

On the Regret of \mathcal{H}_∞ Control

Aren Karapetyan, Andrea Iannelli, and John Lygeros

Abstract—The \mathcal{H}_∞ synthesis approach is a cornerstone robust control design technique, but is known to be conservative in some cases. The objective of this paper is to quantify the additional cost the controller incurs planning for the worst-case scenario, by adopting an approach inspired by regret from online learning. We define the *disturbance-reality gap* as the difference between the predicted worst-case disturbance signal and the actual realization. The regret is shown to scale with the norm of this *gap*, which turns out to have a similar structure to that of the certainty equivalent controller with inaccurate predictions, obtained here in terms of the *prediction error* norm.

I. INTRODUCTION

In this work we focus on the control of linear time-invariant (LTI) systems subject to process noise. The optimal control of such systems for the case of stochastic noise is well studied and is referred to as \mathcal{H}_2 control [1]. The controller in this case is optimal for the expected value of the associated cost. When the noise is non-stochastic, a popular approach is to model it as adversarial and having finite energy. In this case, the \mathcal{H}_∞ controller solves the disturbance attenuation problem by ensuring that the system is internally stable and that the infinity norm of the transfer function mapping disturbances to a measurable performance metric is minimized [2]. For LTI systems with quadratic performance metrics this is equivalent to minimizing the induced 2-norm of this transfer function matrix [3]. The closed-loop system is thus robust to any allowable process noise. This robustness proves vital in many applications, especially safety-critical ones. However, as it plans for the worst, \mathcal{H}_∞ control can in practice result in a conservative performance and incur a high cost. The aim of this work is to analyze and quantify the degree of this conservatism using online learning tools [4]. By doing so we study how the concept of robustness from control theory is reflected in a regret formulation.

Recently there has been an increasing interest in quantifying the performance of control algorithms for dynamical systems in terms of regret [5]. Regret compares the incurred cost of a given (online) algorithm with a clairvoyant one that has full knowledge of the problem. When the latter is restricted to a policy class, its associated regret is referred to as *policy regret*. This is contrasted to *dynamic regret* with no restrictions on the optimal controller [6]. The regret of \mathcal{H}_2 controllers has been extensively studied in the literature

[7], [8], [6]. A number of approaches have been proposed for linear time varying models and general convex costs that achieve sublinear *policy regret* (with respect to the control horizon length) [9], [10]. Similar regret bounds are also achieved for cases with model uncertainties using adaptive control techniques [5], [7]. The effect of predictions on the performance of the controllers in the form of *dynamic regret* has been studied in the context of receding horizon control, both for fixed [11] and time varying costs [12]. The performance of a robust receding horizon controller for general costs has been characterized in [13] in terms of the disturbance gain. Regret-optimal controllers are considered in [14], [15] with a classical \mathcal{H}_∞ approach and in [16], [17] by adopting the framework of system level synthesis.

A number of online learning algorithms have been compared to \mathcal{H}_∞ control in numerical experiments to show its relatively poor performance in the absence of online updates [5], [14]. By solving the disturbance attenuation problem, the \mathcal{H}_∞ controller provides robustness guarantees on the possible effects of the noise on the system. This is contrasted to many recent online learning-inspired algorithms, whose objectives differ from that of the robust controller.

In this work we seek to quantify the additional cost that the \mathcal{H}_∞ controller incurs due to its cautious nature, thus allowing direct analytical comparisons with other algorithms. The upper bound of its *dynamic regret* is shown to scale with the norm of the difference between the worst-case and the observed disturbance signals, which we define as the *disturbance-reality gap*. This is achieved using the game-theoretic formulation of the disturbance attenuation problem, which has been extensively studied in the robust control literature [3], [18]. To the best of our knowledge, this is the first work to derive regret bounds for the \mathcal{H}_∞ control. In addition, it aims to promote the application of game-theoretic concepts in the design of novel online algorithms in control. To put the \mathcal{H}_∞ control results in perspective, the regret of a certainty equivalent (CE) controller with an erroneous prediction of the noise signal is also derived. It is shown to be proportional to the square of the norm of the *prediction error*, in agreement with results from [12]. Moreover, when the value of this norm is equal to the norm of the *disturbance-reality gap*, the CE controller attains a lower upper bound as compared to the \mathcal{H}_∞ one. This reinforces the empirically observed good performance of many “optimistic” algorithms [12], [19]. We show that, while similar in structure to that of the CE, the regret of the \mathcal{H}_∞ controller has additional terms arising due to a mismatch between its stabilizing and the linear quadratic regulator (LQR) optimal state feedback gains. An exemplifying numerical example is finally provided.

