2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

Competitive Control with Delayed Imperfect Information

Chenkai Yu[1], Guanya Shi[2], Soon-Jo Chung[2], Yisong Yue[2], and Adam Wierman[2]

***Abstract*— This paper studies the impact of imperfect information in online control with adversarial disturbances. In particular, we consider both *delayed* state feedback and *inexact* predictions of future disturbances. We introduce a greedy, myopic policy that yields a constant competitive ratio against the offline optimal policy. We also analyze the fundamental limits of online control with limited information by showing that our competitive ratio bounds for the greedy, myopic policy in the adversarial setting match (up to lower-order terms) lower bounds in the stochastic setting.**

I. INTRODUCTION

The design and analysis of controllers with imperfect, delayed information is a long-standing challenge for the fields of online learning and robust control. This paper provides a finite-time analysis of the impact of imperfect information in the context of a disturbed linear system with feedback delay. Specifically, we consider a disturbed online Linear Quadratic Regulator (LQR) optimal control problem with state feedback delay and inexact predictions of future disturbances, governed by $x_{t+1} = Ax_t + Bu_t + w_t$, where x_t, u_t, and w_t are the state, control, and disturbance respectively.

A growing literature at the interface of learning and control has emerged in recent years with the goal of designing controllers under various non-asymptotic learning-theoretic criteria, such as regret [1], [2], [3], [4], [5], dynamic regret [6], [7], [8], [9], and competitive ratio [10], [11], [9]. However, this line of work has made little progress when it comes to using imperfect prediction information, and has not approached the challenge of delayed feedback at all.

The task of designing controllers given imperfect predictions has a long history in the control literature, as well as in real-world applications [12], [13]. The basic idea is that, at state x_t, one has access to imperfect estimates $\hat{w}_{t+i}$ of the true future disturbances w_{t+i} ($i \geq 0$). Of particular relevance is robust Model Predictive Control (MPC) or MPC with uncertainty, which focuses on designing MPC policies given inexact future dynamics predictions [14], [15]. However, prior work focuses on stability and asymptotic performance, and there are no finite-time performance guarantees known.

The topic of control with delayed feedback is less studied, with some results in the robust control theory literature [16], [17], where the focus is on stability and motivation is taken from applications to real-world dynamical systems, e.g., [18], [19]. Here, the basic idea is that, at time t, one has only observed states up to x_{t-d} ($d \geq 0$). As in the case of imperfect predictions, there are no finite-time performance guarantees known to the best of our knowledge.

[1]Columbia University, USA. cyu26@gsb.columbia.edu.
[2]Caltech, USA. {gshi,sjchung,yyue,adamw}@caltech.edu.

Thus, while imperfect predictions and delayed feedback are of long-standing importance in both theory and practice, little progress has been made toward obtaining finite-time performance bounds for metrics like regret and competitive ratio under delayed imperfect information due to the technical challenges associated with proving such bounds.

A. Contributions

We show that a simple, myopic, predictive control has a constant competitive ratio bound in the case of delayed imperfect information (Theorem 4). This bound exponentially increases (decreases) in the number of delays (predictions), which highlights the cost associated with delay and the power of predictions, even when they are inexact. To the best of our knowledge, this result represents the first competitive ratio bounds for either the setting of inexact predictions or delayed feedback. We also prove that this bound is tight by showing that in some systems, the competitive ratio of the optimal online policy can be computed and matches our upper bound.

We would like to emphasize the generality of our result. The model we consider is the general online LQR setting with bounded adversarial disturbance in the dynamics, where only stabilizability is assumed. Further, the prediction errors are assumed to be adversarial. Additionally, our results compare to the globally optimal policies without any constraints (i.e., using the metric competitive ratio), rather than the optimal linear, static policy (i.e., using static regret).

Our result adds further evidence that the structure of LQR allows simple algorithmic ideas to be effective: [3] recently proved that the naive exploration is optimal in online LQR adaptive control problem with unknown $\{A, B\}$, and [7] proved the classic MPC is near-optimal in online LQR control with exact future predictions. Combined with the current paper, there is growing evidence that simple, myopic policies that build on MPC are constant-competitive and near-optimal, even in adversarial settings with delayed imperfect information, which sheds light on key algorithmic principles and fundamental limits in continuous control.

Due to lack of space, most of the proofs are omitted and can be found in an extended version of this paper [20].

B. Related work

There is a growing literature of papers that approach the control of linear dynamical systems with tools and concepts from online learning, with a focus on finite-time non-asymptotic guarantees. This non-asymptotic perspective is important because (1) it can easily integrate with learning theory and (2) finite-time optimality analysis is crucial for many real-world systems with finite horizons. Within this

978-1-6654-5196-3/$31.00 ©2022 AACC

literature, most work focuses on the design of controllers with low regret [1], [2], [3], [7], [8], [21]. The study of competitive ratio has received increased attention recently [11], [10], [22], [9]. Again, note that competitive ratio compares with the offline optimal policy without any constraints while regret compares with the optimal static policy in a specific policy class. The most general such result to this point is [9] which provides competitive bounds of a predictive controller in a time-varying setting. However, the bounds in [9] require a sufficiently large amount of *exact* predictions.

In these lines of work, very few papers focus on settings where the controller has access to predictions of future disturbances. Among these papers, [7], [9] focus on settings where predictions are exact. To the best of our knowledge, the only result for inexact predictions is studied in [8], however [8] only gives dynamic regret results which are data-dependent. In this paper we provide stronger, data-independent results for competitive ratio. Inexact dynamic predictions have received attention in the robust MPC community (e.g., [14], [15]), focusing on stability and asymptotic performance analysis. In contrast, this paper focuses on non-asymptotic analysis from a learning theory perspective. Even outside of control in the related area of online optimization, when predictions are considered, they are typically assumed to be exact [23]. One exception is [24], which uses a less general model of prediction error than the current paper, and the connection to control is unclear.

