\infty^{-1})^{\Delta_{\mathrm{stab}}} \|L_{\mathrm{cl},T_{\mathrm{stab}}\to t_2}\|$ into the constant, since the product involves at most Δ_{stab} terms and is uniformly bounded independently of T.

Case 3: $T_{\mathrm{stab}} < t_1 \le t_2$**.** Here we have

$$\begin{aligned}
\|L_{\mathrm{cl},t_1\to t_2}\| &= \gamma_\infty^{t_2-t_1} (\gamma_\infty^{-1})^{t_2-t_1} \|L_{\mathrm{cl},t_1\to t_2}\| \\
&\overset{(a)}{\le} \gamma_\infty^{t_2-t_1} (\gamma_\infty^{-1})^{\Delta_{\mathrm{stab}}} \|L_{\mathrm{cl},t_1\to t_2}\| \\
&\overset{(b)}{=} \mathcal{O}(\gamma_\infty^{t_2-t_1}),
\end{aligned}$$

where (a) is due to $t_2 - t_1 \le \Delta_{\mathrm{stab}}$ and (b) absorbs $(\gamma_\infty^{-1})^{\Delta_{\mathrm{stab}}} \|L_{\mathrm{cl},t_1\to t_2}\|$ into the constant, since this quantity is uniformly bounded by a constant independent of T.

Combining all cases yields the desired bound. ■

C.3. Proof of Proposition 5.1

Proof Let $w_i = A\psi(r_i) - \psi(r_{i+1})$. By Assumption 2.1, $\|w_i\| \le D_w := (\|A\|+1)D_\psi$. Then,

$$\begin{aligned}
\sum_{t=1}^{T-W} \left\| \sum_{i=t+W-1}^{T-1} K_{t\to i} w_i \right\|_{\Sigma_t}^2 &\le \sum_{t=1}^{T-W} \|\Sigma_t\| \cdot \left(\sum_{i=t+W-1}^{T-1} \|K_{t\to i}\| \cdot \|w_i\| \right)^2 \\
&\le D_w^2 \cdot \max_t \|\Sigma_t\| \cdot \sum_{t=1}^{T-W} \left(\sum_{i=t+W-1}^{T-1} \|K_{t\to i}\| \right)^2 \\
&\le D_w^2 \cdot \left(\|R\| + \|B\|^2 \cdot \|P_\infty\| \right) \cdot \sum_{t=1}^{T-W} \left(\sum_{i=t+W-1}^{T-1} \mathcal{O}(\gamma_\infty^{i-t}) \right)^2,
\end{aligned}$$

where the last equality follows from Lemma 5.2. For the inner sum, changing the summation index via $j = i - t + 1$, we obtain

$$\sum_{i=t+W-1}^{T-1} \mathcal{O}(\gamma_\infty^{i-t}) = \mathcal{O}\left(\sum_{j=W}^{T-t} \gamma_\infty^j\right) = \mathcal{O}\left(\sum_{j=W}^{\infty} \gamma_\infty^j\right) = \mathcal{O}(\gamma_\infty^W).$$

Therefore, taking the square and summing over t gives

$$\sum_{t=1}^{T-W} \left\|\sum_{i=t+W-1}^{T-1} K_{t\to i} w_i\right\|_{\Sigma_t}^2 = \mathcal{O}(\gamma_\infty^{2W} T).$$

■

C.4. Auxiliary lemmas for Lemma 5.3

To prove Lemma 5.3, we first introduce two technical lemmas (Lemmas C.2 and C.3) below for bounding the difference between two intermediate matrices defined as follows:

$$X_{t\to i} := A_{\mathrm{cl},t\to i}^\top P_{i+1} = \left(A_{\mathrm{cl},i}\cdots A_{\mathrm{cl},t+1}\right)^\top P_{i+1},$$
$$\bar{X}_{1\to i-t+1} := \bar{A}_{\mathrm{cl},1\to i-t+1}^\top \bar{P}_{i-t+2} = \left(\bar{A}_{\mathrm{cl},i-t+1}\cdots \bar{A}_{\mathrm{cl},2}\right)^\top \bar{P}_{i-t+2}.$$

The first lemma relates the difference of $\bar{X}_{1\to i-t+1} - X_{t\to i}$ to $\bar{P}_{j-t+2} - P_{j+1}$ for $t+1 \le j \le i$.

