Perturbation-based Regret Analysis of Predictive Control in Linear Time Varying Systems

Yiheng Lin
California Institute of Technology
Pasadena, CA, USA
yihengl@caltech.edu

Yang Hu
Tsinghua University
Beijing, China
huy18@mails.tsinghua.edu.cn

Guanya Shi
California Institute of Technology
Pasadena, CA, USA
gshi@caltech.edu

Haoyuan Sun
California Institute of Technology
Pasadena, CA, USA
hsun2@caltech.edu

Guannan Qu
California Institute of Technology
Pasadena, CA, USA
gqu@caltech.edu

Adam Wierman
California Institute of Technology
Pasadena, CA, USA
adamw@caltech.edu

Abstract

We study predictive control in a setting where the dynamics are time-varying and linear, and the costs are time-varying and well-conditioned. At each time step, the controller receives the exact predictions of costs, dynamics, and disturbances for the future $k$ time steps. We show that when the prediction window $k$ is sufficiently large, predictive control is input-to-state stable and achieves a dynamic regret of $O(\lambda^k T)$, where $\lambda < 1$ is a positive constant. This is the first dynamic regret bound on the predictive control of linear time-varying systems. We also show a variation of predictive control obtains the first competitive bound for the control of linear time-varying systems: $1 + O(\lambda^k)$. Our results are derived using a novel proof framework based on a perturbation bound that characterizes how a small change to the system parameters impacts the optimal trajectory.

1 Introduction

We study the problem of predictive control in a linear time-varying (LTV) system, where the dynamics is given by $x_{t+1} = A_t x_t + B_t u_t + w_t$. Here, $x_t$ is the state, $u_t$ is the control action, and $w_t$ is the disturbance or exogenous input. At each time step $t$, the online controller incurs a time-varying state cost $f_t(x_t)$ and control cost $c_t(u_{t-1})$, and then decides its next control action $u_t$. In deciding $u_t$ the controller makes use of predictions of the next $k$ future disturbances, cost functions, and dynamical matrices, and seeks to minimize its total cost on a finite horizon $T$. Our main results bound the dynamic regret and competitive ratio of predictive controllers in this LTV setting.

Recently, a growing literature has sought to design controllers that achieve learning guarantees such as static regret \footnote{Yiheng Lin, Yang Hu, Haoyuan Sun, Guanya Shi, and Guannan Qu contributed equally to this work.}, dynamic regret \footnote{This work was supported by NSF grants CNS-2106403, NGSDI-2105648, and AitF-1637598, with additional support from Amazon AWS, PIMCO, and the Resnick Sustainability Institute. Yiheng Lin was supported by Kortschak Scholars program.}, and competitive ratio \footnote{35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.}. The most relevant
We develop a novel analysis framework based on a perturbation approach. Specifically, instead of solving for the optimal states and control actions like previous analyses in the LTI setting with quadratic costs, e.g., [4, 5, 7], we bound how much impact an perturbation to the system parameters can have on the optimal solution. This type of perturbation bound (Theorem 3.3) can be shown even when the optimal trajectory cannot be written down explicitly, which allows it to be applied in LTV systems with well-conditioned costs. Then, we utilize this perturbation bound to establish results on dynamic regret and the competitive ratio. In addition, we want to emphasize that the perturbation approach we develop is highly modular and extendable. For instance, if a stronger perturbation bound for some specific class of dynamics and/or cost functions can be shown, the dynamic regret of the predictive controller will improve. Similarly, to further generalize the problem setting (e.g., to include additional constraints), one only needs to establish the corresponding perturbation bounds and the competitive ratio of MPC-style policies have been limited. Despite recent work providing such guarantees in the context of LTI systems with quadratic costs, e.g., [4, 5, 7], the derivation of regret and competitive ratio results for MPC in LTV systems remains open.

Contributions. We provide the first regret and competitive ratio results for a controller in LTV systems with time-varying costs. Specifically, we show that an MPC-style predictive control algorithm (Algorithm 1) achieves a dynamic regret that decays exponentially with respect to the length of prediction window \(k\) in the LTV system (Theorem 4.2): \(O(\lambda^k T)\), where the decay rate \(\lambda\) is a positive constant less than 1. This almost matches the exponential lower bound for improvement from predictions in the LTI setting shown in [3] in the sense that, to achieve any target regret level, the required length of prediction \(k\) shown by our bound differs from the theoretical lower bound by at most a constant factor. With a variation of predictive control (Algorithm 2), we also show the first competitive bound in LTV systems with time-varying well-conditioned costs (Theorem 4.3): \(1 + O(\lambda^k)\), where the decay rate \(\lambda\) is identical with the one in the regret bound.

