On the Regret Analysis of Online LQR Control with Predictions

Runyu Zhang, Yingying Li, and Na Li

Abstract—In this paper, we study the dynamic regret of online linear quadratic regulator (LQR) control with time-varying cost functions and disturbances. We consider the case where a finite look-ahead window of cost functions and disturbances is available at each stage. The online control algorithm studied in this paper falls into the category of model predictive control (MPC) with a particular choice of terminal costs to ensure the exponential stability of MPC. It is proved that the regret of such an online algorithm decays exponentially fast with the length of predictions. The impact of inaccurate prediction on disturbances is also investigated in this paper.

I. INTRODUCTION

Consider a classical finite-horizon discrete-time linear quadratic regulator (LQR) problem:

$$\min_{u_1, \ldots, u_T} \quad \sum_{t=1}^{T-1} (x_t^T Q x_t + u_t^T R u_t) + x_T^T Q x_T$$

s.t. $x_{t+1} = A x_t + B_u u_t + B_d d_t$, \hspace{1cm} $t \geq 1$,

where the cost function parameters $Q$, $R$, and the system disturbances $d_t$ are time-varying, while the system parameters $A, B_u, B_d$ are time-invariant. It is well-known that the optimal control input to (1) at time step $t$ requires the information of all the future, i.e. $\{Q_\tau, R_\tau, d_\tau\}_{\tau \geq t}$ [1]. However, in many real-world applications, e.g., autonomous driving [2], energy systems [3], data center management [4], it is impractical for a decision maker to acquire all the (accurate) future information beforehand. Instead, the decision maker may only have access to some predictions for the near future and the predictions can be inaccurate. Hence, this calls for the study of online LQR problem with limited and inaccurate predictions of the future. Specifically, this paper considers the following online LQR problem: at each time step $t$, the controller has to take an action based on the $W$-step ahead predictions $\{Q_{i|t}, R_{i|t}, d_{i|t}\}_{i=t+1}^{t+W}$ and disturbance predictions $\{d_{i|t}\}_{i=t+1}^{t+W}$ for the next $W$ time steps. For simplicity, we only consider inaccurate disturbance predictions and assume cost predictions are accurate. The goal of online LQR is to minimize the total cost in (1) by only leveraging the predictions and the history.

Among all the online control algorithms that leverage predictions, perhaps model predictive control (MPC) is the most popular one. Although MPC has been intensively studied both for linear systems and nonlinear systems [5]–[11], most studies focus on asymptotic performance such as stability or convergence to some optimal state. Motivated by the aforementioned applications, there is an increasing need to understand the non-asymptotic performances of MPC such as the cost difference compared to the optimal cost over a finite time-horizon. Though there are recent online control papers, e.g., [12]–[14], that study the non-asymptotic behavior of online control algorithms, they do not consider predictions, that is, the controller has to take an action without any knowledge of future $\{Q_{\tau}, R_{\tau}, d_{\tau}\}_{\tau \geq t}$ at time $t$.

In online learning community, on the contrary, there are many papers on the non-asymptotic performance analysis, where the performance is usually measured by regret, e.g., static regrets [15], [16], dynamic regrets [17], etc. But most papers either do not system dynamics or predictions [18]–[21] or only consider simple systems [22], [23] or simplified prediction models where $W$-step ahead predictions are accurate without errors [22], [24].

The setting considered in this paper is closest to the recent papers [24]–[27]. [24], [25] consider a linear dynamical system with time-varying cost functions but no disturbances. They focus on gradient-based online control rather than the more commonly used MPC approach. On the other hand, [26], [27], study the non-asymptotic behavior of MPC algorithm, but they only consider time-invariant cost function and accurate disturbance predictions. But one interesting message from [26] is that MPC turns out to be optimal (or nearly optimal) in stochastic settings (or in adversarial settings) with respect to (dynamic) regrets. This further motivates us to look into the regret analysis of MPC for the setting with time-varying cost functions and inaccurate disturbance predictions.

Contribution. In this paper, we provide an explicit upper bound for the performance of MPC (Theorem 1) in terms of the dynamic regret: the online cost minus the optimal cost in hindsight. The MPC studied in this paper follows the standard MPC framework with a particular choice of terminal cost which ensures the exponential stability of MPC. Our regret bound consists of two parts: the first part decays exponentially with the prediction window $W$ and the second part increases with the prediction errors. The first part indicates the benefits of having more predictions and the second part reflects the negative impact of inaccurate predictions. When there are no prediction errors, the second part is zero and the regret bound decays exponentially with $W$. Further, in the second part of our regret bound, the errors of long-term predictions play an exponentially diminishing effect, i.e. the impact of the error of $d_{t+k|t}$ decays exponentially with $k$ on the regret bound. This indicates that MPC (implicitly) focuses more on the short-term predictions and less on long-term ones, which is desirable in most cases since long-term predictions usually suffer from poor quality. Further, this suggests that our regret upper bound provides useful guidelines in choosing the prediction window $W$ used in MPC in the face of inaccurate predictions.

To develop our regret bound, we provide a general regret formula for any online algorithms for online LQR problems, which is a quadratic function of the differences between the
online control actions and the optimal control actions. Our formula is established by leveraging a cost difference lemma in the (Markov decision processes) literature [28], [29] and the special properties of LQR. Our formula greatly relieves the difficulty of non-asymptotic regret analysis of MPC. Furthermore, the formula can be applied to other online control algorithms and thus can be viewed as a contribution on its own merit.

Notations: The norm $\| \cdot \|$ refers to the $L_2$ norm for both vectors and matrices, $\lambda_{\text{min}}(A)$ denotes the minimum eigenvalue of matrix $A$, and $\lambda_{\text{max}}(A)$ denotes the maximum eigenvalue of $A$. For any symmetric matrices $A$ and $B$, we write $A < B$ if $B - A$ is positive definite.

II. PROBLEM SETUP AND PRELIMINARIES

A. Problem Formulation: Online LQR

As stated in Section I, we consider an online linear quadratic regulator (LQR) problem with process noises/disturbances. The system dynamics is provided by:

$$x_{t+1} = Ax_t + Bu_t + B_dd_t, \quad t \geq 1 \tag{2}$$

where the initial state $x_1 \in \mathbb{R}^n$ is fixed, $x_t \in \mathbb{R}^n$ and $u_t \in \mathbb{R}^m$ denote the state and control input at stage $t$ respectively, and $d_t \in \mathbb{R}^p$ denotes the process noise/disturbance. We consider time-varying quadratic costs at each stage $t$, i.e. $x_t^TQ_t x_t + u_t^TR_t u_t$. The total cost over $T$ stages is defined as

$$J(x, u) = \sum_{t=1}^{T-1} (x_t^TQ_t x_t + u_t^TR_t u_t) + x_T^TQ_T x_T,$$

where $x = [x_1^T, \ldots, x_T^T]^T$, $u = [u_1^T, \ldots, u_{T-1}^T]^T$, $Q_T$ represents the terminal cost, and we impose $R_T = 0$ for notational simplicity.

The control objective is to design control input $u_t$ at each $t$ to minimize the total cost $J(x, u)$. However, the optimal control at time $t$ requires the information of all the future cost functions and disturbances, i.e. $\{d_i, Q_i, R_i\}_{i=t}^T$ (see e.g. Proposition 1), which may not be practical in real-world applications. Nevertheless, some predictions are usually available beforehand, especially for the near future. In this paper, we consider that the predictions of the next $W$ stages are available, i.e. $\{d_{i|t}, Q_{i|t}, R_{i|t}\}_{i=t}^{t+W}$, where $d_{i|t}$ denotes the prediction of the disturbance $d_i$ at stage $t$, and the same applies to $Q_{i|t}$, $R_{i|t}$. The prediction information can be inaccurate and it is worth discussing the impact of the prediction errors. For this paper, we focus on the prediction errors of the disturbances and denote $e_{i|t} := d_{i|t} - d_i$ as the prediction error of $d_i$. Since the prediction $d_{i|t}$ is received at stage $t$, which is $(i-t)$ stages before stage $i$, we also call $e_{i|t}$ as the $(i-t)$-step prediction error of $d_i$. For simplicity, we omit the prediction errors of the cost matrices by considering accurate $W$-stage cost predictions: $Q_{i|t} = Q_i$ and $R_{i|t} = R_i$ for $t \leq i \leq t + W$.

In summary, the online LQR considered in this paper is described as follows: Assuming $A$, $B_u$, $B_d$ is known apriori, at each step $t = 1, 2, \ldots$,

- the controller observes state $x_t$ and receives predictions $\{d_{i|t}, Q_{i|t}, R_{i|t}\}_{i=t}^{t+W}$;
- the controller implements $u_t$ based on the predictions $\{d_{i|t}, Q_{i|t}, R_{i|t}\}_{i=t}^{t+W}$ and the history $\{x_i, d_i, Q_i, R_i\}_{i=1}^t$ and suffers the cost $x_t^TQ_t x_t + u_t^TR_t u_t$;
- the system evolves to the next state $x_{t+1}$ by (2) under the real disturbance $d_t$.

Our goal is to design an online algorithm to reduce the total cost by exploiting the online available information, i.e. the predictions and the history. For example, consider an online control algorithm denoted by $\pi = \{\pi_1, \ldots, \pi_{T-1}\}$, where $\pi_t$ represents the policy at step $t$. Notice that $\pi_t$ only has access to the online available information at $t$, i.e., the control action $u_t$ at time $t$ is determined by

$$u_t = \pi_t(\{x_{\tau}^T\}_{\tau=1}^t, \{d_{\tau}, Q_{\tau}, R_{\tau}\}_{\tau=1}^t, \{d_{i|t}, Q_{i|t}, R_{i|t}\}_{i=t}^{t+W}),$$

We measure the performance of the online algorithm $\pi$ by dynamic regret, which compares the total cost of $\pi$ with the optimal total cost $J^*$ in hindsight, that is,

$$\text{Regret}(\pi) = J(x^\pi, u^\pi) - J^*.$$ 

where $x^\pi, u^\pi$ denote the state and action at step $t$ generated by the online algorithm $\pi$. Let $\pi^* = \{\pi_1^*, \ldots, \pi_{T-1}^*\}$ denote the optimal controller in hindsight that yields the optimal cost $J^*$. Here $\pi_t^*$ is the optimal policy at each step $t$, which will be further discussed in Section II-B.

Dynamic regret is a commonly used performance metric in the literature [24], [26]. The benchmark of the dynamic regret described above is optimal time-varying policies. Notice that another popular regret notion is the static regret, whose benchmark is the optimal time-invariant policy which is a weaker benchmark because the optimal control for finite-time horizon time-varying LQR (II-A) is time-varying.

Throughout the paper, we consider the following assumptions on the dynamics and the cost matrices, which are standard assumptions in the literature.