In contrast to the literature on predictions, there is no work providing non-asymptotic guarantees (either regret or competitive ratio) of policies subject to delayed feedback. The issue has received considerable attention in the robust control community [17], [18], but the focus is typically on closed-loop stability and no finite-time performance bounds exist to the best of our knowledge.

II. MODEL

We consider an online *Linear Quadratic Regulator (LQR)* optimal control problem with adversarial disturbances in the dynamics. In particular, we consider a linear system initialized with $x_0 \in \mathbb{R}^n$ and controlled by $u_t \in \mathbb{R}^m$ at each step $t \in \{0, 1, \ldots, T-1\}$ where T is the total length of the problem. The system dynamics is governed by:

$$x_{t+1} = Ax_t + Bu_t + w_t,$$

where w_t is the disturbance. We assume that w_t is bounded, i.e., $\|w_t\| \leq r$. The goal is to minimize the following cost:

$$J = \sum_{t=0}^{T-1} (x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t) + x_T^\mathsf{T} Q_f x_T, \tag{1}$$

given matrices A, B, Q, R, Q_f. We consider an *online* setting where an adversary *adaptively* selects $\{w_t\}_{t=0}^{T-1}$, and the controller (also adaptively) makes the decision u_t at every time step t, potentially based on *delayed imperfect information* (discussed later in this section).

We make the standard assumptions that $Q, Q_f \succeq 0$, $R \succ 0$, and (A, B) is stabilizable [25], [7], i.e., $\exists K_0 \in \mathbb{R}^{m \times n}$ such that $\rho(A - BK_0) < 1$. We further assume (A^T, Q) is stabilizable to guarantee the stability of the closed-loop [26]. This assumption is more general than the standard assumption that $Q \succ 0$ because $Q \succ 0$ implies the stabilizability of (A^T, Q).

Throughout this paper, we use $\rho(\cdot)$ to denote the spectral radius of a matrix and $\|\cdot\|$ to denote the 2-norm of a vector or the spectral norm of a matrix.

Note that many important problems can be considered as special cases of the above model. One motivating example is the Linear Quadratic (LQ) tracking problem [26].

Example 1 (Linear quadratic tracking): The LQ tracking problem is defined via dynamics $x_{t+1} = Ax_t + Bu_t + \tilde{w}_t$, *and cost function* $J = \sum_{t=0}^{T-1} (x_{t+1} - y_{t+1})^\mathsf{T} Q (x_{t+1} - y_{t+1}) + u_t^\mathsf{T} R u_t$, *where* $\{y_t\}_{t=1}^{T}$ *is the desired trajectory to track.*

To fit LQ tracking into our model, let $\tilde{x}_t = x_t - y_t$. Then, we get $J = \sum_{t=0}^{T-1} \tilde{x}_{t+1}^\mathsf{T} Q \tilde{x}_{t+1} + u_t^\mathsf{T} R u_t$ and $\tilde{x}_{t+1} = A\tilde{x}_t + Bu_t + w_t$, which is a LQR control problem with disturbance $w_t = \tilde{w}_t + Ay_t - y_{t+1}$ in the dynamics. Note that in many LQ tracking problems, delayed observations and imperfect predictions are fundamental challenges [26].

A. Delayed Imperfect Information

The LQR optimal control problem introduced above is typically studied without *predictions* or *delays*. In the classic setting, at each time $t = 0, 1, 2, \ldots$, the controller observes x_t and then decides u_t without knowing w_t. Thus, u_t is a function of all previous information: $u_t = \pi_t(x_0,\ x_1,\ \ldots,\ x_t,\ u_0,\ u_1,\ \ldots,\ u_{t-1})$. In an equivalent formulation, the controller is given x_0 to start and, at each time t, obtains the previous disturbance w_{t-1} (if $t \geq 1$) and then decides u_t. Thus, $u_t = \pi_t(x_0, w_0, w_1, \ldots, w_{t-1}, u_0, u_1, \ldots, u_{t-1})$.

The motivation for this paper is that in many real-world problems (e.g., [12], [13]), predictions of future information are available and using them is crucially important, even though they are typically noisy and delayed. For example, data-driven model-based control is a prominent and successful approach where it is crucial to consider model mismatch due to statistical learning error. Moreover, in many situations, there is state feedback delay in the system, so the controller has to make the decision u_t before the current state x_t is observed. The existence of both imperfect prediction and feedback delay leads to considerable difficulty.

Formally, we model the delayed inexact predictions as follows. At each step t, the revealed information is:

$$x_0,\ u_0, \ldots, u_{t-1},\ w_0, \ldots, w_{t-d-1},\ \hat{w}_{t-d|t}, \ldots, \hat{w}_{T-1|t},$$

or equivalently,

$$x_0,\ u_0, \ldots, u_{t-1},\ x_1, \ldots, x_{t-d},\ \hat{w}_{t-d|t}, \ldots, \hat{w}_{T-1|t},$$

where $d \geq 0$ is the length of feedback delay, and $\hat{w}_{s|t}$ is the prediction of w_s at time t.