We develop a novel analysis framework based on a perturbation approach. Specifically, instead of solving for the optimal states and control actions like previous analyses in the LTI setting with quadratic costs, e.g., [4, 5, 7], we bound how much impact an perturbation to the system parameters can have on the optimal solution. This type of perturbation bound (Theorem 3.3) can be shown even when the optimal trajectory cannot be written down explicitly, which allows it to be applied in LTV systems with well-conditioned costs. Then, we utilize this perturbation bound to establish results on dynamic regret and the competitive ratio. In addition, we want to emphasize that the perturbation approach we develop is highly modular and extendable. For instance, if a stronger perturbation bound for some specific class of dynamics and/or cost functions can be shown, the dynamic regret of the predictive controller will improve. Similarly, to further generalize the problem setting (e.g., to include additional constraints), one only needs to establish the corresponding perturbation bounds and the regret result will follow.

Another important component of the proof is a novel reduction between LTV control and online optimization. Connections between online optimization and control have received increasing attention in recent years, e.g., [1, 5, 14, 15]. Existing reductions rely on the canonical form, which does not apply to LTV systems, and/or formulations of online optimization with memory of multiple prior time steps, which makes the online problem more challenging. The reduction we present here relies
We assume that the algorithm has access to the exact predictions of disturbances, cost functions and disturbances, namely
\[
\min_{x_0,T, u_0, u_{-1}} \sum_{t=1}^{T} (f_t(x_t) + c_t(u_{t-1}))
\]
subject to
\[
x_t = A_t x_{t-1} + B_t u_{t-1} + w_{t-1}, t = 1, \ldots, T,
\]
where \(x_t \in \mathbb{R}^n\), \(u_t \in \mathbb{R}^m\), and \(w_t \in \mathbb{R}^n\) respectively denote the state, the control action, and the disturbance of the system at time steps \(t = 1, \ldots, T\), and \(x(0) \in \mathbb{R}^n\) is a given initial state. By convention, the hitting cost function \(f_t : \mathbb{R}^n \to \mathbb{R}\) and control cost function \(c_t : \mathbb{R}^m \to \mathbb{R}_+\) are assumed to be time-varying and well-conditioned.

2 Background and Setting

We consider a finite-horizon discrete-time online control problem with linear time-varying (LTV) dynamics, time-varying costs, and disturbances, namely
\[
\min_{x_0,T, u_0, u_{-1}} \sum_{t=1}^{T} (f_t(x_t) + c_t(u_{t-1}))
\]
subject to
\[
x_t = A_t x_{t-1} + B_t u_{t-1} + w_{t-1}, t = 1, \ldots, T,
\]
where \(x_t \in \mathbb{R}^n\), \(u_t \in \mathbb{R}^m\), and \(w_t \in \mathbb{R}^n\) respectively denote the state, the control action, and the disturbance of the system at time steps \(t = 1, \ldots, T\), and \(x(0) \in \mathbb{R}^n\) is a given initial state. By convention, the hitting cost function \(f_t : \mathbb{R}^n \to \mathbb{R}\) and control cost function \(c_t : \mathbb{R}^m \to \mathbb{R}_+\) are assumed to be time-varying and well-conditioned. Define the tuple \(\theta_t := (A_t, B_t, w_t, f_{t+1}, c_{t+1})\).

In the classical setting where no predictions are available, after observing state \(x_t\) at time step \(t\), the algorithm needs to decide the control action \(u_t\) before observing \(\theta_t\), which is an unknown random disturbance input. We use the following event sequence to describe this ordering:
\[
 x_0, u_0, \theta_0, x_1, u_1, \theta_1, x_2, \ldots, x_{T-1}, u_{T-1}, \theta_{T-1}, x_T.
\]
We assume that the algorithm has access to the exact predictions of disturbances, cost functions and dynamical matrices in the future \(k\) time steps (which are time-varying); i.e., the event sequence is
\[
 x_0, \theta_0, \theta_1, \ldots, \theta_{k-1}, u_0, \theta_k, u_1, \theta_{k+1}, \ldots, u_{T-k-1}, \theta_{T-1}, u_{T-k}, u_{T-k+1}, \ldots, u_{T-1}.
\]
Here we assume all predictions are exact, and leave the case of inexact predictions for future work. This prediction model has been used in previous works like [4, 24, 26], and is available in many real-world applications such as disturbance estimation in robotics and frequency regulation in power grids. The availability is due to the fact that, in such scenarios as mentioned above, experiments or observations on the dynamics can be conducted repeatedly and consistently, which makes it feasible to train a good predictor based on the data collected from repeated trials.

2.1 Assumptions

As is standard in studies of regret and competitive ratio in linear control problems, we assume the cost functions are well-conditioned.

**Assumption 2.1 (Well-conditioned Costs).** The cost functions satisfy the following constraints:

1. \(f_t(\cdot)\) is \(m_f\)-strongly convex for \(t = 1, \ldots, T\), and \(\ell_f\)-strongly smooth for \(t = 1, \ldots, T - 1\).
2. \(c_t(\cdot)\) is both \(m_c\)-strongly convex and \(\ell_c\)-strongly smooth for \(t = 1, \ldots, T\).
3. \(f_t(\cdot)\) and \(c_t(\cdot)\) are twice continuously differentiable for \(t = 1, \ldots, T\).
4. \(f_t(\cdot)\) and \(c_t(\cdot)\) are non-negative, and \(f_t(0) = c_t(0) = 0\) for \(t = 1, \ldots, T\).