**Assumption 1.** The pair $(A, B_u)$ is stabilizable. All pairs $(A, Q_i)$ are detectable.

**Assumption 2.** There exist positive definite matrices $Q_{\text{min}}$, $Q_{\text{max}}$, $R_{\text{min}}$, $R_{\text{max}}$ such that $Q_i$, for any $1 \leq i \leq T$ and $R_i$ for any $1 \leq i \leq T - 1$ satisfy

$$0 < Q_{\text{min}} \leq Q_i \leq Q_{\text{max}}, \quad 0 < R_{\text{min}} \leq R_i \leq R_{\text{max}}.$$

1Ideally, we would like to also consider inaccurate prediction on $Q_i$, $R_i$, but due to some technical difficulty in analyzing the performance, it is left as our future work. Nevertheless, our setting still finds applications, e.g., when the cost function is set according to financial contracts or cost planning steps-ahead. However, disturbances are often due to volatile nature such as wind. Thus allowing inaccurate predictions greatly broadens the applications of previous settings studied in [23], [24], [26].
The optimal policies $\pi_t$ minimize $J(x,u)$ and thus, can be written as:

$$\pi^*: u_t = \pi_t(x_t, d_t, \ldots, d_{t-1}) = -K_t x_t - \sum_{i=t}^{T-1} K_{t+1} B_{i+1} d_i,$$

where for $i = t, \ldots, T-1$

$$K_t^{d,i} = (R_t + B_u^T P_{t+1} B_u)^{-1} B_u^T \Phi(i+1, t+1)^T P_{t+1}.$$

Furthermore, the exponential stability of LQR can also be established.
In order to analyze the dynamic regret of the MPC, we first seek a different representation of the online algorithm which allows us to compare the online MPC policy and the optimal offline policy in (7). The key observation is that Proposition 1 can also be applied to solve (10) for MPC which allow us to represent the MPC in a similar form as in (7),

\[
\text{MPC : } u_t^{\text{MPC}} = -R_t t^{\text{MPC}} - \sum_{i=t}^{t+W} R_{t+i}^{d,i} B_d d_{i|t}
\]  

(11)

Here \( R_t, R_t^{d,i} \) are constructed in a similar manner as \( K_t, K_t^{d,i} \). First define:

\[
\begin{align*}
\bar{P}_{t+\tau|t} & := P_t + \tau \{ Q_t, R_t \}_{i=t+\tau}^{i=W}, P_{\text{max}} \\
\bar{K}_{t+\tau|t} & := K_t + \tau \{ Q_t, R_t \}_{i=t+\tau}^{i=W}, P_{\text{max}} \\
& = (R_{t+\tau} + B_u^T \bar{P}_{t+\tau+1|t} B_u)^{-1} B_u^T \bar{P}_{t+\tau+1|t} A
\end{align*}
\]

(12)

which are the cost-to-go and control gain matrices given the \( W \) steps ahead prediction \( \{ Q_t, R_t \}_{i=t}^{i=W} \) and the terminal cost \( P_{\text{max}} \) at step \( t + W + 1 \).

Further define the ‘predicted’ state transition matrix at step \( t \):

\[
\bar{\Phi}_t(t+j, t+i) := \begin{cases} 
(A - B_u R_{t+j-1|t}) \cdots (A - B_u R_{t+i|t}), & 0 \leq i < j \leq W \\
I, & 0 \leq i = j \leq W 
\end{cases}
\]

(13)

Algorithm 1 summarize the implementation of this MPC.

**Algorithm 1 Model Predictive Control (MPC) Algorithm**

**Input:** \( Q_{\text{min}}, Q_{\text{max}}, R_{\text{min}}, R_{\text{max}}, A, B_u, B_d \)

Pre-calculate \( P_{\text{max}} \) using (6).

**for** \( t = 1, 2, \ldots, T - 1 \) **do**

Observe \( x_t \) and receive predictions \( \{ d_{i|t}, Q_t, R_t \}_{i=t}^{i=W} \).

if \( t \leq T - W - 1 \) then

Calculate \( \bar{P}_{t+1|t}, \ldots, \bar{P}_{t+W+1|t}, \bar{K}_{t|t}, \ldots, \bar{K}_{t+W|t} \) using (12).

Calculate \( R_t, R_t^{d,i} \) using (13)(14).

Implement \( u_t \) as in (11)

else

Calculate \( P_{t+1}, \ldots, P_t, K_t, \ldots, K_{T-1} \) directly via (4)(5).

Calculate \( K_t^{d,i} \) using (8).

Implement \( u_t \) as in (7)

end if

end for

Before going into the regret analysis in the next section, we first the exponential stability of the MPC algorithm despite the time-varying cost functions.

**Proposition 3. (Exponential Stability of MPC)** Define the state transition matrix for MPC as:

\[
\bar{\Phi}_t \text{(in (13)}, )
\]

\[
\bar{\Phi}^{\text{MPC}}(t, t_0) := \begin{cases} 
(A - B_u R_{t-1}) \cdots (A - B_u R_{t_0}), & t > t_0 \\
I, & t = t_0 
\end{cases}
\]

Then:

\[
||\bar{\Phi}^{\text{MPC}}(t, t_0)|| \leq \tau \rho^{t-t_0},
\]

with \( \tau, \rho \) defined the same as in Proposition 2

Stability results of MPC is well studied in literature [5], [9]–[11], [33], [34] and this proposition almost follows the results in [33] exactly. Nevertheless, a self-contained proof is provided in our online report [35] for reader’s reference.

**IV. DYNAMIC REGRET ANALYSIS**

Recall the definition of \( \rho = \sqrt{1 - \frac{\lambda_{\min}(Q_{\text{min}})}{\lambda_{\max}(P_{\text{max}})}} \) in Proposition 2 and further define another factor:

\[
\gamma = \frac{\lambda_{\max}(A^T P_{\text{max}} A)}{\lambda_{\min}(Q_{\text{min}}) + \lambda_{\max}(A^T P_{\text{max}} A)}.
\]

The regret of the MPC in Algorithm 1 is bounded as follows.

**Theorem 1.** The regret of MPC in (11) is upper bounded by:

\[
\text{Regret(MPC)} \leq \alpha_1 \left[ \frac{1}{1 - \rho} \left( \gamma W + \gamma W \right) \right] + \alpha_2 \left[ \frac{1}{1 - \rho} \left( 1 + \gamma W \right) \right] + \alpha_3 \left[ \frac{1}{1 - \rho} \right] + \alpha_4 \left[ \frac{1}{1 - \rho} \right]
\]

(15)

where \( \alpha_1, \alpha_2, \alpha_3, \alpha_4 \) are constants that only relate to \( R_{\text{min}}, R_{\text{max}}, Q_{\text{min}}, Q_{\text{max}}, A, B_u, \) but not \( W \).

The regret bound (15) consists of two terms. Part I depends on the total magnitude of the disturbances: \( E_d := \sum_{i=1}^{T} \| B_d d_{i|t} \|^2 \). Part II depends on the error of the predictions \( \| B_d e_{i|j} \| \). Before showing the proof, we make the following discussions on interpreting the regrets.

**Factors \( \rho \) and \( \gamma \):** Proposition 2 and 3 show that the factor \( \rho \) can be interpreted as the contraction factor of \( \Phi, \bar{\Phi}^{\text{MPC}} \). We now further explain the meaning of \( \gamma \). In a word, \( \gamma \) captures the contraction of Riccati iteration \( F_{Q,R}(\cdot) \). In [36], they introduce a special metric on positive definite matrices

\[
\delta_{\infty}(P, P') := \| \log(P^{-\frac{1}{2}} PP'^{-\frac{1}{2}}) \|
\]

and show that \( F_{Q,R}(\cdot) \) is contractive under \( \delta_{\infty}(\cdot, \cdot) \), i.e.,

\[
\delta_{\infty}(F_{Q,R}(P), F_{Q,R}(P')) \leq \gamma \delta_{\infty}(P, P')
\]

(16)

(16) is going to play a key role in bounding the difference \( \|K_t - \bar{K}_t\| \) and \( \|K_t^{d,i} - \bar{K}_t^{d,i}\| \) and therefore the cost difference between \( J^{\text{MPC}} \) and \( J^{*} \), which will be further explained later.

**Impact of \( e_{i|j} \):** Let’s take a more careful look at Part II,

\[
\text{Part II} \sim O\left( \sum_{j=1}^{T-1} \sum_{i=j}^{j+W} \rho^{i-j} \| B_d e_{i|j} \|^2 \right)
\]
The coefficient in front of $\|B_d e_{ij}\|$ is dominated by $\rho^{i-j}$, which decays exponentially with $i-j$, which means that the effect of prediction inaccuracy decays exponentially fast as forecast goes to the far future. Since longer-term prediction tends to be less accurate, this suggests that MPC effectively alleviates the impact of multi-step-ahead prediction errors.

### Choice of $W$:
Part I involves the magnitude of the disturbances, and the coefficient in front of $\left(\|x_t\|^2 + \sum_{i=1}^{T} \|B_d d_t\|^2\right)$ is roughly of scale:

$$\text{Part I} \sim O\left(\frac{\gamma_0^{2W}}{\sum_{i=1}^{T} \|B_d d_t\|^2}\right)$$

where $\gamma_0 = \max\{\rho, \gamma\}$. Thus if the prediction is accurate, i.e., $e_{ij} = 0$, the regret decays exponentially fast with respect to $W$. This result matches with the conclusions in [24] [26]. Additionally, when there’s prediction error, the optimal choice of $W$ depends on the tradeoff between $E_d$ and the prediction errors.

For more insightful discussions, we consider non-decreasing $k$-step-ahead prediction errors, i.e. $\|B_d e_{t+1\ldots ij}\| \geq \|B_d e_{ij}\|$. It can be shown that Part I increases with $E_d$ and Part II increases with prediction errors. Further, as $W$ increases, Part I decreases but Part II increases. Thus when Part I dominates the regret bound, i.e. $E_d$ is large when compared with the prediction errors, selecting a large $W$ reduces the regret bound. On the contrary, when Part II dominates the regret bound, i.e. the prediction errors are large when compared with $E_d$, a small $W$ is preferred. The choice of $W$ above are quite intuitive: when the environment is perturbed by large disturbances but the disturbances could be roughly predicted, one should use more predictions to prepare for future changes; however, with poor prediction and mildly perturbed environments, one should ignore these long-term predictions whose quality tends to be low. Our regret upper bound provides a quantitative way to pick a balanced prediction window $W$.