Lemma C.2 *For $t \in [T-W]$ and $t \le i \le t+W-2$,*

$$\bar{X}_{1\to i-t+1} - X_{t\to i} = A_{\mathrm{cl},t\to i}^\top(\bar{P}_{i-t+2} - P_{i+1})$$
$$+ \sum_{j=t+1}^{i} A_{\mathrm{cl},t\to j}^\top (P_{T-W+j-t+2} - P_{j+1})^\top B(R + B^\top P_{T-W+j-t+2}B)^{-1} B^\top \bar{X}_{j-t+1\to i-t+1}.$$

Proof We first observe the recursive structure

$$X_{t\to i} = A_{\mathrm{cl},t+1}^\top X_{t+1\to i}, \quad \bar{X}_{1\to i-t+1} = \bar{A}_{\mathrm{cl},2}^\top \bar{X}_{2\to i-t+1}.$$

Therefore,

$$\begin{aligned}
&\bar{X}_{1\to i-t+1} - X_{t\to i}\\
=&(\bar{A}_{\mathrm{cl},2}^\top - A_{\mathrm{cl},t+1}^\top)\bar{X}_{2\to i-t+1} + A_{\mathrm{cl},t+1}^\top(\bar{X}_{2\to i-t+1} - X_{t+1\to i})\\
=&(\bar{K}_2 - K_{t+1})^\top B^\top \bar{X}_{2\to i-t+1} + A_{\mathrm{cl},t+1}^\top(\bar{X}_{2\to i-t+1} - X_{t+1\to i}).
\end{aligned}$$

By iterative expansion using the above recursive relation, we obtain

$$\begin{aligned}
&\bar{X}_{1\to i-t+1} - X_{t\to i}\\
=&(\bar{K}_2 - K_{t+1})^\top B^\top \bar{X}_{2\to i-t+1} + A_{\mathrm{cl},t+1}^\top(\bar{X}_{2\to i-t+1} - X_{t+1\to i})\\
=&(\bar{K}_2 - K_{t+1})^\top B^\top \bar{X}_{2\to i-t+1}\\
&+ A_{\mathrm{cl},t+1}^\top(\bar{K}_3 - K_{t+2})^\top B^\top \bar{X}_{3\to i-t+1} + A_{\mathrm{cl},t+1}^\top A_{\mathrm{cl},t+2}^\top(\bar{X}_{3\to i-t+1} - X_{t+2\to i})\\
=&A_{\mathrm{cl},t\to i}^\top(\bar{X}_{i-t+1\to i-t+1} - X_{i\to i}) + \sum_{j=t+1}^{i} A_{\mathrm{cl},t\to j-1}^\top(\bar{K}_{j-t+1} - K_j)^\top B^\top \bar{X}_{j-t+1\to i-t+1}\\
=&A_{\mathrm{cl},t\to i}^\top(\bar{P}_{i-t+2} - P_{i+1}) + \sum_{j=t+1}^{i} A_{\mathrm{cl},t\to j-1}^\top(\bar{K}_{j-t+1} - K_j)^\top B^\top \bar{X}_{j-t+1\to i-t+1} \qquad (17)
\end{aligned}$$

where the last equality uses the convention that $A_{\mathrm{cl},t\to t} = I$. We also have