We define $e_{s|t} = w_s - \hat{w}_{s|t}$ as the estimation error, and we assume that the predictor satisfies $\left\|e_{t-d+i|t}\right\| \leq \epsilon_i \|w_{t-d+i}\|$ for all $i \geq 0$ and $t \geq 0$, where ϵ_i is a given parameter that measures the prediction quality at i steps into the unknown. Although predictions are available for every time step, those for far future may have bad quality, i.e., ϵ_i is typically large

for large i. Therefore, a good control policy may not use all predictions in the same way: using only the predictions with smaller estimation error may yield better performance. For example, if $\epsilon_i > 1$, we can simply let $\hat{w}_{t-d+i|t} = 0$, which is a better prediction with $\epsilon_i = 1$. In general, the error bounds have to be multiplicative rather than additive, if one pursues a competitive ratio guarantee. Consider the case where all disturbances w_t are zero, but there are nonzero prediction errors. The optimal offline policy incurs zero cost, while any online policy that uses the (wrong) predictions incurs nonzero cost, hence leading to an infinite ratio.

Our setting of *delayed imperfect information* generalizes many existing settings in the study of LQR control. The classic setting is the special case where $d = 0$ and $\hat{w}_{t+i|t} = 0$ for all i and t. The offline optimal setting considered by [25], [27] is the special case where $d = 0$ and $\epsilon_i = 0$ for all i. The setting considered by [7], where k *exact* predictions without delay are available, corresponds to $d = 0$, $\epsilon_i = 0$ for $0 \le i \le k-1$ and $\hat{w}_{t+i|t} = 0$ for all t and $i \ge k$.

B. Competitive Ratio

We measure the performance using the *competitive ratio*, which bounds the worst-case ratio of the cost of an online policy (Alg) to the cost of the optimal offline policy (Opt) with perfect knowledge of $\{w_t\}_{t=0}^{T-1}$. Formally, we study the so-called *weak* competitive ratio, which allows for an additive horizon-independent constant factor. We say that a policy is (weakly) c-competitive if, given A, B, Q, R, Q_f and r, for any adversarially and adaptively chosen disturbances $\{w_t\}_{t=0}^{T-1}$, we have $\mathsf{Alg} \le c\,\mathsf{Opt} + \kappa$, where κ is a constant, independent of T. We say that the algorithm is *constant competitive* when c is a constant independent of T.

While there has been considerable success in recent years designing control policies that are no-regret, e.g., [1], [2], [7], there have been very few examples of constant competitive controllers. The few results that exist, e.g., [10], [11], tend to have more restrictive assumptions on the dynamics and/or disturbances. Existing results on regret typically focus on *static* regret, which compares a policy to the optimal offline *static* policy in a specific policy class (e.g., linear policies [2]), while the *dynamic* regret and the competitive ratio compare a policy to the general optimal offline policy that may potentially be non-linear and non-static. Although it is possible for an online policy to have sublinear *static* regret, in general, the optimal static linear policy may have cost that is an order-of-magnitude larger than the optimal offline cost [10]. On the other hand, it has been shown [7] that the minimum *dynamic* regret of an online policy can be linear in T even with random disturbances. Thus, it is natural to consider the competitive ratio and pin down the multiplicative constant incurred from not knowing the future.

III. A MYOPIC PREDICTIVE POLICY

In this paper, we study a myopic predictive control policy that is a natural and myopic variant of model predictive control (MPC). When there are predictions, but no delays, MPC is a popular and successful approach [13], [28]. In fact, [7] recently showed that MPC has a near-optimal dynamic regret in the case of exact predictions and no delay. Note that this paper aims to show simple/standard algorithms (either standard MPC or its natural extension) could be competitive and the competitive ratio bounds could be tight, rather than propose a new algorithm.

Suppose the controller uses k predictions. At each time t, the controller optimizes based on $x_t, \hat{w}_{t|t}, \dots, \hat{w}_{t+k-1|t}$:

$$\begin{aligned}&(u_t, \dots, u_{t+k-1}) = \\ &\quad \arg\min_u \left(\sum_{i=t}^{t+k-1} (x_i^\mathsf{T} Q x_i + u_i^\mathsf{T} R u_i) + x_{t+k}^\mathsf{T} \tilde{Q}_f x_{t+k} \right), \\ &\text{s.t. } x_{i+1} = Ax_i + Bu_i + \hat{w}_{i|t}, \ \forall i = t, \dots, t+k-1. \quad (2)\end{aligned}$$

This optimization is myopic in the sense that it assumes that the length of the problem is k instead of T and only uses predicted future disturbances within those k steps. The terminal cost matrix $\tilde{Q}_f$ in (2) may or may not be the same as the terminal cost matrix Q_f of the original problem (1), and can be viewed as a hyperparameter. Similarly, k is also a hyperparameter. Larger k is not necessarily better because the predictions in the far future may have very large errors. In this paper, we let $\tilde{Q}_f = P$, where P is the solution of the discrete algebraic Riccati equation (DARE):

$$P = Q + A^\mathsf{T} P A - A^\mathsf{T} P B (R + B^\mathsf{T} P B)^{-1} B^\mathsf{T} P A. \quad (3)$$

The output of (2) is k control actions corresponding to time $t, t+1, \dots, t+k-1$, respectively, but only the first (u_t) is applied to the system. The rest (i.e., $u_{t+1}, \dots, u_{t+k-1}$) are discarded. The explicit solution of (2) is given below [7].

Proposition 2: The policy defined by (2) at time t is:

$$u_t = -(R + B^\mathsf{T} P B)^{-1} B^\mathsf{T} \left(P A x_t + \sum_{i=0}^{k-1} {F^\mathsf{T}}^i P \hat{w}_{t+i|t} \right),$$

where $F = A - B(R + B^\mathsf{T} P B)^{-1} B^\mathsf{T} P A =: A - BK$.

The closed loop is stable since $\rho(F) < 1$ [7]. As stated above, the policy does not directly apply to the case of delayed imperfect information. To adapt it, we consider two cases: (i) when the number of predictions available is longer than the feedback delay, i.e., $k \ge d$, and (ii) when the delay is longer than the number of predictions available, i.e., $k < d$.