The authors are with the Department of Information Technology and Electrical Engineering, Automatic Control Laboratory, ETH Zürich, 8092, Zürich, Switzerland. Emails: {akarapetyan, iannelli, lygeros}@control.ee.ethz.ch

This work has been supported by the Swiss National Science Foundation under NCCR Automation (grant agreement 51NF40-180545) and under grant no. 200021-178890.

Notation: For a matrix A the spectral radius and the spectral norm are denoted by $\rho(A)$, and $\|A\|$, respectively; $\lambda_{max}(A)$ denotes the maximum and $\lambda_{min}(A)$ the minimum eigenvalue of A . For a vector $w \in \mathbb{R}^{n_a}$, $w_{[1:a]} := [w_1^\top, \dots, w_a^\top]^\top$ and $\|w\|$ denotes its Euclidean norm. The Kronecker product between two matrices is denoted by \otimes .

II. PRELIMINARIES

We consider a LTI system

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad (1)$$

with initial state $x_0 \in \mathbb{R}^n$ and known matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$. The control input is denoted by $u_t \in \mathbb{R}^m$, and $x_t, w_t \in \mathbb{R}^n$ are the state and disturbance vectors, respectively. The state is assumed to be fully observed. The control objective is to minimize the total accumulated cost over a horizon of length T

$$J^T(u, w; x_0) = x_0^\top Q_T x_0 + \sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t, \quad (2)$$

where $Q, Q_T \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$ are design matrices, $u := [u_0^\top, \dots, u_{T-1}^\top]^\top \in \mathbb{R}^{mT}$ and $w := [w_0^\top, \dots, w_{T-1}^\top]^\top \in \mathbb{R}^{nT}$. Moreover, it is assumed that $Q, Q_T \succeq 0$, $R \succ 0$, and the pair (A, Q) is detectable, the pair (A, B) is stabilizable, and $\|x_0\| \leq X$ for some $X \in \mathbb{R}^+$. We consider finite energy disturbance signals [3], [18] in the space $\mathcal{L}_2(0, T) = \{w : N \rightarrow \mathbb{R}^n : \|w\|_{[0, T-1]} < \infty\}$ where $\|w\|_{[0, T-1]} = \left(\sum_{k=0}^{T-1} \|w(k)\|^2\right)^{\frac{1}{2}}$ over a finite horizon and $\mathcal{L}_2(0, \infty) = \{w : \|w\| < \infty\}$ over an infinite horizon¹. Disturbance signals considered in this paper will generally take values in the space of \mathcal{L}_2 signals with total energy less than or equal to 1, denoted by $\tilde{\mathcal{L}}_2$. This allows us to define the infinite horizon cost $J(u, w; x_0) := \lim_{T \rightarrow \infty} J^T(u, w; x_0)$.

A. The \mathcal{H}_∞ Problem

The robust \mathcal{H}_∞ controller minimizes the induced spectral norm of operator T_μ mapping the disturbance signal $w \in \tilde{\mathcal{L}}_2$ to an output signal $z \in \mathcal{L}_2$ and internally stabilizing the system for the infinite horizon case [3]. The optimization problem can be written as

$$\inf_{\mu \in \mathcal{M}} \sup_{w \in \tilde{\mathcal{L}}_2} \frac{\|T_\mu w\|}{\|w\|} = \inf_{\mu \in \mathcal{M}} \sup_{\|w\|=1} \|T_\mu w\| := \gamma^*, \quad (3)$$

where \mathcal{M} is the set of policies with access to current and past state measurements [18]. With an appropriate definition of the output signal z , the above can be written in terms of the total cost²

$$\gamma^* = \inf_{\mu \in \mathcal{M}} \sup_{\|w\|=1} (J(\mu, w; x_0))^{\frac{1}{2}}, \quad (4)$$

or with J replaced by J^T for the finite horizon case [18]. A policy μ^* that attains the minimum in (4) will be referred to as the \mathcal{H}_∞ controller.