Note that assumptions (1) through (3) are quite common [3, 5, 14, 25, 27]. Assumption (4) is less common, but can be satisfied via re-parameterization without loss of generality. Specifically, when the minimizers of state cost \(f_t\) and control cost \(c_t\) are nonzero, we perform the transformation
\[
 x' = x - \arg\min_x f_t(x), u' = u - \arg\min_u c_{t+1}(u),
\]
\[
 w' = w + A_t \arg\min_x f_t(x) + B_t \arg\min_u c_{t+1}(u).
\]
The intuition of this transformation is that, when the minimizer of the cost function for the next step is known, we can always perform a translation in the state and control space to align the minimizer
Furthermore, there exists a positive constant $\varepsilon$. The dynamical system is called controllable if there exists a constant $\varepsilon$ such that

$$\text{Assumption 2 in } [28],$$

where $\varepsilon$ is assumed.

Additionally, we need to assume the dynamics are $\varepsilon$-uniform controllability. It is a natural generalization of its counterpart for LTI systems (see Assumption 2 in [28]). This can be generalized to parallel assumptions for various systems.

**Definition 2.1.** For a dynamical system with linear time-varying dynamics $x_t = A_{t-1}x_{t-1} + B_{t-1}u_{t-1} + w_{t-1}$, $t = 1, \ldots, T$, the transition matrix $\Phi(t_2, t_1) \in \mathbb{R}^{n \times n}$ (from time step $t_1$ to $t_2$) is defined as

$$\Phi(t_2, t_1) := \begin{cases} A_{t_2-1}A_{t_2-2} \cdots A_{t_1} & \text{if } t_2 > t_1 \\ I & \text{if } t_2 \leq t_1 \end{cases}$$

and the controllability matrix $M(t, p) \in \mathbb{R}^{n \times (mp)}$ is defined as

$$M(t, p) := \left[ \Phi(t + p, t + 1)B_1, \Phi(t + p, t + 2)B_{t+1}, \ldots, \Phi(t + p, t + p)B_{t+p} \right].$$

The dynamical system is called controllable if there exists a constant $d \in \mathbb{Z}_+$ such that the controllability matrix $M(t, d)$ is of full row rank for any $t = 1, \ldots, T - d$. The smallest constant $d$ with such property is called the controllability index of the system.

Given the above definition, we can state the key assumption necessary for the analysis of LTV systems. We use a slightly stronger assumption than being merely controllable, which we refer to as $(d, \sigma)$-uniform controllability. It is a natural generalization of its counterpart for LTI systems (see Assumption 2 in [28], where $(d, \sigma)$ is instead named as $(\ell, v)$).

**Assumption 2.2.** There exists positive constants $a$, $b$, and $b'$, such that

$$\|A_t\| \leq a, \quad \|B_t\| \leq b, \quad \text{and } \|B_t^\dagger\| \leq b'$$

hold for all time steps $t = 0, \ldots, T - 1$, where $B_t^\dagger$ denotes the Moore–Penrose inverse of matrix $B_t$.

Furthermore, there exists a positive constant $\sigma$ such that

$$\sigma_{\min}(M(t, \sigma)) \geq \sigma$$

holds for all time steps $t = 0, \ldots, T - d$, where $d$ denotes the controllability index.

Note that Assumption 2.2 implies $\sigma_{\min}(M(t, \sigma)) \geq \sigma$ for all $p \geq d$ because appending more columns to a matrix with full row rank will not reduce its minimum singular value.

The LTV setting we consider is more general than the settings which existing results on regret and competitive ratio have assumed [15, 18, 19, 27]. We highlight the implications of this general setting for enabling applications in the following examples.

**Example 2.1** (Trajectory tracking in LTV systems with well-conditioned costs). Consider a trajectory tracking problem with LTV dynamics and well-conditioned costs, which generalizes the standard linear quadratic tracking problem in [27] with LTI dynamics and quadratic costs. We adopt the LTV dynamics $x_{t+1} = A_tx_t + B_tu_t + w_t$ and general well-conditioned cost functions $f_t(\cdot), c_t(\cdot)$ (see Assumption 2.2). With the desired trajectory $d_{1:T}$, we consider a new state $\hat{x}_t := x_t - d_t$ and a new disturbance $\hat{w}_t := w_t + A_td_t - d_{t+1}$. Thus, using the new state and disturbance, the problem naturally fits into our problem setting with $k$ future predictions of $(A_t, B_t, w_t, f_t, c_t, d_{t+1})$. Note that predictive control with LTV dynamics is practical in nonlinear systems [13, 27] because the nonlinearity could be well approximated by LTV models [27].