$$
\begin{aligned}
&\bar{K}_{j-t+1} - K_j = K_{T-W+j-t+1} - K_j \\
=&(R + B^\top P_{T-W+j-t+2}B)^{-1}B^\top P_{T-W+j-t+2}A - (R + B^\top P_{j+1}B)^{-1}B^\top P_{j+1}A \\
=&(R + B^\top P_{T-W+j-t+2}B)^{-1}B^\top (P_{T-W+j-t+2} - P_{j+1})A \\
&+ \left((R + B^\top P_{T-W+j-t+2}B)^{-1} - (R + B^\top P_{j+1}B)^{-1}\right) B^\top P_{j+1}A \\
=&(R + B^\top P_{T-W+j-t+2}B)^{-1}B^\top (P_{T-W+j-t+2} - P_{j+1})A \\
&- (R + B^\top P_{T-W+j-t+2}B)^{-1}B^T (P_{T-W+j-t+2} - P_{j+1})B(R + B^\top P_{j+1}B)^{-1}B^\top P_{j+1}A \\
=&(R + B^\top P_{T-W+j-t+2}B)^{-1}B^\top (P_{T-W+j-t+2} - P_{j+1})(A - BK_j) \qquad (18)
\end{aligned}
$$

where the third equality uses the identity

$$M_1^{-1} - M_2^{-1} = M_1^{-1}(M_2 - M_1)M_2^{-1}$$

for two invertible matrices M_1 and M_2. Substituting (18) into (17) then completes the proof. ■

Based on Lemma C.2, Lemma 5.2 and Proposition B.4, we obtain the following bound.

Lemma C.3 *For $t \in [T-W]$ and $t \le i \le t+W-2$, we have*

$$\|X_{t\to i} - \bar{X}_{1\to i-t+1}\| = \mathcal{O}(\gamma_\infty^{i-t}\rho_\infty^{W-i+t}).$$

Proof From the expression in Lemma C.2, we have

$$
\begin{aligned}
&\|\bar{X}_{1\to i-t+1} - X_{t\to i}\| \le \|A_{\mathrm{cl},t\to i}^\top\| \cdot \|\bar{P}_{i-t+2} - P_{i+1}\| \\
&+ \sum_{j=t+1}^{i} \|A_{\mathrm{cl},t\to j}^\top\| \cdot \|P_{T-W+j-t+2} - P_{j+1}\| \cdot \|B(R+B^\top P_{T-W+j-t+2}B)^{-1}B^\top\| \cdot \|\bar{X}_{j-t+1\to i-t+1}\|.
\end{aligned}
$$

First by Lemma 5.2,

$$\|\bar{X}_{j-t+1\to i-t+1}\| = \|\left(\bar{A}_{\mathrm{cl},i-t+1}\cdots\bar{A}_{\mathrm{cl},j-t+2}\right)^\top \bar{P}_{i-t+2}\| = \mathcal{O}(\gamma_\infty^{i-j}). \qquad (19)$$

Also by Proposition B.4, we have for $t \in [T-W]$ and $t+1 \le j < t+W-1$,

$$
\begin{aligned}
\|P_{T-W+j-t+2} - P_{j+1}\| &\le \|P_{T-W+j-t+2} - P_\infty\| + \|P_\infty - P_{j+1}\| \\
&= \mathcal{O}(\rho_\infty^{W-j+t-2}) + \mathcal{O}(\rho_\infty^{T-j-1}) \\
&= \mathcal{O}(\rho^{\min\{W-j+t-2, T-j-1\}}) \\
&= \mathcal{O}(\rho_\infty^{W-j+t}). \qquad (20)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|\bar{X}_{1\to i-t+1} - X_{t\to i}\| \le\ & \mathcal{O}(\gamma_\infty^{i-t}) \cdot \mathcal{O}(\rho_\infty^{W-i+t}) \\
&+ \textstyle\sum_{j=t+1}^{i} \mathcal{O}(\gamma_\infty^{j-t}) \cdot \mathcal{O}(\rho_\infty^{W-j+t}) \cdot \|B\|^2 \cdot \|R^{-1}\| \cdot \mathcal{O}(\gamma_\infty^{i-j}) \\
=\ & \mathcal{O}(\gamma_\infty^{i-t}\rho_\infty^{W-i+t}) + \mathcal{O}\left(\gamma_\infty^{i-t} \cdot \textstyle\sum_{j=t+1}^{i} \rho_\infty^{W-j+t}\right) \\
=\ & \mathcal{O}(\gamma_\infty^{i-t}\rho_\infty^{W-i+t})
\end{aligned}
$$