When $k \ge d$, the extension is perhaps straightforward. Here, although the controller does not know the current state x_t, it knows x_{t-d} and $\hat{w}_{t-d|t}, \dots, \hat{w}_{t-1|t}$. Thus, it can estimate the current state. This means that it is possible to simply use this estimation, $\hat{x}_{t|t}$, as a replacement for x_t in the algorithm, which yields the following:

$$u_t = -(R + B^\mathsf{T} P B)^{-1} B^\mathsf{T} \left(P A \hat{x}_{t|t} + \sum_{i=0}^{k-d-1} {F^\mathsf{T}}^i P \hat{w}_{t+i|t} \right). \quad (4)$$

When $k < d$, the extension is not as obvious. In this setting, the quality of the predictions is poor enough that it is better not to use the predictions to estimate the current state. Thus, one cannot simply estimate the current state and run classic MPC. In this case, the key is to view (classic) MPC

from a different perspective: MPC locally solves an optimal control problem by treating known disturbances (predictions) as exact, and treating unknown disturbances as zero [7], [28], [8]. Following this philosophy, in the case when predictions are not enough to be used to estimate the current state, we can instead assume that unknown disturbances are exactly zero. The following theorem derives the optimal policy under this "optimistic" assumption.

Theorem 3: Suppose there are d delays and k exact predictions with $k < d$. Assume all used predictions are exact and other disturbances (with unused predictions) are zero. The optimal policy at time t is:

$$u_t = -(R+B^\mathsf{T}PB)^{-1}B^\mathsf{T}PA\Bigg(A^{d-k}\hat{x}_{t-d+k|t} + \sum_{i=0}^{d-k-1} A^i B u_{t-1-i}\Bigg). \quad (5)$$

In other words, the policy in (5) first obtains the greedy estimation $\hat{x}_{t-d+k|t}$ using predictions $\hat{w}_{t-d|t}, \dots, \hat{w}_{t-d+k-1|t}$, and then estimates the current state by treating $w_{t-d+k} = \cdots = w_{t-1} = 0$. In fact, instead of treating them as zero, we can impose other values or distributions on those disturbances. This would generalize Theorem 3 to a broader class of policies.

To summarize the two cases above, the myopic generalization of MPC we study in this paper is described as follows. Suppose we want to use k predictions. If $k \geq d$, then we estimate the current state x_t and apply (4). If $k < d$, then we estimate the state at time $t-d+k$ and apply (5). If fact, the two cases coincide when $k = d$.

IV. PERFORMANCE BOUNDS

Our main result provides bounds on the competitive ratio for the policy defined in Section III in the case of inexact delayed predictions. We present our general result below and then discuss the special cases of (i) exact predictions and no delay, (ii) inexact predictions and no delay, and (iii) delay but no access to predictions. The special cases illustrate the contrast between inexact and exact predictions as well as the impact of delay.

Theorem 4 (Main result): Let $c = \|P\|\|P^{-1}\|(1+\|F\|)$ and $H = B(R+B^\mathsf{T}PB)^{-1}B^\mathsf{T}$. Suppose there are d steps of delays and the controller uses k predictions. When $k \geq d$,

$$\mathsf{Alg} \leq \left[\frac{\left(c\sum_{i=0}^{d-1}\epsilon_i\|A^{d-i}\| + c\sum_{i=d}^{k-1}\epsilon_i\|F^{i-d}\| + \|F^{k-d}\|\right)^2}{\|H\|^{-1}\lambda_{\min}(P^{-1}-FP^{-1}F^\mathsf{T}-H)} + 1\right]\mathsf{Opt} + O(1).$$

When $k \leq d$,

$$\mathsf{Alg} \leq \left[\frac{\left(c\sum_{i=0}^{k-1}\epsilon_i\|A^{d-i}\| + c\sum_{i=k}^{d-1}\|A^{d-i}\| + 1\right)^2}{\|H\|^{-1}\lambda_{\min}(P^{-1}-FP^{-1}F^\mathsf{T}-H)} + 1\right]\mathsf{Opt} + O(1).$$

The $O(1)$ is with respect to T. It may depend on the system parameters A, B, Q, R, Q_f and the range of disturbances r, but not on T. When $Q_f = P$ the $O(1)$ is zero.

The two cases in the theorem correspond to the two cases in the algorithm: when predictions are of high enough quality to allow estimation of the current state and when they are not. Note that the closed-loop dynamics is stable, i.e., $\rho(F) = \rho(A-BK) < 1$. Therefore, there exists a constant γ such that $\|F^i\| \leq \gamma(\frac{\rho(F)+1}{2})^i$ for all $i \geq 1$ from Gelfand's formula.

In the first case, we see that the quality of predictions in the near future has more impact, especially when $\rho(A) > 1$. In the second case, we see that the amount of delay d exponentially increases the bound if $\epsilon_i > 0$ and $\rho(A) > 1$. We explore further insights by looking at special cases in the following subsections. Before moving on, we provide an overview of the proof of Theorem 4.

A. Proof Sketch for Theorem 4

We first prove Theorem 4 in the case $Q_f = P$. In this case, the $O(1)$ is not needed. Then, we analyze the impact of the terminal cost Q_f and show that it introduces at most an $O(1)$ additional cost. The full proof can be found in the appendix of an extended version of this paper [20].

Lemma 5: The conclusion of Theorem 4 holds if $Q_f = P$.