¹Arguments and subscripts are dropped from this notation when clear from the context

²We abuse the notation slightly, $J(\mu, w; x_0)$ denotes here the cost associated with the control inputs generated by the policy μ

B. Regret Definition

The regret is a metric designed to measure the performance of sequential decision making algorithms under uncertainty [4], [20]. This idea has been extended to dynamical systems subject to process noise [5]. Given a noise signal w , the dynamic regret of an algorithm \mathcal{A} is defined as

$$\mathcal{R}_T(\mathcal{A}, w) = J^T(u^{\mathcal{A}}, w; x_0) - J^T(u^*, w; x_0), \quad (5)$$

where the system dynamics are given by (1), $u^{\mathcal{A}}$ is the control input generated by the algorithm \mathcal{A} , and u^* is the optimal offline control input. This is obtained by solving the following optimization problem

$$\begin{aligned} u^* &= \arg \min_u J^T(u, w; x_0) \\ \text{s.t. } &x_{t+1} = Ax_t + Bu_t + w_t, \end{aligned}$$

given full knowledge of the noise realization w and the cost function sequence. Here the optimal inputs are not restricted to any policy set, that is, we consider here the case of dynamic regret.

III. REGRET BOUNDS FOR \mathcal{H}_∞ CONTROL

A. The Worst-Case Disturbance

In this section, the disturbance signal that attains the highest cost for the \mathcal{H}_∞ controller, μ^* , is characterized. For this we adopt the zero-sum game formulation [18], [3] that makes use of the game theoretical toolkit to derive the optimal controller, and is also relevant for getting regret bounds.

For a given $\gamma > 0$, let $J_\gamma^T(u, w; x_0)$ be defined as follows

$$J_\gamma^T(u, w; x_0) = J^T(u, w; x_0) - \gamma^2 \sum_{t=0}^{T-1} \|w_t\|^2,$$

with the infinite horizon case defined as $T \rightarrow \infty$. Signals u and w are considered to be two adversarial players in the game trying to respectively minimize or maximize the cost $J_\gamma^T(u, w; x_0)$. The following *min-max* inequality defines the upper and lower values of the game

$$\inf_{\mu \in \mathcal{M}} \sup_{w \in \tilde{\mathcal{L}}_2} \{J_\gamma^T(\mu, w; x_0)\}^{\frac{1}{2}} \geq \sup_{w \in \tilde{\mathcal{L}}_2} \inf_{\mu \in \mathcal{M}} \{J_\gamma^T(\mu, w; x_0)\}^{\frac{1}{2}},$$

where the upper value is, effectively, the soft constrained version of the original problem (4). If there exists a policy pair (u^{SP}, w^{SP}) such that the lower and upper values are equal then it constitutes a saddle-point (SP) solution. The cost $J_\gamma^{T, SP}$ that these policies attain is called the value of the game.

1) *Finite horizon:* Consider the following condition on γ ,

$$\Xi = \gamma^2 I - M_{t+1}(\gamma) > 0, \quad \forall t \in [0, T-1], \quad (6)$$

where $M_t(\gamma)$ is a sequence of matrices generated by the following coupled generalised Riccati equations

$$\begin{aligned} \Lambda_t(\gamma) &= I + (BR^{-1}B^T - \gamma^{-2}I)M_{t+1}(\gamma), \\ M_t(\gamma) &= Q + A^T M_{t+1}(\gamma) \Lambda_t^{-1}(\gamma) A, \end{aligned} \quad (7)$$