where the first inequality follows from (19) and (20). ■

C.5. Proof of Lemma 5.3

Proof The proof is based on Lemma C.3. Recall the feedforward gains

$$K_{t\to i} = \Sigma_t^{-1} B^\top (A_{\text{cl},i} \cdots A_{\text{cl},t+1})^\top P_{i+1},$$
$$\bar{K}_{1\to i-t+1} = \bar{\Sigma}_1^{-1} B^\top (\bar{A}_{\text{cl},i-t+1} \cdots \bar{A}_{\text{cl},2})^\top \bar{P}_{i-t+2}$$

for $t \in [T-W]$ and $t \le i \le t+W-2$. We can rewrite their difference as

$$\begin{aligned} K_{t\to i} - \bar{K}_{1\to i-t+1} &= \Sigma_t^{-1} B^\top X_{t\to i} - \bar{\Sigma}_1^{-1} B^\top \bar{X}_{1\to i-t+1} \\ &= \Sigma_t^{-1}(B^\top X_{t\to i} - B^\top \bar{X}_{1\to i-t+1}) + (\Sigma_t^{-1} - \bar{\Sigma}_1^{-1}) B^\top \bar{X}_{1\to i-t+1}. \end{aligned} \tag{21}$$

Using Lemma C.3, the first term in (21) is bounded as

$$\begin{aligned} \|\Sigma_t^{-1}(B^\top X_{t\to i} - B^\top \bar{X}_{1\to i-t+1})\| &\le \|R^{-1}\| \cdot \|B\| \cdot \|X_{t\to i} - \bar{X}_{1\to i-t+1}\| \\ &= \mathcal{O}(\gamma_\infty^{i-t} \rho_\infty^{W-i+t}) \end{aligned} \tag{22}$$

For the second term in (21), by the identity $\Sigma_t^{-1} - \bar{\Sigma}_1^{-1} = \Sigma_t^{-1}(\bar{\Sigma}_1 - \Sigma_t)\bar{\Sigma}_1^{-1}$,

$$\begin{aligned} \|(\Sigma_t^{-1} - \bar{\Sigma}_1^{-1}) B^\top \bar{X}_{1\to i-t+1}\| &\le \|\Sigma_t^{-1}\| \cdot \|\bar{\Sigma}_1 - \Sigma_t\| \cdot \|\bar{\Sigma}_1^{-1}\| \cdot \|B\| \cdot \|\bar{X}_{1\to i-t+1}\| \\ &\le \|B\|^3 \cdot \|R^{-1}\|^2 \cdot \|\bar{P}_2 - P_{t+1}\| \cdot \|\bar{X}_{1\to i-t+1}\| \\ &= \mathcal{O}(\rho_\infty^W \gamma_\infty^{i-t+1}), \end{aligned} \tag{23}$$

where the last equality follows by arguments analogous to those in (20) and (19). Combining (22) and (23), and applying Lemma C.3, we obtain

$$\|K_{t\to i} - \bar{K}_{1\to i-t+1}\| = \mathcal{O}(\gamma_\infty^{i-t} \rho_\infty^{W-i+t}) + \mathcal{O}(\rho_\infty^W \gamma_\infty^{i-t+1}) = \mathcal{O}(\gamma_\infty^{i-t} \rho_\infty^{W-i+t}).$$

Finally, one can verify that $K_t - \bar{K}_1 = (K_{t\to t} - \bar{K}_{1\to 1})A$ and thus

$$\|K_t - \bar{K}_1\| \le \|K_{t\to t} - \bar{K}_{1\to 1}\| \cdot \|A\| = \mathcal{O}(\rho_\infty^W).$$

■

C.6. Proof of Lemma 5.4

To prove Lemma 5.4, we first introduce a lemma.