The proof of the result in the case of $Q_f = P$ follows from a novel difference analysis of the quadratic cost-to-go functions. Here, we focus on the second part of the proof, i.e., reducing the case of $Q_f \neq P$ to the case of $Q_f = P$. To that end, let $\mathsf{Alg}(X)$ be the cost of our algorithm when the terminal cost is X and, similarly, let $\mathsf{Opt}^Y(X)$ be the cost of the policy that is optimal for terminal cost Y when the terminal cost is actually X.

Our analysis proceeds by first bounding the impact of the terminal cost on the gap between the algorithm and the optimal cost via the following lemma.

Lemma 6: For any algorithm,

$$\mathsf{Alg}(Q_f) - \mathsf{Opt}^P(Q_f) \leq \mathsf{Alg}(P) - \mathsf{Opt}^P(P) + O(1).$$

Then, we prove that the terminal cost only has an $O(1)$ impact on the optimal cost using the following lemma.

Lemma 7: The followings are equal up to $O(1)$ difference: $\mathsf{Opt}^P(P)$, $\mathsf{Opt}^P(Q_f)$, $\mathsf{Opt}^{Q_f}(Q_f)$.

Together, these two lemmas imply that

$$\begin{aligned}
&\mathsf{Alg}(Q_f) - \mathsf{Opt}^{Q_f}(Q_f)\\
&\leq \mathsf{Alg}(Q_f) - \mathsf{Opt}^P(Q_f) + O(1)\\
&\leq \mathsf{Alg}(P) - \mathsf{Opt}^P(P) + O(1)\\
&\leq \frac{\mathsf{Alg}(P) - \mathsf{Opt}^P(P)}{\mathsf{Opt}^P(P)}\mathsf{Opt}^P(P) + O(1)\\
&\leq \frac{\mathsf{Alg}(P) - \mathsf{Opt}^P(P)}{\mathsf{Opt}^P(P)}\mathsf{Opt}^{Q_f}(Q_f) + O(1).
\end{aligned}$$

Thus, we can complete the proof by concluding that

$$\mathsf{Alg}(Q_f) \leq \frac{\mathsf{Alg}(P)}{\mathsf{Opt}^P(P)}\mathsf{Opt}^{Q_f}(Q_f) + O(1).$$

B. Exact Predictions Without Delay

In the sections that follow, we explore special cases of Theorem 4 in order to highlight the impact of inexact predictions and delay. First, we present the special case of k accurate, exact predictions and no feedback delays. Formally, we have $d = 0$, $\epsilon_i = 0$ for $0 \le i \le k-1$ and $\hat{w}_{t+i|t} = 0$ for all t and $i \ge k$.

The main result for this setting is given below. It directly follows from the $k \ge d$ case of Theorem 4.

Theorem 8: Suppose there are k exact predictions and no feedback delay. Then:

$$\mathsf{Alg} \le \left[1 + \frac{\|F^k\|^2\|H\|}{\lambda_{\min}(P^{-1} - FP^{-1}F^\mathsf{T} - H)}\right]\mathsf{Opt} + O(1).$$

Thus, the competitive ratio exponentially decreases as k goes up. To illustrate how the primary parameters A, B, Q and R affect the competitive ratio in this case, it is useful to consider the case when $n = m = 1$, as shown in both Corollary 8.1 and Figure 1.

Corollary 8.1: Assume there are k exact predictions and no feedback delay, and let $n = m = 1$ and $Q_f = P$. Then,

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{2A^{4-2k}}{B^2Q/R} \qquad \text{if } A^2 \gg B^2Q/R + 1,$$

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{A^{2k}}{(B^2Q/R)^{2k-1}} \qquad \text{if } B^2Q/R \gg A^2 + 1.$$

Interestingly, in this case, the competitive bound only depends on A^2 and B^2Q/R. It does not depend on the sign of A, nor on B, Q or R as long as B^2Q/R is fixed. Further, when $k \ge 3$, we see that the competitive ratio is small if B, Q are small, R is large, or A is either very large or very small. However, when $k = 0$ or 1, a large A can result in a large competitive ratio. When $k = 0$, a large value of B^2Q/R also results in a large competitive ratio. We see below that this is similar to the case of delay (see Section IV-D).

Proposition 9: Theorem 8 is tight in the sense that there exist systems where the competitive ratio of the optimal online algorithm is $1 + \Theta(\|F^k\|^2)$.

C. Inexact Predictions Without Delay

We next consider the case where predictions are inexact, but there is no feedback delay. The contrast with the previous section highlights the impact of prediction error.

As discussed in Section II-A, the controller should optimize k to utilize predictions with smaller estimation errors while avoiding the use of those with larger errors. The following directly follows from the $d = 0$ case of Theorem 4 and reduces to the exact case when $\epsilon_i = 0$.

Theorem 10: Suppose there are k inexact predictions and no feedback delay. Then,

$$\mathsf{Alg} \le \left[\frac{\|H\|\left(c\sum_{i=0}^{k-1}\epsilon_i\|F^i\| + \|F^k\|\right)^2}{\lambda_{\min}(P^{-1} - FP^{-1}F^\mathsf{T} - H)} + 1\right]\mathsf{Opt} + O(1).$$

This subsection differs from the previous one in that the controller can minimize the bound in Theorem 10 with respect to k. We characterize this optimization in the following result in 1-d systems, and also provide simulation evidence in Section V (see Figure 3).

Corollary 10.1: Suppose there are k inexact predictions and no feedback delay. Assume $n = m = 1$. Given non-decreasing $\{\epsilon_i\}$, to minimize the competitive ratio bound in Theorem 10, the optimal number k of predictions to use is such that:

$$\epsilon_{k-1} < \frac{1 - |F|}{1 + |F|} < \epsilon_k.$$

The following 1-d setting highlights the dependence of the competitive ratio on the system parameters.