for all $t \in [0, T-1]$ and with $M_T(\gamma) = Q_T$. For the case with perfect state measurements, if γ satisfies (6), then J_γ^T

is strictly convex in u and strictly concave in w [18], [3], and $\Lambda_t(\gamma)$ is nonsingular [21]. In this case the game has a SP solution, and its value is equal to $x_0^T M_0(\gamma) x_0$ [18]. The SP policies can be formulated as a feedback on the current state, and are defined for all $0 \leq t < T$ by

$$u_t^\infty = -R^{-1} B^T M_{t+1}(\gamma) \Lambda_t^{-1}(\gamma) A x_t := -K_t^\infty x_t, \quad (8)$$

$$w_t^\infty = \gamma^{-2} M_{t+1}(\gamma) \Lambda_t^{-1}(\gamma) A x_t. \quad (9)$$

As shown in [21], if both players play the optimal strategies, then the resulting system will evolve according to

$$x_{t+1}^\infty = \Lambda_t^{-1}(\gamma) A x_t^\infty, \quad x_0^\infty = x_0 \quad \forall 0 \leq t < T. \quad (10)$$

It then follows from the ordered interchangeability property of zero sum games that the disturbance policy (9) with x_t replaced by x_t^∞ from (10)

$$w_t^* := \gamma^{-2} M_{t+1}(\gamma) \Lambda_t^{-1}(\gamma) A x_t^\infty \quad \forall 0 \leq t < T \quad (11)$$

constitutes an open loop strategy in saddle point equilibrium with (8). Moreover, [22], [21] show that the feedback policy (8) with $\gamma = \underline{\gamma}$, the lowest possible value that satisfies (6), and with initial state $x_0 = 0$, is the \mathcal{H}_∞ minimax controller. For this value of γ the optimal disturbance attenuation problem (4) and its soft constrained version coincide, and $\underline{\gamma} = \gamma^*$. For any other fixed $\gamma > \gamma^*$ the solution becomes suboptimal. In [23] the case for non-zero initial states is considered and it is shown that a saddle point solution also exists for the original hard-constrained problem (4), given that the energy of the disturbance is at its maximum. Here we only consider the initial states x_0 that belong to the following set

$$X_s = \bigcup_{\gamma > \underline{\gamma}} \{x_0 \in \mathbb{R}^n : \|w^*(\gamma)\| = 1\}, \quad (12)$$

such that the maximum energy is achieved with the policy (11). Thus, for all initial states in the set X_s , a pure strategy saddle point exists for the \mathcal{H}_∞ problem. The optimal strategies are given by (8) and (11) with $\gamma = \bar{\gamma}$ satisfying $\|w^*(\bar{\gamma})\| = 1$, i.e. having the maximum allowable energy.

2) *Infinite horizon*: In the infinite horizon case, the minimal non negative-definite, stationary solution to (7) is considered. In this case, (7) become coupled generalized algebraic Riccati equations (ARE-s) with solutions $M(\gamma)$ and $\Lambda(\gamma)$. For the disturbance attenuation problem (4), a saddle point equilibrium exists also in this case and is given for all $0 \leq t < T$ by

$$u_t^\infty = -R^{-1} B^T M(\bar{\gamma}) \Lambda^{-1}(\bar{\gamma}) A x_t := -K^\infty x_t, \quad (13)$$

$$w_t^\infty = \bar{\gamma}^{-2} M(\bar{\gamma}) \Lambda^{-1}(\bar{\gamma}) A x_t, \quad (14)$$

where $\bar{\gamma}$ is obtained through a trial-and-error method to satisfy $\|w^*(\bar{\gamma})\|^2 = 1$, along with certain conditions that allow the minimization problem to be well-posed [24], [25]. The worst-case open loop disturbance signal can then be defined similar to the finite horizon case

$$w_t^* := \bar{\gamma}^{-2} M(\bar{\gamma}) \Lambda^{-1}(\bar{\gamma}) A x_t^\infty \quad \forall t \geq 0, \quad (15)$$

where

$$x_{t+1}^\infty = \Lambda^{-1}(\bar{\gamma}) A x_t^\infty, \quad x_0^\infty = x_0 \quad \forall t \geq 0. \quad (16)$$