Lemma C.4 *Consider the MPC policy given in Theorem 4.1. For $t \in [T-W]$, we have*

$$u_t^{\text{MPC}} = -\bar{K}_1 \tilde{x}_t - q_t(\mathbf{w}), \quad q_t(\mathbf{w}) := \textstyle\sum_{i=t}^{t+W-2} \bar{K}_{1\to i-t+1} w_i,$$

where $\tilde{x}_t = x_t - \psi(r_t)$ and $w_t = A\psi(r_t) - \psi(r_{t+1})$. Then,

(i) *The state trajectory $\tilde{x}_t$ for $t \in [T-W]$ can be expressed as*

$$\tilde{x}_{t+1}(\mathbf{w}) = \textstyle\sum_{i=1}^t \bar{A}_{\text{cl},1}^{t-i} w_i - \sum_{i=1}^t \bar{A}_{\text{cl},1}^{t-i} B q_i(\mathbf{w}).$$

(ii) *The feedforward term is uniformly bounded for all $t \in [T-W]$ as*

$$\|q_t(\mathbf{w})\| \le D_q := \|R^{-1}\| \cdot \|B\| \cdot \|P_\infty\| \cdot D_w \cdot (1-\gamma_\infty)^{-1}.$$

Proof The first statement directly follows from recursive substitution. To bound $\|q_t(\mathbf{w})\|$, recall that

$$\begin{aligned} \bar{K}_{1\to i-t+1} &= K_{T-W\to T-W+i-t+1} \\ &= \Sigma_{T-W}^{-1} B^\top (A_{\text{cl},T-W+i-t+1} \cdots A_{\text{cl},T-W+1})^\top P_{T-W+i-t+2} \end{aligned}$$

defined for $t \le i \le t+W-2$ with $t \in [T-W]$. Therefore, by Lemma 5.2,

$$\begin{aligned}\|q_t(\mathbf{w})\| &= \left\|\sum_{i=t}^{t+W-2} \bar{K}_{1\to i-t+1} w_i\right\| \\ &\le \|\Sigma_{T-W}^{-1}\| \cdot \|B\| \cdot \max_i \|P_{T-W+i-t+2}\| \cdot D_w \cdot \sum_{i=t}^{t+W-2} \|A_{\mathrm{cl},T-W\to T-W+i-t+1}\| \\ &\le \|R^{-1}\| \cdot \|B\| \cdot \|P_\infty\| \cdot D_w \cdot \sum_{i=t}^{t+W-2} \mathcal{O}(\gamma_\infty^{i-t+1}) \le D_q.\end{aligned}$$

■

Proof We bound the norm of the lifted MPC trajectory as follows:

$$\begin{aligned}\|\tilde{x}_{t+1}(\mathbf{w})\| &\overset{(a)}{=} \left\|\sum_{i=1}^{t} \bar{A}_{\mathrm{cl},1}^{t-i} w_i - \sum_{i=1}^{t} \bar{A}_{\mathrm{cl},1}^{t-i} B q_i(\mathbf{w})\right\| \\ &\overset{(b)}{\le} \kappa_\infty \left(1 + \|A\| + \max_i \|q_i(\mathbf{w})\|\right) \cdot \sum_{i=1}^{t} \gamma_\infty^{t-i} \\ &\overset{(c)}{\le} \kappa_\infty (1 + \|A\| + D_q) \cdot (1-\gamma_\infty)^{-1} := D_x,\end{aligned}$$

where (a) follows from Lemma C.4, (b) applies Lemma B.4, which ensures that for any $W \ge \Delta_{\mathrm{stab}}$ the decay bound $\|\bar{A}_{\mathrm{cl},1}^{t-i}\| \le \kappa_\infty \gamma_\infty^{t-i}$ holds, and (c) uses again Lemma C.4. ■

C.7. Proof of Proposition 5.2

Proof We first bound the feedback gain deviation. From Lemma 5.4, we have

$$\sum_{t=1}^{T-W} \left\|(K_t - \bar{K}_1)(x_t - \psi(r_t))\right\|_{\Sigma_t}^2 \le D_x^2 \cdot \max_t \|\Sigma_t\| \cdot \sum_{t=1}^{T-W} \|K_t - \bar{K}_1\|^2 = \mathcal{O}(\rho_\infty^{2W} T),$$

where we use the convergence bound $\|K_t - \bar{K}_1\| = \mathcal{O}(\rho_\infty^W)$ from Proposition B.4.