**Example 2.2** (Power grid frequency regulation). Consider the frequency regulation problem in [27], where state $x = [\theta^T, \omega^T]^T$ represent the status of a power plant, and power generation $p_m \in \mathbb{R}^n$ is the control action. The continuous-time dynamics is given by

$$\begin{bmatrix} \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M(t)^{-1}L & -M(t)^{-1}D \end{bmatrix} \begin{bmatrix} \theta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ M(t)^{-1} \end{bmatrix} p_m. \quad x(t)$$
We study a classical predictive control (PC) algorithm inspired by model predictive control, which expresses a sub-trajectory of the offline optimal trajectory as the solution of an optimization problem. Here $M(t)$ denotes the rotational inertia matrix, which is time-varying and is determined by the proportion of renewable power in total power generation at time $t$, and can be accurately predicted in a certain time horizon [2][3]. $L$ and $D$ are known system parameters. Using standard discretization techniques, we can formulate a discrete-time linear time-varying system $x_{t+1} = A_t x_t + B_t u_t + w_t$, where $A_t$ and $B_t$ are determined by $\hat{A}(t)$ and $\hat{B}(t)$. The cost functions are quadratic costs which penalizes frequency deviation [8]. This setting fits into our predictive control algorithm, since the controllers have accurate predictions of $A_t$ and $B_t$ in the near future due to predictability of $M(t)$.

### Algorithm 1 Predictive Control ($PC_k$)

1: for $t = 0, 1, \ldots, T - k - 1$ do
2:   Observe current state $x_t$ and receive predictions $\tilde{y}_{t:T+k-1}$.
3:   Solve and commit control actions $u_t := \tilde{\psi}_k^t (x_t, \tilde{w}_{t:T+k-1}; F)_{0_t}$.
4:   At time step $t = T - k$, observe current state $x_t$ and receive predictions $\tilde{y}_{t:T-1}$.
5:   Solve and commit control actions $u_{t:T-1} := \tilde{\psi}_k^t (x_t, \tilde{w}_{t:T-1}; 0)_{0_{T-1}}$.

Here $\psi$ is a function that expresses the terminal cost function regularizing the final state. Here we additionally require that the terminal cost $H$ has the form $H(x) = \alpha(||x||)$, where $\alpha : \mathbb{R} \to \mathbb{R}$ is a convex function (i.e. continuous increasing function with 0 at the origin, see Appendix A for definition) that is twice continuously differentiable. For each time step $\tau = 1, \ldots, k$, $y_\tau \in \mathbb{R}^n$ is the predictive state, and $v_\tau \in \mathbb{R}^m$ is the predictive control action. To make the algorithm well-defined, at time step $t = T - k$, $PC_k$ can finish the rest of the trajectory optimally by committing $u_{T-k:T-1} = \tilde{\psi}(x_{T-k}, \tilde{w}_{T-k:T-1}; 0)$.

The pseudo-code of predictive control is given in Algorithm 1.

It is also important in our framework to study the behavior of predictive control under some fixed terminal point. So, for prediction length $p \geq 1$ and time step $t$, we define an auxiliary optimization problem with a strict terminal constraint $y_p = z$ as follows:

$$\psi_p^t(x, \xi, z) := \arg \min_{y_0, \xi_0, \xi_{p-1}} \sum_{\tau=1}^{p} f_{t+\tau}(y_{\tau}) + \sum_{\tau=1}^{p} c_{t+\tau}(v_{\tau-1})$$

s.t. $\begin{align*}
y_{\tau} &= A_{t+\tau-1} y_{\tau-1} + B_{t+\tau-1} v_{\tau-1} + \xi_{\tau-1}, \tau = 1, \ldots, p, \\
y_0 &= x, \quad y_p = z,
\end{align*}$

where the optimal value is denoted by $\psi_p^t(x, \xi, z)$. We define this auxiliary optimization problem besides $\tilde{\psi}_k^t (x, \xi)$. Because we need to fix both the initial state and the terminal state, for example, when expressing a sub-trajectory of the offline optimal trajectory as the solution of an optimization problem, $\psi$ also allows us to study the impact of the perturbation at the terminal point on the optimal trajectory directly, which will be useful in the proof of dynamic regret (Theorem 4.2) and competitive ratio (Theorem 4.3).
Throughout the paper, we use \( \{(x_t, u_t)\}_{t=1}^T \) to denote the trajectory of predictive control, and use \( \{(x^*_t, u^*_t)\}_{t=1}^T \) to denote the offline optimal trajectory (i.e., the optimal solution of (1)). We also use several standard definitions and notations in linear algebra and optimization, which we detail in Appendix A for clarity. In particular, we use vector 2-norms and induced matrix 2-norms throughout this paper unless otherwise specified.

Two standard metrics will be utilized to assess the performance of our predictive control algorithms, namely dynamic regret \( \sup_{x(0),w} (\text{cost}(\text{ALG}) - \text{cost}(\text{OPT})) \) (the additional cost of our algorithm against the optimal algorithm) and competitive ratio \( \sup_{x(0),w} \frac{\text{cost}(\text{ALG})}{\text{cost}(\text{OPT})} \) (the worst-case ratio of the cost of our algorithm over that of the optimal algorithm).