### A. Proof Sketch

Due to the space limit, we only give a proof sketch of Theorem 1 here. The detailed proof is available in our online report [35]. Roughly speaking the proof can be decomposed into 4 steps:

- **Step 1**: We first introduce our general regret bound derived from ‘cost difference lemma’. Cost difference lemma is a very well-know lemma in reinforcement learning (RL) community and plays an important role in deriving many RL algorithms such as Trust Region Policy Optimization (TRPO [28]). In [29], they also use it to prove gradient dominance for standard, time-invariant LQR problem. However, previous results are mostly for time-invariant systems and we haven’t found an exact formula that applies the lemma to LQR problems with time-varying cost functions and disturbances. Thus we now state our result:

**Lemma 1. (General Regret Formula)** For an online control algorithm $\pi$ that satisfies:

$$u_t = \pi_t (x_t, \{d_t, Q_t, R_t\}_{\tau=1}^{\tau=1}, \{d_{i[t]} Q_{i[t]}, R_{i[t]}\}_{i=t}^{i=t+W}),$$

i.e., control action does not rely on history $\{x_t\}_{\tau=1}^{\tau=1}$, the regret can be written as:

$$\text{Regret}(\pi) = \sum_{i=1}^{T-1} (u_t^\tau - u_t^*)^\top (R_t + B_t^\top P_{t+1} B_t) (u_t^\tau - u_t^*)$$

where:

$$u_t^\tau = \pi_t (x_t^\tau, \{d_{\tau+t}, Q_{\tau+t}, R_{\tau+t}\}_{\tau=1}^{\tau=1}, \{d_{[i[t]]}, Q_{[i[t]]}, R_{[i[t]]}\}_{i=t}^{i=t+W}),$$

$$u_t^* = \pi_t^* (x_t^\tau, d_t, \ldots, d_{t-1})$$

and that $\{x_t^\tau\}$ is the trajectory generated by the online control algorithm $\pi$.

The proof of Lemma 1 is given in the Appendix. Lemma 1 successfully decouples the regret into the summation of a quadratic cost on ‘control action error’ $u_t^\tau - u_t^*$, so instead of analyzing $\text{Regret}(\pi)$ that entangles control action in all time step, we only need to bound the ‘control action error’ at each time step separately. Note Lemma 1 holds for policy $\pi$ that can be more general than MPC algorithm.

- **Step 2**: From (7) (11), the ‘control action error’ of MPC algorithm can be decomposed as:

$$u_t^\text{MPC} - u_t^* = \sum_{i=t-W+1}^{i=t-1} (K_{t,i} d_t - K_{t,i}^d d_i)$$

where:

$$K_{t,i} d_t = \sum_{i=t-W+1}^{i=t} (K_{t,i}^d - K_{t,i}^d d_i)$$

$$K_{t,i}^d = \sum_{i=t-W+1}^{i=t} (K_{t,i}^d - K_{t,i}^d d_i)$$

- **Step 3**: We name the term $\sum_{t-W+1=1}^{t} K_{t_i d_i}$ as ‘truncation error’, since this term appears when we throw away future disturbances after step $t + W$. The term $(K_{t,i} - K_{t,i}^d) x_t^\text{MPC}$ + $\sum_{i=t-W+1}^{i=t} (K_{t,i}^d - K_{t,i}^d d_i)$ is named ’matrices approximation error’ because this term will disappear if the controller has knowledge of all future $Q_t, R_t$’s. Note that in [26]’s setting, they study constant $Q_t = Q, R_t = R$, in which case matrices approximation error is zero and need not be taken into consideration. The term $\sum_{i=t-W+1}^{i=t} (K_{t,i} - K_{t,i}^d) e_{i[t]}$ is called ‘prediction error’ for the reason that it is related to the prediction accuracy at time step $t$.
Step 3: We’ll look into truncation error and prediction error first, because they are easier to analyse. It has been pointed out in Corollary 1 that $\|K_i^{d,i}\|$ decays exponentially w.r.t. $t - n$. Note that for truncation error $\sum_{t=W+1}^{t} K_t^{d,i} d_t$ the summation start from $t + W + 1$, thus roughly speaking this term is of order $O(1^{W+1})$.

As for prediction error, similar to Proposition 2 and Corollary 1, we can show that

$$\|\Phi(t+j; t+i)\| \leq \tau \rho_{i-j}, \ j \geq i$$

$$\|\Phi^{d,i}\| \leq \frac{\tau}{\lambda_{\max}(P_{\max})} \rho^{t-i}, \ i \geq t$$

Thus $\|\Phi_{d,i}\|$ is roughly of scale $\rho^{t-i}$, which suggests that the effect caused by prediction error decays exponentially as prediction goes to the far future.

Step 4: ‘Matrices approximation error’ is a little bit harder to handle, because we not only need to bound the error term $\|K_t - K_t^{d,i}\|, \|K_t^{d,i} - K_t^{d,i}\|$ but also need to make sure that $\{x_t^{MPC}\}$, which is the trajectory generated by the MPC algorithm, does not explode and become unbounded. The latter is guaranteed by the exponential stability of MPC algorithm, while bounding the error terms $\|K_t - K_t^{d,i}\|, \|K_t^{d,i} - K_t^{d,i}\|$ requires the contractivity of Ricatti iteration shown in (16). This is the point where factor $\gamma$ emerges in the regret.

Lemma 2.

$$\|K_t - K_t^{d,i}\| \leq \alpha_3 \gamma^W$$

$$\|K_t^{d,i} - K_t^{d,i}\| \leq \alpha_4 \gamma^{W-i} \rho^{t-i}, \ t \leq i \leq t + W$$

where,

$$\alpha_3 = 2 \frac{\|A\|_2}{\lambda_{\min}(R_{\min})} \frac{\lambda_{\max}(P_{\max})^4}{\lambda_{\min}(Q_{\min})^2} \left(\|B_u R_{\min}^{-1} B_u^T\| + 1\right) \frac{1}{1 - \gamma}$$

$$\alpha_4 = 2 \frac{\|B_u\|_2}{\lambda_{\min}(R_{\min})} \frac{\lambda_{\min}(Q_{\min})^2}{\lambda_{\min}(Q_{\min})^2} \left(\|B_u R_{\min}^{-1} B_u^T\| + 1\right) \frac{1}{1 - \gamma}$$

The key step in the proof of the Lemma 2 is to bound $\|P_{i+1|t} - P_{i+1}\|$ which is provided below

Lemma 3.

$$\|P_{i+1|t} - P_{i+1}\| \leq \gamma^{t+W-i} \frac{\lambda_{\max}(P_{\max})^2}{\lambda_{\min}(Q_{\min})}, \ t \leq i \leq t + W.$$

Lemma 3 is a direct corollary from (16). Note that

$$P_t = F_{Q_t, R_t}(P_{t+1}), \ P_{i+1|t} = F_{Q_t, R_t}(P_{i+1|t})$$

Applying (16) immediately gives us:

$$\delta_\infty(P_{i+1|t}, P_t) \leq \gamma \delta_\infty(P_{i+1|t+1}, P_{i+1}) \leq \cdots \leq \gamma^{t+W+1-i} \delta_\infty(P_{t+W+1|t+1}, P_{t+W+1})$$

which could be used to derive Lemma 3.

Combining all 4 steps together will give the upper bound in Theorem 1. The rigorous proof and computation are quite cumbersome and we refer readers to the Appendix of our online report [35] for more details.

Fig. 1: (Left): Regret for accurate prediction on $d_t$; (Right): Regret for accurate prediction on $d_t$, 'snr' represents the signal-to-noise ratio of the predictions.

V. NUMERICAL SIMULATIONS

To numerically test

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ B_u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ B_d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The stage cost function $Q_t, R_t$ are randomly picked as

$$Q_t = q_t I, \ R_t = r_t$$

where $q_t$’s are picked randomly from Unif[2,3], and $r_t$’s are picked from Unif[5,6]. Disturbances $d_t$ are drawn randomly from standard Gaussian distribution, $d_t \sim N(0,1)$. We consider the following two settings:

Accurate Prediction on $d_t$: The numerical result for accurate prediction case is shown in Figure 1 (Left). It displays the relationship of dynamical regret and prediction window size $W$. According to (15), when setting prediction errors $e_{ij} = 0$, the regret should be exponentially decreasing w.r.t. $W$, which matches the simulation result.

Noisy Prediction on $d_t$: The numerical result is shown in Figure 1 (Right). For this setting, we set the prediction errors as $e_{ij} \sim \text{snr} \times N(0,1)$, where $\text{snr}$ denotes the signal-to-noise ratio of the predictions. Smaller snr indicates more accurate predictions. The figure suggests that for more accurate predictions, choosing a fairly large prediction window will boost the performance of the controller; while for large snr values, a longer prediction does not benefit too much, because the prediction error term is the dominant term of the regret.

VI. CONCLUSION

This paper studies the role of predictions on dynamic regrets of online linear quadratic regulator with time-varying cost functions and disturbances. Besides giving an explicit dynamic regret upper bound for model predictive control algorithm in this setting, we also proposed a regret analysis framework based on ‘cost difference lemma’ that might be applicable for more general class of control algorithms than MPC. This paper leads to many interesting future directions, some of which are briefly discussed below. The first direction is to further allow noisy predictions on $Q_t, R_t$. The second direction is to consider time varying system parameters $A_t, B_{u,d}$, e.g. [11], [37]. Moreover, it might be interesting to consider the setting where system parameters are unknown, possibly by applying learning based control tools, e.g., [36], [38], [39].
REFERENCES