We next bound the feedforward gain deviation. Since $\|w_i\| \le D_w$, we get

$$\begin{aligned}&\sum_{t=1}^{T} \left\|\sum_{i=t}^{t+W-2} (K_{t\to i} - \bar{K}_{1\to i-t+1}) w_i\right\|_{\Sigma_t}^2 \\ \le& D_w^2 \cdot \max_t \|\Sigma_t\| \cdot \sum_{t=1}^{T} \left(\sum_{i=t}^{t+W-2} \|K_{t\to i} - \bar{K}_{1\to i-t+1}\|\right)^2 \\ \le& D_w^2 \cdot \max_t \|\Sigma_t\| \cdot \sum_{t=1}^{T} \left(\sum_{i=t}^{t+W-2} \mathcal{O}(\gamma_\infty^{i-t} \rho_\infty^{W-i+t})\right)^2 \\ \le& \mathcal{O}(1) \cdot \sum_{t=1}^{T} \left(\sum_{j=0}^{W-2} \mathcal{O}(\gamma_\infty^{j} \rho_\infty^{W-j})\right)^2 \\ \le& \mathcal{O}(1) \cdot \sum_{t=1}^{T} \left(\sum_{j=0}^{W-2} \mathcal{O}(\lambda_\infty^{W})\right)^2 = \mathcal{O}(W^2 \lambda_\infty^{2W} T),\end{aligned}$$

where $\lambda_\infty := \max\{\gamma_\infty, \rho_\infty\}$, and we use the bound from Lemma 5.3 in the second inequality. Note that when $\gamma_\infty < \rho_\infty$, we have $\sum_{j=0}^{W-2} \mathcal{O}\left((\gamma_\infty/\rho_\infty)^j\right) = \mathcal{O}\left((1-\gamma_\infty/\rho_\infty)^{-1}\right) = \mathcal{O}(1)$ and hence the bound can be improved to $\mathcal{O}(\rho_\infty^{2W} T)$.

■

Appendix D. Additional Experiments: Two-wheeled Robots

We here illustrate that the proposed online controller (DDPC) can be adapted to work for nonlinear systems that are not Koopman-linearizable. We consider a two-wheeled mobile robot with nonlinear kinematic dynamics (Li et al., 2019):

$$\begin{aligned} z_{x,t+1} &= z_{x,t} + \Delta t \cdot \cos(z_{\delta,t}) \cdot v_t, \\ z_{y,t+1} &= z_{y,t} + \Delta t \cdot \sin(z_{\delta,t}) \cdot v_t, \\ z_{\delta,t+1} &= z_{\delta,t} + \Delta t \cdot w_t, \end{aligned}$$

where (z_x, z_y) denotes the robot's position, v and w represent tangential and angular velocities, z_δ is the heading angle (relative to the x-axis), and $\Delta t = 0.025\,\mathrm{s}$. The objective is to track a heart-shaped reference trajectory $(\mathbf{r}_x, \mathbf{r}_y, \mathbf{r}_\delta)$ given below while minimizing control effort:

$$\begin{aligned} r_{x,t} &= 16\sin^3(t-6), \\ r_{y,t} &= 13\cos(t) - 5\cos(2t-12) - 2\cos(3t-18) - \cos(4t-24), \\ r_{\delta,t} &= \arctan\left(\frac{r_{y,t+1} - r_{y,t}}{r_{x,t+1} - r_{x,t}}\right). \end{aligned}$$

We set the cost matrices $Q_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ and $R = 1.3 \times 10^{-3} I_2$. As this nonlinear model is generally not Koopman-linearizable, we introduce a regularization term for implicit model selection and a slack variable for numerical stability (Dörfler et al., 2022). The regularized DDPC then becomes

$$\begin{aligned} \min_{u_{1:W|t}, z_{1:W|t}, g, \sigma_z} \quad & \textstyle\sum_{i=1}^{W} \|z_{i|t} - r_{t+i-1}\|_{Q_z}^2 + \|u_{i|t}\|_R^2 + \lambda_g \|g\|_1 + \lambda_z \|\sigma_z\|_2^2 \\ \text{s.t.} \quad & H_{\mathrm{d}} g = [\mathbf{u}_{\mathrm{ini}}^\mathsf{T}, u_{1:W|t}^\mathsf{T}, (\mathbf{z}_{\mathrm{ini}} + \sigma_z)^\mathsf{T}, z_{1:W|t}^\mathsf{T}]^\mathsf{T}. \end{aligned} \tag{Reg-DDPC}$$