Corollary 10.2: Assume there are k inexact predictions and no feedback delay, and let $n = m = 1$ and $Q_f = P$.

If $A^2 \gg B^2Q/R + 1$,

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{2A^4}{B^2Q/R}\left(\sum_{i=0}^{k-1}\frac{\epsilon_i}{|A|^i} + \frac{1}{|A|^k}\right)^2.$$

If $B^2Q/R \gg A^2 + 1$,

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{B^2Q}{R}\left(\sum_{i=0}^{k-1}\frac{\epsilon_i|A|^i}{(B^2Q/R)^i} + \frac{|A|^k}{(B^2Q/R)^k}\right)^2.$$

The dependence of the competitive ratio on A, B, Q, R is similar to the case of exact predictions. In particular, we find that the prediction quality in the near future is (exponentially) more important than further in the future, which is consistent with the robust MPC literature [15].

In the exact prediction case we show that Theorem 8 is tight with respect to $\|F^k\|$. In contrast, in the inexact case the tightness of ϵ_i and $\|F^i\|$ in Theorem 10 remains as an open question.

D. Delay Without Predictions

The last special case we consider is the case with delays but no (usable) predictions. This case separates the impact of delay from that of predictions. Here, $\hat{w}_{t-d+i|t} = 0$ for all t and $i \ge 0$. When $k \le d$, via Theorem 4 we have that:

$$\mathsf{Alg} \le \left[\frac{\|H\|\left(c\sum_{i=1}^{d}\|A^i\| + 1\right)^2}{\lambda_{\min}(P^{-1} - FP^{-1}F^\mathsf{T} - H)} + 1\right]\mathsf{Opt} + O(1).$$

Depending on whether the spectral radius $\rho(A) < 1$, there are two simplifications one can make: (i) if $\rho(A) < 1$, then $\|A^i\| \le \kappa a^i$ for $a = (\rho(A) + 1)/2 < 1$ for a constant κ, and (ii) if $\rho(A) > 1$, $\|A^i\| \le \|A\|^i$.

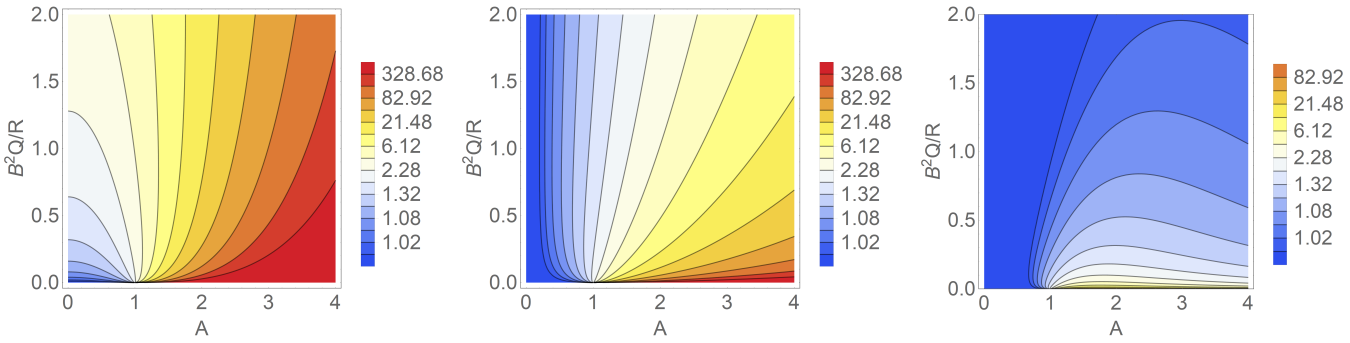
Theorem 11: Suppose there are d delays and no predictions are available. If $\rho(A) < 1$, then the competitive ratio is bounded by a constant irrelevant to the length of delay:

$$\mathsf{Alg} \le \left[\frac{\|H\|\left(c\kappa\frac{a}{1-a} + 1\right)^2}{\lambda_{\min}(P^{-1} - FP^{-1}F^\mathsf{T} - H)} + 1\right]\mathsf{Opt} + O(1).$$

If $\rho(A) > 1$, then the competitive ratio bound grows exponentially fast with the of number of delay steps:

$$\mathsf{Alg} \le \left[\frac{\|H\|\left(c\frac{\|A\|^{d+1} - \|A\|}{\|A\| - 1} + 1\right)^2}{\lambda_{\min}(P^{-1} - FP^{-1}F^\mathsf{T} - H)} + 1\right]\mathsf{Opt} + O(1).$$

As in the previous subsections, it is useful to consider the one-dimensional case to get insights about the impact of the system parameters.



(a) $k = 0$. (b) $k = 1$. (c) $k = 3$.

Fig. 1: Illustration of the competitive ratio bound in Corollary 8.1. The system is one-dimensional ($n = m = 1$) without delays ($d = 0$). With no predictions ($k = 0$), the bound is small only if both A and B^2Q/R are small. When $k = 1$, it is small if A is small or B^2Q/R is large. When $k = 3$, it is small if A is either small or large, or if B^2Q/R is large. The bound is tight in the sense that in some systems, it equals the relative cost of the optimal online policy.