B. Regret Analysis

In this section an upper bound for the regret of the \mathcal{H}_∞ problem is obtained. We consider first the infinite horizon case with the minimax controller (13), then the finite one with the controller (8). For a given control input u and disturbance signal w , at a generic timestep i (with $0 \leq i < T$) the cost-to-go function is defined as

$$J_i^T(u^\infty, w; x_i) = x_i^\top Q_T x_T + \sum_{t=i}^{T-1} x_t^\top Q x_t + u_t^\top R u_t, \quad (17)$$

and $J(u, w; x_i) := \lim_{T \rightarrow \infty} J_i^T(u, w; x_i)$.

1) *Infinite horizon*: The following result is introduced to characterize the cost in the infinite horizon case.

Lemma 3.1: For all F with $\rho(F) < 1$, $P_i \succ 0 \quad \forall i \geq 0$ and $Q \succ 0$, the iteration $P_{i+1} = F^\top P_i F + Q$ converges to a unique value P .

Following [6], [11], we claim that $J(u^\infty, w; x_i)$ can be expressed in the form of an extended quadratic function, as formulated in the following lemma.

Lemma 3.2: The infinite horizon cost of the \mathcal{H}_∞ controller (13), solving the disturbance attenuation problem (4), is given by $J(u^\infty, w; x_i) = x_i^\top P^\infty x_i + x_i^\top v_i^\infty + q_i^\infty$, with $P^\infty \in \mathbb{R}^{n \times n}$, $v_i^\infty \in \mathbb{R}^n$, $q_i^\infty \in \mathbb{R} \quad \forall i \geq 0$ given in (19).

Proof: The finite horizon cost for the controller (13) is claimed to be given by $J_i^T(u^\infty, w; x_i) = x_i^\top P_i^\infty x_i + x_i^\top v_i^\infty + q_i^\infty$, with some $P_i^\infty \in \mathbb{R}^{n \times n}$, $v_i^\infty \in \mathbb{R}^n$, $q_i^\infty \in \mathbb{R}$. Indeed, for $i = T$ this holds trivially, with $P_T^\infty = Q_T$ and $v_T^\infty, q_T^\infty = 0$. Then if the claim holds at $i + 1$ the cost-to-go at i satisfies

$$\begin{aligned} J_i^T(u^\infty, w; x_i) &= x_i^\top Q x_i + u_i^{\infty \top} R u_i^\infty \\ &+ (A x_i + B u_i^\infty + w_i)^\top P_{i+1}^\infty (A x_i + B u_i^\infty + w_i) \\ &+ (A x_i + B u_i^\infty + w_i)^\top v_{i+1}^\infty + q_{i+1}^\infty \\ &= u_i^{\infty \top} (R + B^\top P_{i+1}^\infty B) u_i^\infty \\ &+ 2u_i^{\infty \top} B^\top (P_{i+1}^\infty A x_i + P_{i+1}^\infty w_i + \frac{v_{i+1}^\infty}{2}) \\ &+ x_i^\top Q x_i + (A x_i + w_i)^\top P_{i+1}^\infty (A x_i + w_i) \\ &+ (A x_i + w_i)^\top v_{i+1}^\infty + q_{i+1}^\infty. \end{aligned} \quad (18)$$

Substituting $u_i^\infty = -K^\infty x_i$, defining $F^\infty = A - B K^\infty$ and grouping terms leads to

$$\begin{aligned} J^T(u^\infty; x_i) &= x_i^\top \underbrace{(F^{\infty \top} P_{i+1}^\infty F^\infty + Q + K^{\infty \top} R K^\infty)}_{P_i^\infty} x_i + \\ &x_i^\top \underbrace{(2P_{i+1}^\infty w_i + v_{i+1}^\infty)}_{v_i^\infty} + \underbrace{w_i^\top v_{i+1}^\infty + w_i^\top P_{i+1}^\infty w_i + q_{i+1}^\infty}_{q_i^\infty}. \end{aligned} \quad (19)$$

The claim then follows by induction. To complete the proof, we note that $\rho(F^\infty) < 1$ [18] and invoke Lemma 3.1 to replace P_i^∞ and P_{i+1}^∞ by P^∞ in (19). We note that while

P^∞ , v_i^∞ and q_i^∞ are independent of x_i , the last two depend on the noise realisation w . \square

Remark: Lemma 3.2 also holds for any stabilising state feedback matrix K^s , with the coefficients in the cost-to-go appropriately defined.