When the data library H_{d} is collected from a nonlinear system that is not Koopman-linearizable, it typically has full row rank, rendering the equality constraint ineffective. To address this, we employ the l_1-norm regularizer $\|g\|_1$ to encourage the optimization to select more "important" trajectories for prediction. This serves as a convex relaxation of the l_0-norm to reduce model complexity.

In addition, modeling error may cause the initial trajectory to fall outside the range space of H_{d}. To maintain feasibility, we introduce a slack variable σ_z and penalize large deviations from the initial trajectory through a relatively large weight λ_z. In our simulation, we set $\lambda_g = 2$ and $\lambda_z = 3 \times 10^6$. Also note that for general nonlinear systems, the collected data captures only the local dynamics over the data collection region. Therefore, the data-driven constraint in (Reg-DDPC) can be viewed as a local linear approximation. Thus, to improve prediction accuracy and tracking performance, our strategy is to switch the data library based on the robot's current state.

Since the nonlinearity of the two-wheeled robot mainly arises from the coupled terms $\sin(z_\delta) \cdot v$ and $\cos(z_\delta) \cdot v$ in the dynamics of z_x and z_y, the switching strategy is based on the robot's orientation. Specifically, during offline data collection, we collect 4 trajectories of length 1500 with initial states $(0, 0, \frac{\pi}{4})$, $(0, 0, \frac{3\pi}{4})$, $(0, 0, \frac{5\pi}{4})$ and $(0, 0, \frac{7\pi}{4})$. The input signals v and w are uniformly sampled from $[10, 20]$ and $[-\frac{\pi}{6}, \frac{\pi}{6}]$, respectively. For each prediction horizon W, we construct 4 corresponding data libraries H_{d} with $T_{\mathrm{ini}} = 5$, each approximately representing the local behavior of the nonlinear system within orientation intervals $[0, \frac{\pi}{2})$, $[\frac{\pi}{2}, \pi)$, $[\pi, \frac{3\pi}{2})$, and $[\frac{3\pi}{2}, 2\pi)$. During online tracking, the data library associated with the robot's current orientation is selected to improve tracking performance.

The results of tracking the heart-shaped curve over two cycles with prediction horizons $W = 6$, 9, 12 are shown in Figure 2. We observe that the robot tracks the reference trajectory relatively well in smooth segments but exhibits noticeable deviation or oscillation at sharp turns. As we can see, a longer prediction horizon enhances tracking accuracy in smooth regions by reducing drift and allows the robot to recover more quickly after deviating at sharp corners.

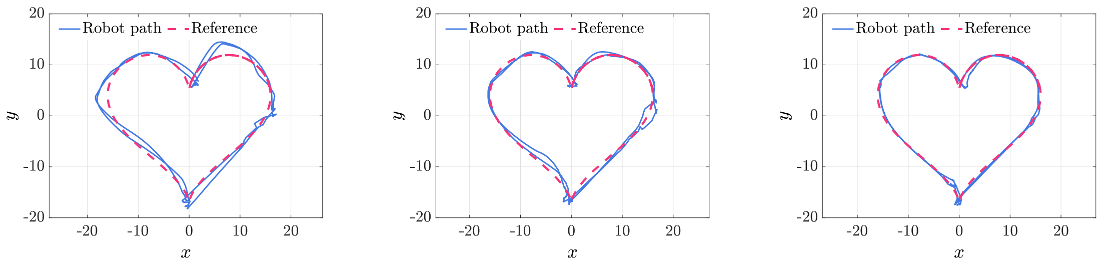

Figure 2: Tracking performance of the two-wheel robot under varying prediction horizons: $W = 6$ (left), $W = 9$ (middle), and $W = 12$ (right). The red curve denotes the reference trajectory, and the blue curve shows the robot's actual path.