Corollary 11.1: Assume there are d delays and no predictions. Let $n = m = 1$ and $Q_f = P$. Then,

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{2A^{4+2d}}{B^2Q/R} \qquad \text{if } A^2 \gg B^2Q/R + 1,$$

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{A^{2d+2}B^2Q/R}{(|A|-1)^2} \quad \text{if } B^2Q/R \gg A^2 + 1,\ |A| > 1,$$

$$\frac{\mathsf{Alg}}{\mathsf{Opt}} \le 1 + \frac{B^2Q/R}{(1-|A|)^2} \qquad \text{if } |A| < 1.$$

Contrary to the case with $k \ge 3$ inexact predictions, when there are delays, a large value of B^2Q/R or A does not lead to a small competitive ratio. Instead, in the case of feedback delay, it results in a large competitive ratio. This is consistent with results from robust control theory: the less stable the open loop is ($|A|$ is larger), the more impact delay has [16].

Proposition 12: Theorem 11 is tight in the sense that there exist systems such that the competitive ratio of the optimal online algorithm is at least $1 + \Theta(\|A^d\|^2)$.

V. NUMERICAL EXAMPLES

To illustrate our results, we end the paper with numerical examples that highlight the impact of delayed, inexact predictions. To that end, we consider a 2-d tracking problem with the following trajectory [6], illustrated in Figure 2:

$$y_t = \begin{bmatrix} 16\sin^3\left(\frac{t}{4}\right) \\ 13\cos\left(\frac{t}{4}\right) - 5\cos\left(\frac{2t}{4}\right) - 2\cos\left(\frac{3t}{4}\right) - \cos\left(\frac{4t}{4}\right) \end{bmatrix}.$$

We consider following double integrator dynamics:

$$p_{t+1} = p_t + 0.2v_t + h_t, \quad v_{t+1} = v_t + 0.2u_t + \eta_t,$$

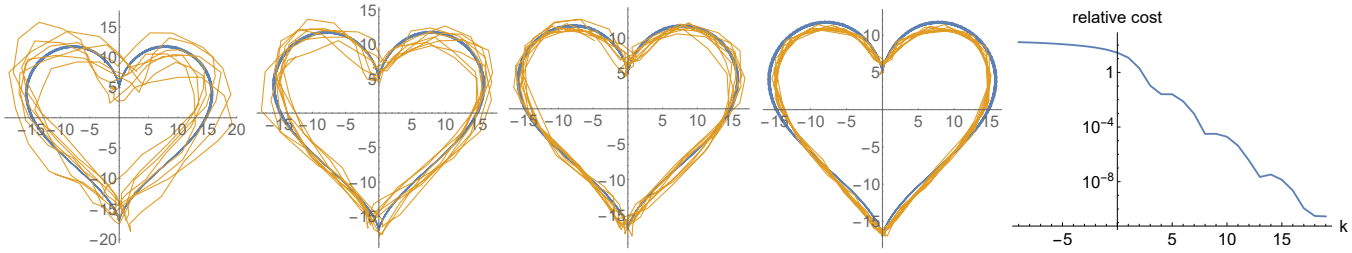
where $p_t \in \mathbb{R}^2$ is the position, v_t is the velocity, u_t is the control, and $h_t, \eta_t \sim \mathrm{U}[-1,1]^2$ are i.i.d. noises. The objective is to minimize $\sum_{t=0}^{T-1} \|p_t - y_t\|^2 + 0.0016\|u_t\|^2$, where we let $T = 200$. This problem can be converted to the standard LQR with disturbance w_t by letting $x_t = \left[\begin{smallmatrix} p_t \\ v_t \end{smallmatrix}\right]$ and $\tilde{w}_t = \left[\begin{smallmatrix} h_t \\ \eta_t \end{smallmatrix}\right]$ and then using the reduction in the LQ tracking example in Section II. Note that the disturbances are the combination of a deterministic trajectory and i.i.d. noise. In contrast, our theoretical results focus on more challenging adversarial disturbances. Nonetheless, the numerical results are consistent with our theorems.

In our first experiment, we study the effect of the number of delays or predictions. For simplicity, we exclude the effect of inexactness of the predictions — a prediction is either exact ($\epsilon_i = 0$) or uninformative ($\hat{w}_{t-d+i|t} = 0$). In this case, each exact prediction cancels a step of delay so, delays can be viewed as "negative" predictions. Figure 2 shows the performance of the proposed myopic policy in Section III with different numbers of predictions or delays. We see that the cost exponentially decreases (increases) as the number of predictions (delays) increases.

In the second experiment, we study the effect of inexact predictions and show that the controller needs to optimize how many predictions are used — it is better to use only a few predictions and ignore those that are too noisy. Specifically, we let $\epsilon_i = \frac{1}{5}i^2$, i.e., the noise level of predictions grows quadratically fast with the number of steps into the future. Each estimation error $e_{t+i|t} = w_{t+i} - \hat{w}_{t+i|t}$ is independently sampled from $\mathrm{U}[-\epsilon_i\|w_{t+i}\|, \epsilon_i\|w_{t+i}\|]^4$. This process is repeated 8 times, with each instance depicted by an orange line and their maximum represented by a red line. As shown in Figure 3, with exact predictions, the cost will decrease as the number of predictions increase (the blue line); while with inexact predictions, using fewer predictions may yield better performance.

VI. CONCLUDING REMARKS

Our results presens the first constant-competitive policy for general LQR control with adversarial disturbances and delayed imperfect information. We also show that in the case of exact predictions with no delay, or in the case of delay with no predictions, the competitive ratio bounds of the proposed myopic policy match the lower bound. However, in the inexact prediction case, the tightness of ϵ_i in our bounds remains as an open question. Other important extensions include nonlinear dynamics and time-variant linear systems, which can also lead to studying online learning of robust controllers under model mismatch.



(a) $d = 5,\ k = 0.$ (b) $d = 1,\ k = 0.$ (c) $d = 0,\ k = 1.$ (d) $d = 0,\ k = 8.$ (e) Relative cost: $\mathsf{Alg}/\mathsf{Opt} - 1.$

Fig. 2: Tracking results with k exact predictions and no delay, or with d steps of delay with no predictions. (a-d) show the desired trajectory (blue) and actual (orange) trajectories. (e) shows both delay and prediction on the x-axis, with the negative part corresponding to delay. The y-axis is in log-scale.