[1] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, Dynamic programming and optimal control. Athena scientific Belmont, MA, 1995, vol. 1, no. 2.
[2] K.-D. Kim and P. R. Kumar, “An mpc-based approach to provable system-wide safety and liveness of autonomous ground traffic,” IEEE Transactions on Automatic Control, vol. 59, no. 12, pp. 3341–3356, 2014.
[3] S. Kouro, P. Cortés, R. Vargas, U. Ammann, and J. Rodríguez, “Model predictive control—a simple and powerful method to control power converters,” IEEE Transactions on Industrial Electronics, vol. 56, no. 6, pp. 1825–1838, 2009.
[4] N. Li, L. Cazic, B. Boutilier, T. Lu, E. Wong, B. Roy, M. Ryu, and G. Inwalle, “Data center cooling using model-predictive control,” in Advances in Neural Information Processing Systems, 2018, pp. 3814–3823.
[5] M. Diehl, R. Amrit, and J. B. Rawlings, “A lyapunov function for economic optimizing model predictive control,” IEEE Transactions on Automatic Control, vol. 56, no. 3, pp. 703–707, 2010.
[6] M. Ellis, H. Durand, and P. D. Christofides, “A tutorial review of economic model predictive control methods,” Journal of Process Control, vol. 24, no. 8, pp. 1156–1178, 2014.
[7] R. Amrit, J. B. Rawlings, and D. Angeli, “Economic optimization using model predictive control with a terminal cost,” Annual Reviews in Control, vol. 35, no. 2, pp. 178–186, 2011.
[8] L. Grüne, “Economic receding horizon control without terminal constraints,” Automatica, vol. 49, no. 3, pp. 725–734, 2013.
[9] D. Angeli, R. Amrit, and J. B. Rawlings, “On average performance and stability of economic model predictive control,” IEEE transactions on automatic control, vol. 57, no. 7, pp. 1615–1626, 2012.
[10] L. Grüne and M. Stieler, “Asymptotic stability and transient optimality of economic mpc without terminal conditions,” Journal of Process Control, vol. 24, no. 8, pp. 1187–1196, 2014.
[11] L. Grüne and S. Pirkelmann, “Economic model predictive control for time-varying system: Performance and stability results,” Optimal Control Applications and Methods, vol. 41, no. 1, pp. 42–64, 2020.
[12] Y. Avni, D. Angeli, and A. Wierman, “Tracking adversarial targets,” in International Conference on Machine Learning, 2014, pp. 369–377.
[13] A. Cohen, A. Hasidim, T. Koren, N. Lazic, Y. Mansour, and K. Talwar, “Online linear quadratic control,” ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80, Stockholmsmässan, Stockholm Sweden: PMLR, 10–15 Jul 2018, pp. 1029–1038. [Online]. Available: http://proceedings.mlr.press/v80/cohen18b.html
[14] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, “Online control with adversarial disturbances,” ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97, Long Beach, California, USA: PMLR, 09--15 Jun 2019, pp. 111–119. [Online]. Available: http://proceedings.mlr.press/v97/agarwal19c.html
[15] E. Hazan, “Introduction to online convex optimization,” Found. Trends Optim., vol. 2, no. 3–4, p. 157–325, Aug. 2016. [Online]. Available: https://doi.org/10.1561/2400000003
[16] S. Shalev-Shwartz et al., “Online learning and online convex optimization,” Foundations and trends in Machine Learning, vol. 4, no. 2, pp. 107–194, 2011.
[17] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, “Online optimization: Competing with dynamic comparators,” in Artificial Intelligence and Statistics, 2015, pp. 398–406.
[18] A. Rakhlin and K. Sridharan, “Online learning with predictable sequences,” Journal of Machine Learning Research, vol. 30, 08 2012.
[19] N. Chen, A. Agarwal, A. Wierman, S. Barman, and L. L. Andrew, “Online convex optimization using predictions,” in Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, 2015, pp. 191–204.
[20] M. Badiei, N. Li, and A. Wierman, “Online convex optimization with ramp constraints,” in 2015 54th IEEE Conference on Decision and Control (CDC). IEEE, 2015, pp. 6730–6736.
[21] N. Chen, J. Comden, Z. Liu, A. Gandhi, and A. Wierman, “Using predictions in online optimization: Looking forward with an eye on the past,” ACM SIGMETRICS Performance Evaluation Review, vol. 44, no. 1, pp. 193–206, 2016.
[22] Y. Li, G. Qu, and N. Li, “Using predictions in online optimization with switching costs: A fast algorithm and a fundamental limit,” in 2018 Annual American Control Conference (ACC), 2018, pp. 3008–3013.
[23] G. Goel and A. Wierman, “An online algorithm for smoothed regression and lqr control,” Proceedings of Machine Learning Research, vol. 89, pp. 2504–2513, 2019.
[24] Y. Li, X. Chen, and N. Li, “Online optimal control with linear dynamics and predictions: Algorithms and regret analysis,” in Advances in Neural Information Processing Systems, 2019, pp. 14 887–14 899.
[25] Y. Li and N. Li, “Leveraging predictions in smoothed online convex optimization via gradient-based algorithms,” in Advances in Neural Information Processing Systems, 2020.
[26] C. Yu, G. Shi, S.-J. Chung, Y. Yue, and A. Wierman, “The power of predictions in online control,” 2020.
[27] G. Goel and B. Hassibi, “The power of linear controllers in lqr control,” arXiv preprint arXiv:2002.02574, 2020.
[28] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, “Trust region policy optimization,” in International conference on machine learning, 2015, pp. 1889–1897.
[29] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi, “Global convergence of policy gradient methods for the linear quadratic regulator,” in ICML, 2018.
[30] B. D. Anderson and J. B. Moore, Optimal control: linear quadratic methods. Courier Corporation, 2007.
[31] M. A. Müller and F. Allgöwer, “Economic and distributed model predictive control: Recent developments in optimization-based control,” SICE Journal of Control, Measurement, and System Integration, vol. 10, no. 2, pp. 39–52, 2017.
[32] A. Ferramosca, J. B. Rawlings, D. Limón, and E. F. Camacho, “Economic mpc for a changing economic criterion,” in 49th IEEE Conference on Decision and Control (CDC). IEEE, 2010, pp. 6131–6136.
[33] P. Falcone, F. Borrelli, E. Tseng, J. Asgari, and D. Hrovat, “Linear time-varying model predictive control and its application to active steering systems: Stability analysis and experimental validation,” International Journal of Robust and Nonlinear Control, vol. 18, pp. 862–875, 05 2008.
[34] D. Angeli, A. Casavola, and F. Tesedos, “Theoretical advances on economic model predictive control with time-varying costs,” Annual Reviews in Control, vol. 41, pp. 218–224, 2016.
[35] R. Zhang, Y. Li, and N. Li, On the Regret Analysis of Online LQR Control with Predictions, 2020. [Online]. Available: https://nali.seas.harvard.edu/files/nali/files/2021acconlinecontrol.pdf
[36] K. Krauth, S. Tu, and B. Recht, “Finite-time analysis of approximate policy iteration for the linear quadratic regulator,” in Advances in Neural Information Processing Systems, 2019, pp. 8512–8522.
[37] L. Grüne and S. Pirkelmann, “Closed-loop performance analysis for economic model predictive control of time-varying systems,” in 2017 IEEE 56th Annual Conference on Decision and Control (CDC). IEEE, 2017, pp. 5563–5569.
[38] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, “Regret bounds for robust adaptive control of the linear quadratic regulator,” in Advances in Neural Information Processing Systems, 2018, pp. 4188–4197.
[39] Y. Ouyang, M. Gagrani, and R. Jain, “Control of unknown linear systems with thompson sampling,” in 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2017, pp. 1198–1205.

APPENDIX

A. Proof of Proposition 1 and Lemma 1

We first define the optimal cost-to-go function:

\[ V^*_t(x) := \min_{\{u_t, \ldots, u_{T-1}\}} \sum_{k=t}^{T-1} \left( x_k^T Q x_k + u_k^T R_k u_k + x_{k+1}^T Q_{k+1} x_{k+1} \right) \mid x_t = x \]

s.t. \( x_{k+1} = Ax_k + Bu_k + B_k d_k \)

For simplicity we further define the following variables:

\[ A^{(n)} := \begin{bmatrix} I \\ A \\ y_t^T \end{bmatrix}, \quad y_t^{T+1} := \begin{bmatrix} y_t^1 \\ y_t^2 \\ y_t^{T+1} \end{bmatrix} \]
where \( y_i^{t:T} \in \mathbb{R}^{(T-t+1)n} \) and the recursive relationship between \( y_i^t \) satisfies:
\[
\begin{align*}
y_i^t &= x, \\
y_i^{t+1} &= Ay_i^t + B_a d_{t+1}, & i = 1, 2, \ldots, T - t.
\end{align*}
\]

Further define:
\[
y_i^{t+1:T}(x) := \begin{bmatrix} y_i^t \\ y_i^{t+1} \end{bmatrix}.
\]

By the definition of \( A^{(n)} \), \( y_i^{t+1} \) and \( y_i^{t+1:T} \), we have that:
\[
y_{t+1}^{t+1:T}(Ax + B_a u + B_d d_t) = y_i^{t+1:T}(x) + A^{(T-t)} B_a u.
\]

Additionally, by the definition of \( V_i^* \), we have the following Bellman equation:
\[
V_i^*(x) = \min_{u} x^T Q_t x + u^T R_t u + V_{t+1}^*(Ax + B_a u + B_d d_t).
\]

The optimal policy is given by:
\[
u_i^* = -G_i y_i^{t+1:T}(x),
\]
where
\[
G_i := (R_t + B_a^T A^{(T-t)} A^{(T-t)} + 1 B_a^T A^{(T-t)} V_{t+1}^T y_i^{t+1:T}).
\]

Proof. We use proof of induction. It is quite obvious that \( V_{2}^*(x) = x^T Q_2 x \). Suppose proposition holds for time step \( t + 1 \), i.e.
\[
V_{t+1}^*(x) = y_i^{t+1:T}(x) V_{t+1}^T y_i^{t+1:T}(x).
\]

Here for the sake of simplicity we write \( y_i^{t+1:T} (x; d_1, \ldots, d_{T-t}) \) as \( y_i^{t+1:T}(x) \).

By Bellman equation (19) we know that
\[
V_i^*(x) = \min_{u} x^T Q_t x + u^T R_t u + V_{t+1}^*(Ax + B_a u + B_d d_t)
\]
\[
= \min_{u} x^T Q_t x + u^T R_t u + y_i^{t+1:T}(Ax + B_a u + B_d d_t)
\]
\[
+ (y_i^{t+1:T}(x) + A^{(T-t)} B_a u)^T V_{t+1}^T y_i^{t+1:T}(x) + A^{(T-t)} B_a u.
\]

Minimizing over \( u \) we get that:
\[
u_i^* =
- (R_t + B_a^T A^{(T-t)} V_{t+1}^T y_i^{t+1:T})^{-1} B_a^T A^{(T-t)} V_{t+1}^T y_i^{t+1:T}(x)
\]
\[(22)\]

Substituting (22) into the Bellman equation we get:
\[
V_i(x) = x^T Q_t x + y_i^{t+1:T}(x) f_t(V_{t+1}) y_i^{t+1:T}(x)
\]
\[
= y_i^{t+1:T}(x) ^T \begin{bmatrix} Q_t & 0 \\ 0 & f_t(V_{t+1}) \end{bmatrix} y_i^{t+1:T},
\]
which completes the proof.

The next proposition shows that Proposition 4 is consistent with classical LQR results.

Proposition 5. (Consistency with Standard LQR) Define \( P_t \in \mathbb{R}^{n \times n} \) as:
\[
P_t := A^{(T-t+1)} V_{t+1} A^{(T-t+1)}.
\]

Then we have:
\[
P_t = Q_t + A^T P_{t+1} A - A^T P_{t+1} B_a (R_t + B_a^T P_{t+1} B_a)^{-1} B_a^T P_{t+1} A,
\]
which is the same matrix as the value function matrix for LQR defined in (4).

We omit the proof for Proposition 5, since it can be done by simple algebraic manipulation.

We are now ready to prove Proposition 1. Though there are some very similar results in literature, we’ll give our own proof for the optimal offline controller. Note that we have already shown in Proposition 4 that the optimal policy is given by (21), thus we only need to show that (7) and (21) are equivalent.