### 3 A Perturbation Approach

In order to study the regret and competitive ratio of controllers in LTV systems, we develop a new analysis based on a perturbation approach, which we introduce in this section. This approach is based on developing bounds on how much the solutions to (2) and (3) change with respect to perturbations to the initial/terminal states and the disturbance sequence. Our perturbation bounds are related to the concept of incremental stability defined in [44], but not exactly the same because we consider the optimal trajectory in a finite horizon whereas the incremental stability focuses on asymptotic behavior over an infinite horizon. Simply stated, the key to our approach is to derive the perturbation bound in Theorem 3.3, which states that if the target variable we are concerned with is the \( h \)-th predictive state/control input, while the perturbation occurs at the \( \tau \)-th time step, then the impact on the target variable is exponentially small with respect to the time difference \( |h - \tau| \).

Proving such a result directly is challenging because of the complexity of the LTV dynamical constraints in (2) and (3). Thus, we develop a novel reduction from LTV systems to fully-actuated systems, i.e., systems where the controller can steer the system to any state in the whole space \( \mathbb{R}^n \) freely at every time step. This special case is a form of online optimization called smoothed online convex optimization (SOCO), and has received considerable attention recently, e.g., [24] [27] [53]. We exploit the controllability of the dynamics to analyze the LTV system in chunks of \( d \) time steps. A sequence of \( d \) time steps combined together can be thought as a fully-actuated system and thus we can formulate a SOCO problem, which is \((1/d)\)-times as long as the original LTV system. In this section, we first show the perturbation bound for SOCO in Section 3.1, and then we leverage our reduction to derive a result for general LTV systems in Section 3.2.

#### 3.1 Smoothed Online Convex Optimization

The classic setting of SOCO is an online game played by an agent against an adversary: at each time step \( t \), the adversary reveals a hitting cost function \( f_t \), a switching cost function \( g_t \), and a disturbance (or exogenous input) \( \hat{w}_t \). The agent picks a decision point \( \hat{x}_t \in \mathbb{R}^n \), and incurs a stage cost of \( f_t(\hat{x}_t) + g_t(\hat{x}_t, \hat{x}_{t-1}, \hat{w}_{t-1}) \). The agent seeks to minimize the total cost it incurs throughout the game. The offline optimal cost is defined as the minimum cost if the agent has full knowledge of the costs and disturbances at the start of the game. Instead of analyzing the performance of an online algorithm directly, our focus is on studying how the perturbations of the system parameters (initial state, terminal state, and disturbances) impact the offline optimal solution. These results are critical for deriving the guarantees for predictive control in the online setting in Section 4.

To begin, observe that when the initial state \( \hat{x}_0 \), terminal state \( \hat{x}_p \), and the disturbances \( \hat{w} \) are given, the optimal \( p \)-step trajectory of SOCO can be obtained from the unconstrained optimization problem

\[
\hat{\psi}(\hat{x}_0, \hat{w}, \hat{x}_p) := \arg \min_{\hat{x}_1, \ldots, \hat{x}_{p-1}} \sum_{\tau=1}^{p-1} f_\tau(\hat{x}_\tau) + \sum_{\tau=1}^p g_\tau(\hat{x}_\tau, \hat{x}_{\tau-1}, \hat{w}_{\tau-1}),
\]

where the objective is a convex function of the decision variables \( \hat{x}_1, \ldots, \hat{x}_{p-1} \). Since (4) is an unconstrained optimization problem, the gradient of its objective equals zero at \( \hat{\psi}(\hat{x}_0, \hat{w}, \hat{x}_p) \). Using this, we can further show that the directional derivative of \( \hat{\psi}(\hat{x}_0, \hat{w}, \hat{x}_p) \) along some direction \( e \) denoted by \( \chi \), satisfies the linear equation \( M \chi = \delta \), where symmetric matrix \( M \) is the Hessian of the objective and vector \( \delta \) is determined by the direction \( e \). A special structure of the objective of (4) is that the correlations only occur in two consecutive time steps. This implies that its Hessian \( M \) is block
We now build upon the SOCO perturbation result to derive a perturbation result for LTV systems. In particular, we show an exponentially-decaying perturbation bound for our LTV system by reducing it to SOCO and apply Theorem 5.1. As we have discussed, LTV systems are more difficult than SOCO because the dynamics prevent the online agent from picking the next state \(x_{t+1}\) freely at a given state \(x_t\). We overcome this obstacle by redefining the decision points as illustrated in Figure 4. Specifically, given state \(x_t\) at time step \(t\) as the last decision point, we then ask the online agent to decide state \(x_{t+d}\) at time step \((t+d)\) rather than \(x_{t+1}\) at time step \((t+1)\).