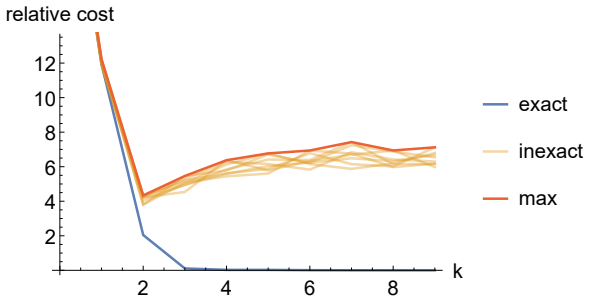


Fig. 3: The impact of inexact predictions. The relative cost ($\mathsf{Alg}/\mathsf{Opt} - 1$) of MPC using k exact (blue) or inexact (orange) predictions is shown.

REFERENCES

[1] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," in *Neural Information Processing Systems (NeurIPS)*, 2018.

[2] N. Agarwal, B. Bullins, E. Hazan, S. M. Kakade, and K. Singh, "Online control with adversarial disturbances," in *International Conference on Machine Learning (ICML)*, 2019.

[3] M. Simchowitz and D. J. Foster, "Naive exploration is optimal for online LQR," *arXiv preprint arXiv:2001.09576*, 2020.

[4] M. Simchowitz, K. Singh, and E. Hazan, "Improper learning for non-stochastic control," *arXiv preprint arXiv:2001.09254*, 2020.

[5] G. Shi, K. Azizzadenesheli, S.-J. Chung, and Y. Yue, "Meta-adaptive nonlinear control: Theory and algorithms," *arXiv preprint arXiv:2106.06098*, 2021.

[6] 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. 14 858–14 870.

[7] C. Yu, G. Shi, S.-J. Chung, Y. Yue, and A. Wierman, "The power of predictions in online control," *Neural Information Processing Systems (NeurIPS)*, 2020.

[8] R. Zhang, Y. Li, and N. Li, "On the regret analysis of online lqr control with predictions," *arXiv preprint arXiv:2102.01309*, 2021.

[9] Y. Lin, Y. Hu, H. Sun, G. Shi, G. Qu, and A. Wierman, "Perturbation-based regret analysis of predictive control in linear time varying systems," *arXiv preprint arXiv:2106.10497*, 2021.

[10] G. Shi, Y. Lin, S.-J. Chung, Y. Yue, and A. Wierman, "Beyond no-regret: Competitive control via online optimization with memory," *Neural Information Processing Systems (NeurIPS)*, 2020.

[11] 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.

[12] G. Shi, X. Shi, M. O'Connell, R. Yu, K. Azizzadenesheli, A. Anandkumar, Y. Yue, and S.-J. Chung, "Neural lander: Stable drone landing control using learned dynamics," in *International Conference on Robotics and Automation (ICRA)*, 2019.

[13] N. Lazic, C. Boutilier, T. Lu, E. Wong, B. Roy, M. Ryu, and G. Imwalle, "Data center cooling using model-predictive control," in *Advances in Neural Information Processing Systems*, 2018, pp. 3814–3823.

[14] A. Bemporad and M. Morari, "Robust model predictive control: A survey," in *Robustness in identification and control*. Springer, 1999, pp. 207–226.

[15] M. Cannon and B. Kouvaritakis, "Optimizing prediction dynamics for robust mpc," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1892–1897, 2005.

[16] K. Zhou and J. C. Doyle, *Essentials of robust control*. Prentice hall Upper Saddle River, NJ, 1998, vol. 104.

[17] J. H. Kim and H. B. Park, "H∞ state feedback control for generalized continuous/discrete time-delay system," *Automatica*, vol. 35, no. 8, pp. 1443–1451, 1999.

[18] A. K. Bejczy, W. S. Kim, and S. C. Venema, "The phantom robot: predictive displays for teleoperation with time delay," in *Proceedings., IEEE International Conference on Robotics and Automation*. IEEE, 1990, pp. 546–551.

[19] G. Shi, W. Hönig, X. Shi, Y. Yue, and S.-J. Chung, "Neural-swarm2: Planning and control of heterogeneous multirotor swarms using learned interactions," *IEEE Transactions on Robotics*, 2021.

[20] C. Yu, G. Shi, S.-J. Chung, Y. Yue, and A. Wierman, "Competitive control with delayed imperfect information," *arXiv preprint arXiv:2010.11637*, 2020.

[21] M. Nonhoff and M. A. Müller, "Online gradient descent for linear dynamical systems," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 945–952, 2020.

[22] G. Goel and B. Hassibi, "Competitive control," *arXiv preprint arXiv:2107.13657*, 2021.

[23] Y. Lin, G. Goel, and A. Wierman, "Online optimization with predictions and non-convex losses," *arXiv preprint arXiv:1911.03827*, 2019.

[24] 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.

[25] G. Goel and B. Hassibi, "The power of linear controllers in LQR control," *arXiv preprint arXiv:2002.02574*, 2020.

[26] B. D. Anderson and J. B. Moore, *Optimal control: linear quadratic methods*. Courier Corporation, 2007.

[27] D. Foster and M. Simchowitz, "Logarithmic regret for adversarial online control," in *International Conference on Machine Learning*. PMLR, 2020, pp. 3211–3221.

[28] E. F. Camacho and C. B. Alba, *Model predictive control*. Springer Science & Business Media, 2013.