The proof of Proposition 1 depends on the following lemma:

Lemma 4. Define \( Y_t \in \mathbb{R}^{(T-t+1)n \times n} \) as \( Y_T := V_T A^{(T-t)} \), then \( Y_t \) has an explicit form:
\[
Y_t = \begin{bmatrix} Q_t \\ Q_{t+1} (A - B_a K_t) \\ Q_{t+2} (A - B_a K_{t+1}) (A - B_a K_t) \\ \vdots \\ Q_T (A - B_a K_{T-1}) \cdots (A - B_a K_t) \end{bmatrix},
\]
where \( K_t \) is the optimal control gain for standard LQR problem as defined in (5).

Proof. From eq(20) and Proposition 5, with some simple algebraic computation we can derive the recursive relationship between \( Y_t \):
\[
Y_t = \begin{bmatrix} Y_{t+1} A - Y_t B_a (R_t + B_a^T P_{t+1} B_a)^{-1} B_a^T P_{t+1} A \\ Q_t \\ Y_{t+1} (A - B_a K_t) \end{bmatrix}
\]
\[(24)\]
Applying this relationship recursively we can get that:

\[
Y_t = \begin{bmatrix}
Q_t \\
Q_{t+1}(A - B_uK_t) \\
Q_{t+2}(A - B_uK_{t+1})(A - B_uK_t) \\
\vdots \\
Q_T(A - B_uK_{T-1}) \cdots (A - B_uK_t)
\end{bmatrix}
\]

which completes the proof.

Proof. (Proposition 1) According to (21),

\[
u_t = -G_1y_{t+1:T}(x_t)
= -(R_t + B_u^T P_{t+1} B_u)^{-1} B_u^T \begin{bmatrix}
A^{(T-t)} V_{t+1} y_{t+1:T}(x_t) \\
\vdots \\
A^{(T-t)} V_{T} y_{T}(x_t)
\end{bmatrix}
= -(R_t + B_u^T P_{t+1} B_u)^{-1} B_u^T Y_{t+1}^\top A x_t
+ A^{(T-t)} B_d d_t + \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix} B_d d_{T-1}
+ \cdots + \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix} B_d d_{T-1}
\]

Since \(Y_{t+1} = V_{t+1} A^{(T-t)}\), we have that:

\[
Y_{t+1}^\top A^{(T-t)} = P_{t+1}
\]

Additionally, from (24) we have:

\[
Y_{t+1}^\top \begin{bmatrix}
A^{(T-t-1)} \\
\vdots \\
0
\end{bmatrix} = \begin{bmatrix}
Q_{t+1} (A - B_uK_{t+1})^\top Y_{t+2} \\
\vdots \\
0
\end{bmatrix} \begin{bmatrix}
0 \\
A^{(T-t-1)} \\
\vdots \\
A^{(T-t-k)}
\end{bmatrix}
= (A - B_uK_{t+1})^\top \cdots (A - B_uK_{t+k})^\top P_{t+k+1}
= \Phi(t + k + 1, t + 1)^\top P_{t+k+1}
\]

Thus we have that:

\[
u_t = -(R_t + B_u^T P_{t+1} B_u)^{-1} B_u^T Y_{t+1}^\top A x_t
+ A^{(T-t)} B_d d_t + \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix} B_d d_{T-1}
+ \cdots + \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix} B_d d_{T-1}
\]

which completes the proof.

Define the value function for a given policy \(\pi\):

\[
V_t^\pi(x) := \sum_{k=t}^{T-1} (x_k^\top Q_k x_k + u_k^\top R_k u_k) + x_T^\top Q_T x_T
\]

where \(I_k\) denotes the information available for the controller at time step \(k\). Specifically for MPC considered in this paper, \(I_k = \{x_k, \{d_{i|t}, Q_t, R_t\}_{k+1}^{k+W}\}\).

Similarly we could also define the \(Q\)-function:

\[
Q_t^\pi(x, u) := \sum_{k=t}^{T-1} (x_k^\top Q_k x_k + u_k^\top R_k u_k) + x_T^\top Q_T x_T
\]

where \(\{x_t^\pi, u_t^\pi\}\) are trajectory generated by starting at initial state \(x\) and imposing policy \(\pi\).

We state a general version of cost difference lemma.

Lemma 5. (Cost Difference Lemma) For two policies \(\pi_1, \pi_2\), the difference of their regrets can be represented by:

\[
V_t^{\pi_2}(x) - V_t^{\pi_1}(x) = \sum_{i=1}^{T-1} Q_t^{\pi_1}(x_t^{\pi_2}, u_t^{\pi_2}) - V_t^{\pi_1}(x_t^{\pi_2})
\]

where \(\{x_t^{\pi_1}, u_t^{\pi_2}\}\) are trajectory generated by starting at initial state \(x\) and imposing policy \(\pi_2\).

The proof of Lemma 5 can be found in literature [28], [29]. By applying Lemma 5, we get the proof of Lemma 1.

Proof: (Lemma 1)

\[
\text{Regret}(\pi) = V_t^{\pi_1}(x) - V_t^{\pi}(x) = \sum_{i=1}^{T-1} Q_t^{\pi_1}(x_t^{\pi_1}, u_t^{\pi_1}) - V_t^{\pi_1}(x_t^{\pi_1})
= \sum_{i=1}^{T-1} \left[ x_t^{\pi_1^\top} Q_i x_t^{\pi_1} + u_t^{\pi_1^\top} R_i u_t^{\pi_1} + V_{i+1}^{\pi_1} (A x_t^{\pi_1} + B u_t^{\pi_1} + B d_i) \right]
- \left[ x_t^{\pi_1^\top} Q_i x_t^{\pi_1} - u_t^{\pi_1^\top} R_i u_t^{\pi_1} - V_{i+1}^{\pi_1} (A x_t^{\pi_1} + B u_t^{\pi_1} + B d_i) \right]
= \sum_{i=1}^{T-1} \left[ u_t^{\pi_1^\top} R_i u_t^{\pi_1} + \left( y_t^{i+1:T} (x_t^{\pi_1})^\top + A^{(T-t)} B u_t^{\pi_1} \right)^\top V_{i+1}^{\pi_1} \left( y_t^{i+1:T} (x_t^{\pi_1}) + A^{(T-t)} B u_t^{\pi_1} \right) \right]
- \left[ u_t^{\pi_1^\top} R_t u_t^{\pi_1} - \left( y_t^{i+1:T} (x_t^{\pi_1})^\top + A^{(T-t)} B u_t^{\pi_1} \right) \right]
= \sum_{i=1}^{T-1} \left[ (u_t^{\pi_1} - u_t^{\pi_1})^\top (R_t + B_u^T P_{t+1} B_u)(u_t^{\pi_1} - u_t^{\pi_1}) \right]
+ 2(u_t^{\pi_1} - u_t^{\pi_1})^\top (R_t + B_u^T P_{t+1} B_u)(u_t^{\pi_1} + G_t y_t^{i+1:T} (x_t^{\pi_1}))
= \sum_{i=1}^{T-1} \left( u_t^{\pi_1} - u_t^{\pi_1} \right)^\top (R_t + B_u^T P_{t+1} B_u)(u_t^{\pi_1} - u_t^{\pi_1})
B. Proof of Theorem 1

In this section, we are going to give a rigorous proof for our main result Theorem 1. Proof of sketch in the main text has already given a clear structure how it is proved, what is left is purely tedious calculation and analysis. The major difficulty arises in the state trajectory generated by the MPC algorithm \( x_{t}^{MPC} \) showing up in (18). We need to rewrite \( x_{t}^{MPC} \) in terms of initial state \( x_{1} \), disturbances \( d_{i} \) and prediction errors \( e_{i,j} \), which makes the analysis quite tedious, but the structure should be clear. Thus for readers who are only looking for an intuitive understanding of the proof, we would suggest to first skip some of the detailed calculation.

As stated in the previous paragraph, the first step towards the proof is to rewrite \( x_{t}^{MPC} \) in terms of the initial state, disturbances and prediction errors, which is stated in the following proposition.

**Proposition 6.** Let \( \{x_{t}^{MPC}\} \) be the trajectory generated by MPC algorithm defined as in Eq (11), then:

\[
x_{t}^{MPC} = M_{t}[t]B_{d}d_{i} + \min_{t+1 \leq i \leq T} M_{t}[i]B_{d}d_{i} + \sum_{i=1}^{t-1} M_{i,[t]}B_{d}d_{i} - \sum_{j=1}^{t-1} \sum_{i=j} M_{j,[t]}B_{d}d_{i} \tag{26}
\]

where:

\[
M_{i,[t]} = \sum_{j=\max\{1,i-W\}}^{\min\{t-1,i\}} \Phi_{t+j}^{MP}(t+j+1) (-B_{u}R_{j,i}^{d,j} + 1 \{i=j\})
\]

Additionally,

\[
\|M_{i,[t]}\| \leq c_{4} \rho^{t-i-1}, 1 \leq i \leq t-1
\]

\[
\|M_{i,[t]}\| \leq c_{5} \rho^{t-i+1}, i \geq t
\]

where,

\[
c_{4} = \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\tau^{2} \frac{1}{1 - \rho^{2}} + \tau
\]

\[
c_{5} = \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\tau^{2} \frac{1}{1 - \rho^{2}}
\]

**Proof.** Recall from (11) - (14) the MPC scheme, \( x_{t}^{MPC} \) can be written as:

\[
x_{t}^{MPC} = Ax_{t-1}^{MPC} + B_{u}x_{t-1}^{MPC} + B_{d}d_{t-1}
\]

\[
= (A-B_{u}\bar{K}_{t-1})x_{t-1}^{MPC} - B_{u}\bar{K}_{t-1}^{d,t}B_{d}d_{t-1} - \cdots - B_{u}\bar{K}_{t-W}^{d,t-W}B_{d}d_{t-W} + \cdots
\]

\[
= (A-B_{u}\bar{K}_{t-1})x_{t-1}^{MPC} - B_{u}\bar{K}_{t-W}^{d,t-W}B_{d}d_{t-W} + \cdots - B_{u}\bar{K}_{t-1}^{d,t}B_{d}d_{t-1} - \cdots - B_{u}\bar{K}_{t-W}^{d,t-W}B_{d}d_{t-W} + \cdots
\]

\[
= (A-B_{u}\bar{K}_{t-1})x_{t-1}^{MPC} - (B_{u}\bar{K}_{t-1}^{d,t-1} + 1)B_{d}d_{t-1} - \cdots - B_{u}\bar{K}_{t-W}^{d,t-W}B_{d}d_{t-W} + \cdots
\]

Applying the above equation iteratively we get (26).