The optimal offline controller, as defined in Section II-B, has access to all future disturbances w and minimizes the cost function (2) without constraining the inputs to a policy set. It is shown in [6] that for the infinite horizon case the controller has the following form

$$u_t^* = -Kx_t - (R + B^\top PB)^{-1}B^\top \sum_{i=0}^{\infty} (F^\top)^i P w_{t+i}, \quad (20)$$

where $K = -(R + B^\top PB)^{-1}B^\top PA$, P is the solution of the discrete ARE and $F := A - BK$, with $\rho(F) < 1$. Moreover, the cost-to-go at a state x_i , $J(u^*, w; x_i)$ has the same extended quadratic structure as in (19) with the following coefficients

$$P = F^\top PF + Q + K^\top RK, \quad (21)$$

$$v_i = F^\top (2Pw_i + v_{i+1}) = 2 \sum_{j=0}^{\infty} (F^\top)^{j+1} P w_{i+j}, \quad (22)$$

$$q_i = q_{i+1} + w_i^\top v_{i+1} + w_i^\top P w_i - G_i^\top H G_i, \quad (23)$$

where $G_i := \sum_{j=0}^{\infty} (F^\top)^j P w_{i+j}$ and $H := B(R + B^\top PB)^{-1}B^\top$ [6], [11]. Using the result of Lemma 3.2, the regret of the \mathcal{H}_∞ controller for a given w is then

$$\begin{aligned} \mathcal{R}(\mathcal{H}_\infty, w) &:= \lim_{T \rightarrow \infty} \mathcal{R}_T(\mathcal{H}_\infty, w) \\ &= J(u^\infty, w; x_0) - J(u^*, w; x_0) \\ &= x_0^\top (P^\infty - P)x_0 + x_0^\top (v_0^\infty - v_0) + q_0^\infty - q_0. \end{aligned} \quad (24)$$

The *disturbance-reality gap* Δw is defined as

$$\Delta w := w - w^*. \quad (25)$$

This vector is the difference between the disturbance w , experienced by the system and the worst-case disturbance, w^* , assumed by the \mathcal{H}_∞ controller, defined in (15). The main result of this paper is formulated in the following theorem.

Theorem 3.1: The \mathcal{H}_∞ controller, that solves the disturbance attenuation problem (4) attains dynamic regret

$$\mathcal{R}(\mathcal{H}_\infty, w) \leq k_1 \|\Delta w\| + k_2 \|\Delta w\|^2$$

for all initial states in (12) and constants $k_1, k_2 \in \mathbb{R}^+$, given below in (26).

Proof: Regret (24) can be written in terms of Δw and w^*

$$\begin{aligned} \mathcal{R}(\mathcal{H}_\infty, w) &= x_0^\top (P^\infty - P)x_0 \\ &+ \sum_{i=0}^{\infty} \left(2x_0^\top \left((F^{\infty\top})^{i+1} P^\infty - (F^\top)^{i+1} P \right) (w_i^* + \Delta w_i) \right. \\ &+ (w_i^* + \Delta w_i)^\top (v_{i+1}^\infty - v_{i+1}) + G_i^\top H G_i \\ &\left. + (w_i^* + \Delta w_i)^\top (P^\infty - P)(w_i^* + \Delta w_i) \right). \end{aligned}$$