Since \(d\) is the controllability index, \(x_{t+d}\) can be picked freely from the whole space \(\mathbb{R}^n\) regardless of \(x_t\). We also use the principle of optimality, e.g., if \(y_{0:k}, v_{0:k-1}\) is the optimal solution to \(\psi^p_t(x, \xi, z)\), then \(y_{i:j}, v_{i:j-1}\) is the optimal solution to \(\psi^p_{t+i}(y_i, \xi_{t+j-1}, y_j)\) for any \(0 \leq i < j \leq k\). Therefore, the trajectory between time \(t\) and \((t+d)\) can be recovered by solving \(\psi^p_t(x_t, w_{t:t+d-1}, x_{t+d})\). So we are able to formulate a valid SOCO problem on the sequence of time steps \(t, t+d, t+2d, \ldots\).

Naturally, the hitting cost at time step \((t+d)\) remains the same, while the switching cost becomes \(\xi^d_t(x_t, w_{t:t+d-1}, x_{t+d})\), where the function \(\xi^d_t(x, z)\) is defined as

\[
\xi^d_t(x, z) := \min \left\{ \psi^p_t(x, \xi, z) \right\}.
\]

An illustration of the reduction can be found in Figure 4. We would like to point out that our reduction from optimal control to SOCO is novel in that it leverages the principle of optimality to apply to more general LTV settings, as opposed to the reduction via control canonical forms in [3] that is specific to LTI systems. Unlike the switching costs in \([14, 27, 35, 38]\) which are explicitly defined as the \(\ell_2\)-distance or squared \(\ell_2\)-distance, the switching cost \(\xi^p_t(x, z)\) here is defined implicitly as the optimal value of an optimization problem. Lemma 5.2 shows that the switching cost defined in \([5]\) satisfies the requirements of Theorem 5.1 which allows us to obtain the desired perturbation bound.

**Lemma 3.2.** Under Assumption 2.1 and 2.2 for integer \(p \geq d\), we have

1. \(\psi_t^p(x, \xi, z)\) is \(L_1(p)\)-Lipschitz in \((x, \xi, z)\);
2. \(\xi_t^p(x, \xi, z)\) is convex and \(L_2(p)\)-strongly smooth in \((x, \xi, z)\).

Here \(L_1(p) = C(p)(1 + \ell \cdot C(p)/m_c)\), \(L_2(p) = \ell \cdot C(p)^2 + \ell^2 \cdot C(p)^4/m_c\), where \(\ell = \max(\ell_f, \ell_c)\).

\[
C(p) = \begin{cases} 
O(a^{3p}) & \text{if } a > 1; \\
O(p^2) & \text{if } a = 1; \\
O(1) & \text{if } a < 1.
\end{cases}
\]
We now demonstrate the power of the perturbation approach in Section 3.2 by obtaining bounds on the controller’s trajectory and the offline optimal trajectory. The first inequality because an all zero trajectory can be expressed as the optimal solutions of the optimization problem at the next time step due to dynamical constraints. However, given \((d, \sigma)-\)uniform controllability, the controller is able to enforce an arbitrary target state after \(d\) time steps, which prompts the transformation to a SOCO problem with a decision point in every \(d\) time steps.

In Lemma 3.2, we use \(O(\cdot)\) to hide quantities \(a, b, \) and \(1/\sigma\); the precise expression of \(C(p)\) and the proof of Lemma 3.2 can be found in Appendix E. Using the reduction from LTV to SOCO, we obtain a perturbation bound for the LTV systems \((2)\) and \((3)\) in Theorem 3.3, the proof of which is deferred to Appendix D.

**Theorem 3.3.** Consider the optimization problem defined in \((2)\) and \((3)\) and with a horizon length \(p \geq d.\) Under Assumptions 2.1 and 2.2 given any \((x, \zeta, z)\) and \((x', \zeta', z')\),

\[
\| \tilde{\psi}^P_t (x, \zeta; F) y_h - \tilde{\psi}^P_t (x', \zeta'; F) y_h \| \leq C \left( \lambda^h \| x - x' \| + \sum_{r=0}^{p-1} \lambda^{h-r} \| \zeta_r - \zeta'_r \| \right)
\]

\[
\| \psi^P_t (x, \zeta, z) y_h - \psi^P_t (x', \zeta', z') y_h \| \leq C \left( \lambda^h \| x - x' \| + \sum_{r=0}^{p-1} \lambda^{h-r} \| \zeta_r - \zeta'_r \| + \lambda^{p-h} \| z - z' \| \right)
\]

hold for all time steps \(t.\) Here we define \(L_0 = \max_{d \leq p \leq 2d-1} L_2 (p),\) and the constants are given by

\[
\lambda = \left( 1 - 2 \left( \sqrt{1 + (2L_0/m) + 1} \right)^{-1} \right)^{1-1} \quad C = \frac{2L_0}{m} \cdot \left( 1 - 2 \left( \sqrt{1 + (2L_0/m) + 1} \right)^{-1} \right)^{1-1}.
\]

Theorem 3.3 allows us to bound the distance between any two trajectories so long as they can be expressed as the optimal solutions of the optimization problem \((2)\) or \((3)\). For example, to bound the norm of each state in the predictive trajectory \(\tilde{\psi}^P_t (x, \zeta; F),\) we only need to set \(x' = 0, \zeta' = 0\) in the first inequality because an all zero trajectory can be expressed as \(\tilde{\psi}^P_t (0, 0; F).\) The formal statement of this result can be found in Appendix E.