We now look at the bound on \( \|M_{i,[t]}\| \). For \( 1 \leq i \leq t-1 \),

\[
M_{i,[t]} = \sum_{j=\max\{1,i-W\}}^{\min\{t-1,i\}} \Phi_{t+j}^{MP}(t+j+1) (-B_{u}\bar{K}_{j,i}^{d,j} + 1 \{i=j\}) + \Phi_{t-1,i}^{MP}(t-1,i)
\]

Proposition 3 has already showed that \( \|\Phi_{t}^{MP}(t,t_{0})\| \leq \tau \rho^{t-t_{0}} \). Thus:

\[
\|M_{i,[t]}\| \leq \sum_{j=\max\{1,i-W\}}^{t-1} \tau \rho^{t-j} \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\rho^{j} \tau^{1-i} + \tau \rho^{t-j-1} \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\rho^{t-j} + \tau \rho^{t-j-1}
\]

\[
= \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\tau^{2} \frac{1}{1 - \rho^{2}} + \tau \rho^{t-1-i} + \tau \rho^{t-j-1} \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\tau^{2} \frac{1}{1 - \rho^{2}}
\]

For \( i \geq t \),

\[
M_{i,[t]} = \sum_{j=\max\{1,i-W\}}^{t-1} \Phi_{t+j}^{MP}(t+j+1)(-B_{u}\bar{K}_{j,i}^{d,j})
\]

\[
\leq \sum_{j=\max\{1,i-W\}}^{t-1} \tau \rho^{t-j} \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\rho^{j} \tau^{t-j+1} + \tau \rho^{t-j} \|B_{u}R_{\min}^{-1}\|B_{d}^{T}\|\lambda_{\max}(P_{\max})\tau^{2} \frac{1}{1 - \rho^{2}} \rho^{j} \tau^{t-j+1}
\]

\[
= c_{5} \rho^{t-i+1}
\]

Now that \( x_{t}^{MPC} \) is re-written in terms of \( x_{1}, d_{i}, e_{i,j} \), the next step is to write the `control action error` \( u_{t}^{MPC} - u_{t} \) in these terms as well.

**Proposition 7.**

\( u_{t}^{MPC} - u_{t} = N_{t+1,i}x_{t+1} + \sum_{i=1}^{t-1} N_{i,i}B_{d}d_{i} + \sum_{i=j}^{t+W} L_{i,j}B_{d}e_{i,j} \)

where:

\[
N_{t,i} = \begin{cases} (K_{t} - \bar{K}_{t})\Phi_{t}^{MP}(t,1) & i = 0 \\ (K_{t} - \bar{K}_{t})M_{i,t} & 1 \leq i \leq t-1 \\ (K_{t} - \bar{K}_{t})M_{i,t} + (K_{d,i}^{d,i} - \bar{K}_{t}^{d,i}) & t \leq i \leq t+W \\ (K_{t} - \bar{K}_{t})M_{i,t} + K_{d,i}^{d,i} & i \geq t+W + 1 \end{cases}
\]

\[
L_{i,j} = \begin{cases} -\bar{K}_{t}^{d,i} & i = j-1, i \geq j \\ \Phi_{t}^{MP}(t,j+1)B_{u}\bar{K}_{j,i}^{d,j} & j \leq t-1, i \geq j \\ \bar{K}_{t}^{d,i} & 1 \leq i \leq t-1, j = t \end{cases}
\]

**Proof.**

\[
u_{t}^{MPC} - u_{t}^{*} = (K_{t} - \bar{K}_{t})x_{t}^{MPC} + \sum_{i=t}^{t+W} (K_{d,i}^{d,i} - \bar{K}_{d,i}^{d,i})d_{i}
\]

\[
+ \sum_{i=t+W}^{t+W+1} K_{d,i}^{d,i}d_{i} - \sum_{i=t}^{t+W} \Phi_{t}^{MP}(t,t_{0})e_{i,t_{0}}^{MPC}
\]
\begin{align*}
(K_t - \bar{K}_t)x_t + &\sum_{i=t}^{t+W} (K_t^{d,i} - \bar{K}_t^{d,i})d_i + \sum_{i=t+1}^{T-1} K_t^{d,i}d_i \\
= &\ (K_t - \bar{K}_t)\bar{U}(t, 1)x_1 + \sum_{i=t}^{t+W} (K_t - \bar{K}_t)M_{i|t}B_dd_i \\
- &\sum_{j=t}^{t+W} \sum_{i=j} \left\{ (K_t - \bar{K}_t)\bar{U}(t, j + 1)B_u\bar{K}^{d,i}_j e_{i|j} \\
+ &\sum_{i=t}^{t+W} (K_t^{d,i} - \bar{K}_t^{d,i})B_dd_i + \sum_{j=1}^{T-1} K_t^{d,i}d_i \\
= &N_0|x_1 + \sum_{i=1}^{T-1} N_{i|t}B_dd_i + \sum_{j=1}^{T-1} \sum_{i=j} L_{(i,j)|t}B_dd_{i|j}
\end{align*}

Combining the above two equations completes the proof.

The next proposition bounds the norms of the matrices appeared in Proposition 7.

**Proposition 8.**

\[
\norm{N_{i|t}} \leq c_1^W \rho^{t-1} \\
\norm{N_{i|t}} \leq c_2^W \rho^{i-1}, 1 \leq i \leq t-1 \\
\norm{N_{i|t}} \leq c_3^W \rho^{i+t}, t \leq i \leq t + W \\
\norm{L_{(i,j)|t}} \leq c_4^W \rho^{2j-1}, j \leq t - 1 \\
\norm{L_{(i,t)|t}} \leq c_5^W \rho^{t-1}
\]

where the constant \(c\) can be expressed by:

\[
\begin{align*}
\norm{K_t^{d,i}} &= c_1^W \rho^{i-t}, i \geq t \\
\norm{K_t - \bar{K}_t} &= c_2^W \\
\norm{K_t^{d,i} - \bar{K}_t^{d,i}} &= c_3^W \rho^{i-t} \\
\norm{M_{i|t}} &= c_4^W \rho^{i-t-1}, 1 \leq i \leq t-1 \\
\norm{M_{i|t}} &= c_5^W \rho^{i+t}, i \geq t
\end{align*}
\]

**Proof.** From Corollary 1, Lemma 2, 6 we have that:

\[
\begin{align*}
\norm{K_t^{d,i}} &= c_1^W \rho^{i-t}, i \geq t \\
\norm{K_t - \bar{K}_t} &= c_2^W \\
\norm{K_t^{d,i} - \bar{K}_t^{d,i}} &= c_3^W \rho^{i-t} \\
\norm{M_{i|t}} &= c_4^W \rho^{i-t-1}, 1 \leq i \leq t-1 \\
\norm{M_{i|t}} &= c_5^W \rho^{i+t}, i \geq t
\end{align*}
\]

For \(1 \leq i \leq t-1,\)

\[
\begin{align*}
\norm{N_{i|t}} &= \norm{K_t - \bar{K}_t} \norm{M_{i|t}} \\
&\leq c_2^W c_4^W \rho^{i-t-1} \leq c_7^W \rho^{i-t-1}
\end{align*}
\]

For \(t \leq i \leq t + W,\)

\[
\begin{align*}
\norm{N_{i|t}} &= \norm{K_t - \bar{K}_t} \norm{M_{i|t}} + \norm{K_t^{d,i} - \bar{K}_t^{d,i}} \\
&\leq c_2^W c_4^W \rho^{i-t-1} + c_3^W \rho^{i-t} \\
&\leq \rho c_2^W \rho^{i-t} \leq c_8^W \rho^{i-t} \\
&\leq \rho^i \rho^{i-1} \leq c_9^W \rho^{i-t-1}
\end{align*}
\]

For \(i \geq t + W + 1,\)

\[
\begin{align*}
\norm{N_{i|t}} &= \norm{(K_t - \bar{K}_t)M_{i|t}} + \norm{K_t^{d,i}} \\
&\leq c_2^W c_4^W \rho^{i-t-1} + c_3^W \rho^{i-1} \\
&\leq \rho c_2^W \rho^{i-1} \leq c_9^W \rho^{i-t-1}
\end{align*}
\]

For \(j \leq t - 1, i \geq j\)

\[

\]

For \(i \geq t,\)

\[

\]

We are now ready to give a rigorous proof for Theorem 1. According to Lemma 1, the regret is a quadratic sum over the ‘control action error’ \((u^{\text{MPC}}_t - u^*_t)\). Most of the part of the proof below is to translate this quadratic sum over the ‘control action error’ to a quadratic sum over \(x_1, d_1, e_{i|j}\), which serves as another main source of cumbersome calculation.

**Proof.** (Theorem 1) Applying the regret formula in Lemma 1, we have that

\[
\text{Regret}(\pi) = V^\pi_1(x) - V^*_1(x) \\
= \sum_{i=1}^{T-1} (u^{\text{MPC}}_i - u^*_i)R_t + \sum_{i=1}^{T-1} (u^{\text{MPC}}_i - u^*_i)B_uK^{i,j}_t \\
\leq \norm{R}^\max + \norm{B_u}^\max \sum_{i=1}^{T-1} \norm{u^{\text{MPC}}_i - u^*_i}^2 \\
\leq \sum_{i=1}^{T-1} \norm{N_{i|t}x_1 + \sum_{i=1}^{T-1} N_{i|t}B_dd_i + \sum_{i=1}^{T-1} \sum_{i=1}^{T-1} L_{(i,j)|t}B_dd_{i|j}}^2 \\
\leq 2\norm{N_{i|t}x_1 + \sum_{i=1}^{T-1} N_{i|t}B_dd_i}^2 + 2\sum_{i=1}^{T-1} \sum_{i=1}^{T-1} \norm{L_{(i,j)|t}B_dd_{i|j}}^2
\]

We first bound the term:

\[
\sum_{i=1}^{T-1} \sum_{i=1}^{T-1} \norm{N_{i|t}x_1 + \sum_{i=1}^{T-1} N_{i|t}B_dd_i}^2 \\
\leq \sum_{i=1}^{T-1} \norm{N_{i|t}} \norm{x_1} + \sum_{i=1}^{T-1} \norm{N_{i|t}} \norm{B_dd_i}^2 \\
\leq \sum_{i=1}^{T-1} \norm{a_{i|t}^*x_1} + \sum_{i=1}^{T-1} \norm{a_{i|t}^*B_dd_i} \\
\leq \sum_{i=1}^{T-1} \norm{N_{i|t}x_1 + \sum_{i=1}^{T-1} N_{i|t}B_dd_i}
\]

The last inequality holds for the reason that Proposition 8 implies

\[
\norm{N_{i|t}} \leq c \cdot a_{i|t}^*
\]

where

\[
a_{i|t}^* = \gamma^W \min_{i} \norm{(0, i-t, W)} \rho^\max \{i-t-i-1\}
\]

Applying Lemma 8 in Appendix G, we have

\[
\sum_{i=1}^{T-1} \sum_{i=1}^{T-1} \norm{N_{i|t}x_1 + \sum_{i=1}^{T-1} N_{i|t}B_dd_i}
\]