4 **Performance Guarantees for Predictive Control**

We now demonstrate the power of the perturbation approach in Section 3.2 by obtaining bounds on regret and competitive ratio. The key intuition behind our analysis is the following: at time step \(t,\) if the predictive controller with prediction window \(k\) is given the knowledge of \(x^*_t\) and \(x^*_{t+k},\) it can fully recover the offline optimal states and control inputs for the future \(k\) time steps, \(x_{t+1:t+k}\) and \(u_{t+1:t+k-1}^*,\) from \(\psi^P_t (x^*_t, w_{t+1:t+k-1}, x^*_{t+k})\). However, without the knowledge of the offline optimal states, the predictive controller solves \(\psi^P_t (x_t, w_{t+1:t+k-1}, x_{t+k})\) instead, where \(x_{t+k}\) is implicitly determined by the \(k\)-th predictive state of \(\tilde{\psi}^P_t (x_t, w_{t+1:t+k-1}; F).\) We overcome this gap with our perturbation approach (specifically, Theorem 3.3 and its corollaries), which allows us to bound the distance between the controller’s trajectory and the offline optimal trajectory.
4.1 Dynamic Regret

We first bound the dynamic regret of predictive control. For this analysis, a key observation is that the offline optimal trajectory is given by $x^* = \hat{\psi}_0^T (x_0, w_{0:T-1}; 0)_{y_{1:T}}$. Furthermore, the optimal trajectory starting at time step $t$ with state $x_t$ is equivalent to the trajectory of predictive control with prediction window $(T-t)$ and no terminal cost, i.e. $\hat{\psi}_t^{T-t} (x_t, w_{t:T-1}; 0)_{y_{1:T-t}}$. Using Theorem 3.3, we can bound the change in decision points against the change in prediction window $k$. Lemma 4.1 formalizes this:

**Lemma 4.1.** For any integers $p, h$ such that $p \geq h \geq 1$ and time step $t < T-p$, we have

$$\| \hat{\psi}_{t}^p (x_t, w_{t+T-p}; F)_{y_h} - \hat{\psi}_{t+1}^{p+1} (x_t, w_{t+T+p}; F)_{y_h} \| \leq 2CP^{p-h} \left( C \| x_t \| + \frac{2C}{1-\lambda} \sup_{0 \leq \tau \leq T-1} \| w_\tau \| \right).$$

Then, we can follow the road map below to bound the dynamic regret of predictive control $PC_k$:

(a) Given the well-conditioned state/control costs, it suffices to bound the distance between $PC_k$’s trajectory and the offline optimal trajectory (i.e., $\| x_t - x_t^* \|$) to show the dynamic regret result. See inequalities (32) and (33) in Appendix H for technical details.

(b) At each time step $t$, the optimal next state (under an imaginary terminal cost $F$) from the current state $x_t$ is given by $\hat{\psi}_t^{T-t} (x_t, w_{t:T-1}; F)_{y_i}$. However, reaching the optimal next state from $x_t$ requires full knowledge of the future costs, dynamics, and disturbances. Although $PC_k$ cannot reach the optimal next state due to incomplete knowledge of the future, it can leverage the predictions of future $k$ steps to decide a near-optimal control action from state $x_t$. By cumulatively summing up the bounded difference in Lemma 4.1 and applying Theorem 3.3, one can show the suboptimality, measured by the distance $\| x_{t+1} - \hat{\psi}_t^{T-t} (x_t, w_{t:T-1}; F)_{y_i} \|$, is in the order of $O(\lambda^k)$. See inequality (19) in Appendix H for technical details.

(c) Using the exponentially-decaying LTV perturbation bound in Theorem 3.3, we can convert the per-step suboptimality bounds to a global suboptimality bound on $\| x_t - x_t^* \|$ that is also in the order of $O(\lambda^k)$. See inequalities (20) and (21) in Appendix H for technical details.

The $O(\lambda^k)$ upper bound on the distance between the algorithm’s trajectory and the offline optimal trajectory leads to the regret bound in Theorem 4.2:

**Theorem 4.2.** Suppose $\| w_t \| \leq D$ for some constant $D$ at each time step $t$. Let $\lambda, C, L_0$ be the decay rate and constants defined in Theorem 3.3. If prediction window $k \geq d$ is sufficiently large, such that

$$k \geq 1 + \log \left( \frac{1}{1-\delta} \cdot \left( \frac{2C}{1-\lambda} \right) \right) \left( \log \left( \frac{1}{\lambda} \right) \right)$$

for some positive constant $\delta \in (0, 1)$, then the trajectory of $PC_k$ satisfies:

1. **(Input-to-state Stability)** The norm of each state $x_t$ is upper bounded by

$$\| x_t \| \leq \left( \frac{C}{\delta} \cdot (1-\delta)^{T-2k}\lambda^{k} \| x_0 \| + \left( \frac{2C}{\delta} \right) \left( 1+ \frac{2C}{1-\lambda} \right) D \right) \quad \text{if } 0 < t \leq T - k$$

$$\left( \frac{C}{\delta} \cdot (1-\delta)^{T-2k}\lambda^{k} \| x_0 \| + \left( \frac{2C}{\delta} \right) \left( 1+ \frac{2C}{1-\lambda} \right) D \right) \quad \text{if } T - k < t \leq T.$$  

2. **(Dynamic Regret)** The dynamic regret of $PC_k$ is upper bounded by

$$\text{cost}(PC_k) - \text{cost}(OPT) = O \left( \frac{D + \frac{\lambda^k (\| x_0 \| + D)}{\delta}}{\lambda^k T + \lambda^k \| x_0 \|^2} \right).$$

where the notation hides quantities $a, b, \ell_f, \ell_c$, $C, 1/(1-\lambda)$ and $L_0$.

An implication of Theorem 4.2 is that to obtain $o(1)$ dynamic regret when the norm of disturbances are uniformly upper bounded, it suffices to use a prediction window of length $\Theta(\log T)$. This parallels the result shown in [4], although in a more general setting.

4.2 Competitive Ratio

We now focus on bounding the competitive ratio of predictive control. Here, we study a modification of the predictive control algorithm we have considered to this point. In particular, we introduce a
Algorithm 2 Predictive Control with Replan Window $h$ ($PC_{(k,h)}$)

1: Suppose $T = n_0 h + m_0$, where integers $n_0 \geq 0, k - h + 1 \leq m_0 \leq k$.
2: for $t = 0, h, \ldots, n_0 (h - 1)$ do
3: Observe current state $x_t$ and receive predictions $\theta_{t:t+k-1}$.
4: Solve and commit control actions $u_{t:t+k-1} := \psi^k_t(x_t, w_{t:t+k-1}; F)_{\theta_{t:t+k-1}}$.
5: At time step $t = n_0 h$, observe current state $x_t$ and receive predictions $\theta_{t:T-1}$.
6: Solve and commit control actions $u_{t:T-1} := \psi^m_{n_0 h} (x_t, w_{t:T-1}; 0)_{\theta_{t:T-1}}$.

replan window $h$, as defined in Algorithm 2, which we denote as $PC_{(k,h)}$. This style of algorithm has been considered previously in the SOCO literature, where it has been shown to obtain a constant competitive ratio in some settings where MPC does not.

Our analysis approach highlights why this modification is beneficial for competitive ratio. Specifically, we obtain the competitive ratio bound by applying a potential method building on. We define the potential function as the squared distance between the algorithm’s trajectory and the offline optimal trajectory, i.e., $\phi_t(x_t, x^*_t) = \|x_t - x^*_t\|^2$, which is standard in the literature. We study how this potential function changes over time. Intuitively, we need to upper bound the increment of this potential function by the offline optimal cost to obtain a competitive ratio result. To achieve this, the algorithm needs to “move closer” to the offline optimal trajectory rather than “moving further away” from it. Recall that Theorem 3.3 gives that

$$\|\psi^k_t(x_t, w_{t:t+k-1}; F)_{\theta_h} - \psi^k_t(x^*_t, w_{t:t+k-1}; F)_{\theta_h}\| \leq C \lambda^h \|x_t - x^*_t\|. \quad (7)$$

When the algorithm commits the first predictive state ($h = 1$), the left hand side of (7) might be larger than $\|x_t - x^*_t\|$ when $\lambda > 1$. Thus, the algorithm must “wait” until the right hand side of (7) becomes smaller than $\|x_t - x^*_t\|$. This is accomplished in Algorithm 2 via the replan window $h$.

Our main result for this section is the following competitive ratio bound for $PC_{(k,h)}$.

**Theorem 4.3.** Let $\lambda, C, L_0$ be the decay rate and constants defined in Theorem 3.3. In Algorithm 2 if the replan window $h$ satisfies $h \geq \max\{\log ((1 + \varepsilon)C)/\log (1/\lambda), d\}$ for some positive constant $\varepsilon$, and the prediction window $k$ satisfies $k \geq h + d$, then it has competitive ratio

$$\sup_{x(0),w} \frac{\text{cost}(PC_{(k,h)})}{\text{cost}(OPT)} = 1 + O \left( e^{-1} \left( \frac{L_0 + \ell_f}{m_f} \right)^{1/2} \cdot C \lambda^{k-1-h} \right),$$

where the notation only hides a small numerical constant.

Note that when the constant $\varepsilon$ and the replan window $h$ are fixed, the competitive ratio is on the order of $1 + O(\rho^h)$ as the length of prediction $k$ tends to infinity. One potential line of future work is to understand if the replan window is necessary. It may be possible to either strengthen the constants given in Theorem 3.3 or improve our proof approach so as to eliminate the requirement on $h$.

**References**