Thus we have that:

\[
\sum_{t=1}^{T} \|N_{0|i}x_1 + \sum_{i=1}^{T-1} N_{i|i} B_{d} d_{i}\|^2 \leq c^2 \left( 1 + \frac{\gamma W}{1 - \rho} \right)^2 \left( \|x_1\|^2 + \sum_{t=1}^{T} \|B_{d} d_{t}\|^2 \right).
\]

We now bound

\[
\sum_{t=1}^{T-1} \sum_{j=1}^{T-1} L_{(i,j)|t} \|B_{d} e_{i|j}\|^2 \leq c^2 \left( \frac{1}{1 - \rho} + \frac{\gamma W}{1 - \rho^2} \right) \left( \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \|B_{d} e_{ij}\|^2 \right)^2,
\]

where

\[
\alpha_{t,j} = \begin{cases} 
\rho^{t-2j+i-1} \gamma W, & j \leq t - 1 \\
\rho^{t-i}, & j = t \\
0, & j \geq t + 1 
\end{cases}
\]

Applying Lemma 8 we have that:

\[
\sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \sum_{i=j}^{T-1} \sum_{i',j'} \alpha_{t,j} \alpha_{t',j'} \|B_{d} e_{i|j}\|^2 \leq c^2 \left( \frac{1}{1 - \rho} + \frac{\gamma W}{1 - \rho^2} \right) \left( \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \|B_{d} e_{ij}\|^2 \right)^2.
\]

For fixed \(t\),

\[
\sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \sum_{i=j}^{T-1} \sum_{i'=j}^{T-1} \alpha_{t,j} \alpha_{t',j'} \|B_{d} e_{i|j}\|^2 \leq c^2 \frac{1}{1 - \rho} \left( 1 + \frac{\gamma W}{1 - \rho} \right) \left( \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \|B_{d} e_{ij}\|^2 \right)^2
\]

For fixed \(i, j\),

\[
\sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \sum_{i=j}^{T-1} \sum_{i'=j}^{T-1} \alpha_{t,j} \alpha_{t',j'} \|B_{d} e_{i|j}\|^2 \leq c^2 \left( \frac{1}{1 - \rho} + \frac{\gamma W}{1 - \rho^2} \right) \left( \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \|B_{d} e_{ij}\|^2 \right)^2.
\]

Thus,

\[
\|x_1\|^2 + \sum_{t=1}^{T} \|B_{d} d_{t}\|^2 \leq c^2 \left( \frac{1}{1 - \rho} + \frac{\gamma W}{1 - \rho^2} \right) \left( \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \|B_{d} e_{ij}\|^2 \right)^2.
\]

Combining the above inequalities we have that:

\[
\text{Regret}(\pi) \leq c^2 \left( \frac{1}{1 - \rho} + \frac{\gamma W}{1 - \rho^2} \right) \left( \sum_{t=1}^{T} \|B_{d} d_{t}\|^2 \right)^2 + c^2 \left( \frac{1}{1 - \rho} + \frac{\gamma W}{1 - \rho^2} \right) \left( \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \|B_{d} e_{ij}\|^2 \right),
\]

which completes the proof.
C. Proof of Lemma 2

We define the following auxiliary variables:

\[ X_t^{d,s} = \Phi(s + 1, t + 1) \top P_{s+1}, \ s \geq t \]
Similarly, we can define \( \bar{X}_{r[t]}^{d,s} \):

\[ \bar{X}_{r[t]}^{d,s} = \bar{\Phi}_s(s + 1, r + 1) \top \bar{P}_{s+1[t]} \]

It is not hard to verify that

\[ K_t^{d,s} = (R_t + B_u^r \bar{P}_{r+1[t]}B_u) - B_u^r X_t^{d,s} \]
\[ \bar{K}_t^{d,s} = (R_t + B_u^r \bar{P}_{r+1[t]}B_u) - B_u^r \bar{X}_{r[t]}^{d,s} \]

The following proposition relates the difference of \( \bar{X}_{r[t]}^{d,s} - X_t^{d,s} \) to \( \bar{P}_{s[t]} - P_s \).

**Proposition 9.** For \( 1 \leq i \leq W \),

\[ \bar{X}_{r[t]}^{d,s} - X_t^{d,s} = \Phi(s + 1, t + 1) \top (\bar{P}_{s+1[t]} - P_{s+1}) \]
\[ - \sum_{r = t+1}^{s} \Phi(r + 1, t + 1) \top (\bar{P}_{r+1[t]} - P_{r+1}) B_u \]
\[ (R_t + B_u^r \bar{P}_{r+1[t]}B_u) - B_u^r \bar{X}_{r[t]}^{d,s} \]

**Proof.**

\[ \bar{X}_{r[t]}^{d,s} = (A - B_u \bar{K}_{r+1[t]}^\top) \times X_{r+1[t]}^{d,s} \]
\[ X_t^{d,s} = (A - B_u K_{r+1}^\top) \times X_{r+1}^{d,s} \]

Thus,

\[ X_t^{d,s} - K_t^{d,s} = ((A - B_u \bar{K}_{r+1[t]}^\top) - (A - B_u K_{r+1}^\top)) X_{r+1[t]}^{d,s} \]
\[ + (A - B_u K_{r+1}^\top) \times (X_{r+1[t]}^{d,s} - X_{r+1}^{d,s}) \]
\[ = -(K_{r+1[t]} - K_{r+1}^\top) B_u^r X_{r+1[t]}^{d,s} \]
\[ + (A - B_u K_{r+1}^\top) \times (X_{r+1[t]}^{d,s} - X_{r+1}^{d,s}) \]
\[ = \cdots \]
\[ = - \sum_{r = t+1}^{s} \Phi(r, t + 1) (K_{r[t]} - K_r) \top B_u^r X_{r[t]}^{d,s} \]
\[ + \Phi(s + 1, t + 1) \top (X_{s+1[t]}^{d,s} - X_{s}^{d,s}) \]

Since,

\[ K_{r[t]} - K_r = (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r \bar{P}_{r+1[t]} A \]
\[ - (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r (P_{r+1[t]} - P_r) A \]
\[ = (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r (P_{r+1[t]} - P_r) A \]
\[ + B_u^r (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r A \]
\[ + B_u (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r (P_{r+1[t]} - P_r) A \]
\[ = (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r (P_{r+1[t]} - P_r) A \]
\[ - (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r (P_{r+1[t]} - P_r) A \]

Substitute the equation into the previous equation, we get

\[ \bar{X}_{r[t]}^{d,s} - X_t^{d,s} = \Phi(s + 1, t + 1) \top (\bar{P}_{s+1[t]} - P_{s+1}) \]
\[ - \sum_{r = t+1}^{s} \Phi(r + 1, t + 1) \top (\bar{P}_{r+1[t]} - P_{r+1}) B_u \]
\[ (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r \bar{X}_{r[t]}^{d,s} \]

which completes the proof. \( \square \)

**Proposition 10.**

\[ \| \bar{X}_{r[t]}^{d,s} - X_t^{d,s} \| \leq \frac{\lambda_{\max}(P_{\max})}{(1 - \gamma)\lambda_{\min}(P_{\min})^2} \left( \| B_u R_{\min}^{-1} B_u^\top \| + 1 \right) \rho^{s-t} \gamma^{W-s+t} \]

**Proof.** From Proposition 9, we have that

\[ \| \bar{X}_{r[t]}^{d,s} - X_t^{d,s} \| \leq \| \Phi(s + 1, t + 1) \top \| \left( \| \bar{P}_{s+1[t]} - P_{s+1} \| \right) \]
\[ + \sum_{r = t+1}^{s} \| \Phi(r + 1, t + 1) \top \| \left( \| \bar{P}_{r+1[t]} - P_{r+1} \| \right) \]
\[ \| B_u (R_r + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^\top \| \| \bar{X}_{r[t]}^{d,s} \| \]

It is easy to show from definition that

\[ \| \bar{X}_{r[t]}^{d,s} \| \leq \lambda_{\max}(P_{\max}) \cdot \rho^{s-t} \gamma^{W-s+t} \]

Thus, by applying Proposition 12, we have that

\[ \| \bar{X}_{r[t]}^{d,s} - X_t^{d,s} \| \leq \| \Phi(s + 1, t + 1) \top \| \| \bar{P}_{s+1[t]} - P_{s+1} \| \]
\[ + \sum_{r = t+1}^{s} \| \Phi(r + 1, t + 1) \top \| \| \bar{P}_{r+1[t]} - P_{r+1} \| \]
\[ \| B_u R_{\min}^{-1} B_u^\top \| \| \bar{X}_{r[t]}^{d,s} \| \]

Thus, we can conclude that

\[ \| \bar{X}_{r[t]}^{d,s} - X_t^{d,s} \| \leq \| \Phi(s + 1, t + 1) \top \| \left( \| \bar{P}_{s+1[t]} - P_{s+1} \| \right) \]
\[ + \sum_{r = t+1}^{s} \| \Phi(r + 1, t + 1) \top \| \left( \| \bar{P}_{r+1[t]} - P_{r+1} \| \right) \]
\[ \| B_u R_{\min}^{-1} B_u^\top \| \| \bar{X}_{r[t]}^{d,s} \| \]

which proves the Proposition 10. \( \square \)

We are now ready to prove Lemma 2

**Proof.** Lemma 2

\[ K_t^{d,s} = (R_t + B_u^r P_{r+1[t]}B_u)^{-1} B_u^r X_t^{d,s} \]
\[ \bar{K}_t^{d,s} = (R_t + B_u^r \bar{P}_{r+1[t]}B_u)^{-1} B_u^r \bar{X}_{t}^{d,s} \]

Thus, we can conclude that

\[ \| \bar{K}_t^{d,s} - K_t^{d,s} \| \leq \frac{\| B_u \| L}{\lambda_{\min}(R_{\min})^2} \| \bar{X}_{t}^{d,s} - X_t^{d,s} \| \]
\[ + \frac{\| B_u \| L}{\lambda_{\min}(R_{\min})^2} \| \bar{P}_{t+1[t]} - P_{t+1} \| \| B_u R_{\min}^{-1} B_u^\top \| \| X_t^{d,s} \| \]

which completes the proof. \( \square \)
which completes the proof.

Proposition 12. Given two sequences 
\( \{Q_i, R_t\}_{t=1}^T \) that satisfy Assumption 2, and that:
\[ Q_t = Q, \quad R_t = R, \quad i = t, t+1, \ldots, t+W. \]

Then corresponding \( P_t, \bar{P}_t \) satisfies:
\[ \delta_\infty(\bar{P}_t, P_t) \leq \gamma^{t+W+1-i} \log \left( \frac{\lambda_{\text{max}}(P_{\text{max}})}{\lambda_{\text{min}}(Q_{\text{min}})} \right), \quad t \leq i \leq t+W, \]
where \( \gamma = \frac{\lambda_{\text{max}}(A^T P_{\text{max}} A)}{\lambda_{\text{min}}(Q_{\text{min}})} \). Furthermore, by applying Lemma 7 in Appendix F, we have:
\[ \| \bar{P}_t - P_t \| \leq \gamma^{t+W+1-i} \frac{\lambda_{\text{max}}(P_{\text{max}})^2}{\lambda_{\text{min}}(Q_{\text{min}})}, \quad t \leq i \leq t+W. \]

**Proof.** Recall that
\[ P_t = F_{Q_t, R_t}(P_{t+1}), \quad \bar{P}_t = F_{Q_t, R_t}(\bar{P}_{t+1}), \quad t \leq i \leq t+W. \]
Define \( \alpha := \max \{ \lambda_{\text{max}}(A^T P_{\text{max}} A), \lambda_{\text{max}}(A^T \bar{P}_{\text{max}} A) \} \), Lemma D.2 in [36] implies:
\[ \delta_\infty(\bar{P}_t, P_t) \leq \frac{\alpha}{\lambda_{\text{min}}(Q_{\text{min}})} \delta_\infty(\bar{P}_{t+1}, P_{t+1}) \leq \frac{\lambda_{\text{max}}(A^T P_{\text{max}} A)}{\lambda_{\text{min}}(Q_{\text{min}})} \delta_\infty(\bar{P}_{t+1}, P_{t+1}) = \gamma \delta_\infty(\bar{P}_{t+1}, P_{t+1}), \quad t \leq i \leq t+W. \]
Applying this inequality recursively, we have:
\[ \delta_\infty(\bar{P}_t, P_t) \leq \gamma \delta_\infty(\bar{P}_{t+1}, P_{t+1}) \leq \cdots \leq \gamma^{t+W+1-i} \delta_\infty(\bar{P}_{t+W+1}, P_{t+W+1}) \leq \gamma^{t+W+1-i} \log \left( \frac{\lambda_{\text{max}}(P_{\text{max}})}{\lambda_{\text{min}}(Q_{\text{min}})} \right). \]

Then by applying Lemma 7 in Appendix F, we have:
\[ \| \bar{P}_t - P_t \| \leq \gamma^{t+W+1-i} \frac{\lambda_{\text{max}}(P_{\text{max}})^2}{\lambda_{\text{min}}(Q_{\text{min}})}, \quad t \leq i \leq t+W, \]
which completes the proof. 

**Proof.** (Lemma 3)
Lemma 3 is simply a corollary of Proposition 12. 

E. Exponential Stability

In this section we will look into the exponential stability of both finite time LQR (Proposition 2) and MPC algorithm (Proposition 3).

**Proof.** (Proposition 2)
Let \( V_t^{\text{LQR}}(x) := x^T P_t x \) be the value function for standard finite time horizon LQR problem. For arbitrary \( t, i, x \), let
\[ x_t := x, \quad x_{t+i} := (A - B_a K_{t+j}) \cdots (A - B_a K_{t+2})(A - B_a K_{t+1})x. \]
According to Bellman optimality equation, we have that
\[ V_{t+1}^{\text{LQR}}(x_{t+1}) = V_t^{\text{LQR}}(x_t) - x_t^T Q_t x_t - x_t^T K_t R_t K_t^T x_t \leq V_t^{\text{LQR}}(x_t) - x_t^T Q_t x_t = x_t^T P_t x_t - x_t^T \bar{Q}_t \bar{P}_t x_t \leq \left( 1 - \frac{\lambda_{\text{min}}(Q_{\text{min}})}{\lambda_{\text{max}}(P_{\text{max}})} \right) x_t^T P_t x_t \leq \left( 1 - \frac{\lambda_{\text{min}}(Q_{\text{min}})}{\lambda_{\text{max}}(P_{\text{max}})} \right) V_t^{\text{LQR}}(x_t). \]
Similarly,
\[
V_{t+2}^{LQR}(x_{t+2}) \leq \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right) V_{t+1}^{LQR}(x_{t+1}) \leq \cdots
\]
\[
\Rightarrow V_{t+i}^{LQR}(x_{t+i}) \leq \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^i V_t^{LQR}(x_t)
\]
Since we have:
\[
x_{t+i}^T P_{\min} x_{t+i} \leq V_{t+i}^{LQR}(x_{t+i})
\]
\[
\leq \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^i V_t^{LQR}(x_t)
\]
\[
\leq \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^i x_t^T P_{\max} x_t
\]
\[
\Rightarrow \|x_{t+i}\| \leq \lambda_{\max}(P_{\max}) \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^i \|x_t\|
\]
where \(x_t = x\), \(x_{t+i} = (A - B_u K_{t+i}) \cdots (A - B_u K_{t+2})(A - B_u K_{t+1})\). The inequality holds for arbitrary \(x\), thus we have:
\[
\| (A - B_u K_{t+i}) \cdots (A - B_u K_{t+2})(A - B_u K_{t+1}) \|
\]
\[
\leq \sqrt{\frac{\lambda_{\max}(P_{\max})}{\lambda_{\min}(P_{\min})} \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^i}
\]
which proves the proposition.

The proof for exponential stability of MPC algorithm is similar to the proof provided above.

**Proof.** (Proposition 3)

First we define a time-varying Lyapunov function:
\[
V_t^{\mathrm{MPC}}(x) := x^T P_{t|t} x
\]
\[
= \min_{s=t}^{t+W} \left( x_s^T Q_s x_s + u_s^T R_s u_s \right) + x_{t+W+1}^T P_{\max} x_{t+W+1}
\]
\[s.t. \quad x_{t+1} = Ax_t + B_u u_s
\]
\[x_t = x
\]
Applying Proposition 11 we can show that:
\[
Q_{\min} \leq P_{t|t} \leq P_{\max}
\]
For arbitrary \(t, j, x\), let:
\[
x_{t+j} := (A - B_u R_{t+j}) \cdots (A - B_u R_{t+2})(A - B_u R_{t+1}) x
\]
\[
\Rightarrow V_t^{\mathrm{MPC}}(x_{t+j}) \leq \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^j V_t^{\mathrm{MPC}}(x_t)
\]
\[
\Rightarrow \|x_{t+j}\|^2 \leq \lambda_{\max}(P_{\max}) \left(1 - \frac{\lambda_{\min}(Q_{\min})}{\lambda_{\max}(P_{\max})}\right)^j \|x_t\|^2
\]
which completes the proof.

**F. Properties of Invariant Metric**

**Lemma 6.** Suppose \(A, B\) are positive definite, and
\[
L \preceq A, B \preceq U,
\]
where \(L, U\) are both positive definite matrices. Then:
\[
\delta_{\infty}(A, B) \leq \frac{\lambda_{\max}(U)}{\lambda_{\min}(L)}
\]

**Proof.**
\[
\delta_{\infty}(A, B) = \|\log(A^{-1/2} B A^{-1/2})\| = \max \left\{ \log(\lambda_{\max} \left(A^{-1/2} B A^{-1/2}\right)), \log(\lambda_{\min} \left(A^{-1/2} B A^{-1/2}\right)) \right\}
\]
Since
\[
A^{-1/2} B A^{-1/2} \preceq A^{-1/2} \left(\frac{\lambda_{\max}(U)}{\lambda_{\min}(L)} \right) A^{-1/2} \leq \lambda_{\max}(U) A^{-1} \leq \lambda_{\max}(U) L^{-1} \leq \lambda_{\max}(U) I
\]
\[
A^{-1/2} B A^{-1/2} \succeq A^{-1/2} \left(\frac{\lambda_{\min}(L)}{\lambda_{\max}(U)}\right) A^{-1/2} = \lambda_{\min}(L) A^{-1} \leq \lambda_{\min}(L) U^{-1} \leq \lambda_{\min}(L) I
\]
Let

Proof.

It is easy to verify that

which completes the proof.

Lemma 7. Suppose \( A, B \) are positive definite, and

\[ A \preceq U, \quad \delta_{\infty}(A, B) \leq c, \]

where \( U \) is a positive definite matrix. Then,

\[ \| A - B \| \leq \lambda_{\max}(U) e^{e^{-1}} - \frac{1}{c} \delta_{\infty}(A, B). \]

Proof.

\[ \| A - B \| \leq \| A^{1/2}(I - A^{-1/2}BA^{-1/2})A^{1/2} \| \]

\[ \leq \| A^{1/2} \| \| I - A^{-1/2}BA^{-1/2} \| \]

\[ \leq \lambda_{\max}(U) \| I - A^{-1/2}BA^{-1/2} \| \]

Since

\[ e^{e^{-1}}(A, B)I - I \leq \lambda_{\min}(A^{-1/2}BA^{-1/2})I - I \leq A^{-1/2}BA^{-1/2} - I \]

\[ \leq \lambda_{\max}(A^{-1/2}BA^{-1/2})I - I \leq e^{\delta_{\infty}(A, B)}I - I \]

Thus

\[ \| I - A^{-1/2}BA^{-1/2} \| \leq \max \left\{ e^{\delta_{\infty}(A, B)} - 1, 1 - e^{-\delta_{\infty}(A, B)} \right\} \]

\[ \leq \max \left\{ e^{\delta_{\infty}(A, B)} - 1, \delta_{\infty}(A, B) \right\}. \]

It is easy to verify that \( e^x - 1 \leq e^{e^{-1}}x \) for \( 0 \leq x \leq c \). Thus, we have

\[ e^{\delta_{\infty}(A, B)} - 1 \leq \frac{e^{e^{-1}}}{c} \delta_{\infty}(A, B), \]

which completes the proof.

G. Others

Lemma 8. \( y_1, \ldots, y_n \in \mathbb{R} \), and

\[ a_t = a_t^1 y_1 + \cdots + a_t^n y_n, \quad t = 1, 2, \ldots, T, \]

where \( a_t^i \geq 0, \quad i = 1, 2, \ldots, n \). Then:

\[ \sum_{t=1}^{T} a_t^2 \leq \sum_{i} \left( \left( \sum_{t=1}^{T} a_t^i \right) \left( \sum_{j=1}^{n} a_j^i \right)^{y_i} \right) \]

\[ \leq \max_{i} \left( \sum_{t=1}^{T} a_t^i \right) \left( \sum_{j=1}^{n} a_j^i \right)^{y_i^2} \]

Proof. Define:

\[ a_t := [a_t^1, \ldots, a_t^n]^\top, \quad y := [y_1, \ldots, y_n]^\top, \]

then

\[ \sum_{t=1}^{T} a_t^2 = y^\top \left( \sum_{t=1}^{T} a_t a_t^\top \right) y. \]

Let

\[ A := \sum_{t=1}^{T} a_t a_t^\top. \]