Since the \mathcal{H}_∞ controller is in saddle point equilibrium with w^* , the optimal offline controller with knowledge of the future disturbances, will attain the same cost as the \mathcal{H}_∞ controller if $\Delta w = 0$. Hence $R(\mathcal{H}_\infty, w^*) = 0$, leaving

$$\begin{aligned} \mathcal{R}(\mathcal{H}_\infty, w) &= \\ &\lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} \underbrace{\left(2x_0^\top \left((F^{\infty\top})^{i+1} P^\infty - (F^\top)^{i+1} P \right) \Delta w_i \right)}_{a_i} \\ &+ \underbrace{w_i^{*\top} (v_{i+1}^\infty - v_{i+1})}_{b_i} + \underbrace{\Delta w_i^\top (v_{i+1}^\infty - v_{i+1})}_{c_i} \\ &+ \underbrace{\Delta w_i^\top (P^\infty - P) \Delta w_i}_{d_i} + \underbrace{2\Delta w_i^\top (P^\infty - P) w_i^*}_{e_i} \\ &+ \underbrace{G_i^{\Delta w\top} H G_i^{\Delta w}}_{f_i} + \underbrace{2G_i^{\Delta w\top} H G_i w^*}_{g_i}, \end{aligned}$$

where $v_i^{\Delta w} := F^\top (2P\Delta w_i + v_{i+1}^{\Delta w})$, $v_i^{\infty\Delta w} := F^{\infty\top} (2P^\infty \Delta w_i + v_{i+1}^{\infty\Delta w})$ and $G_i^{\Delta w} := \sum_{j=0}^{T-i-1} (F^\top)^j P \Delta w_{i+j}$; the corresponding expressions with w^* and w are defined analogously. This reformulation of regret is then used to upper bound it in terms of the norm of Δw . For d_i and e_i

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} d_i &= \Delta w^\top \left(\Delta P \otimes I_{nT} \right) \Delta w \leq 2\|\Delta w\|^2 \|\bar{P}\| \\ \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} e_i &= 2\Delta w^\top \left(\Delta P \otimes I_{nT} \right) w^* \leq 4\|\Delta w\| \|\bar{P}\|, \end{aligned}$$

where $\Delta P := P^\infty - P$, $\|\bar{P}\| := \max\{\|P^\infty\|, \|P\|\}$, and using $\|w^*\| = 1$ and the fact [26] that $\|A \otimes B\| = \|A\| \|B\|$. The sum of terms a_i can be written as

$$\sum_{i=0}^{T-1} a_i = 2x_0^\top \Delta L_v \Delta w,$$

where $\Delta L_v \in \mathbb{R}^{n \times nT}$ is a block matrix with the term $(F^{\infty\top})^i P^\infty - (F^\top)^i P$ on its i -th block column for all $1 \leq i \leq T$. From Gelfand's formula it can be shown that there exists a constant $c > 1$ such that $\|F^i\| \leq c\lambda^i$ and $\|(F^\infty)^i\| \leq c(\lambda^\infty)^i$ for all $i \geq 1$ with $\lambda := \frac{1+\rho(F)}{2} < 1$, $\lambda^\infty = \frac{1+\rho(F^\infty)}{2}$, since $\rho(F) < 1$, $\rho(F^\infty) < 1$. Hence

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} a_i &\leq 2c\|x_0\| \|\bar{P}\| \left(\frac{\lambda^\infty}{1-\lambda^\infty} + \frac{\lambda}{1-\lambda} \right) \|\Delta w\| \\ &\leq 4cX \|\Delta w\| \|\bar{P}\| \left(\frac{\bar{\lambda}}{1-\bar{\lambda}} \right), \end{aligned}$$

where $\bar{\lambda} := \max\{\lambda^\infty, \lambda\}$. For the term with c_i

$$\sum_{i=0}^{T-1} c_i = 2\Delta w_{[0:T-2]}^\top \Delta L_u w_{[1:T-1]},$$

where $\Delta L_u \in \mathbb{R}^{n(T-1) \times n(T-1)}$ is an upper triangular block Toeplitz matrix, such that for all $1 \leq i < T$ and $i \leq j < T$,

the matrix on the i -th block row and j -th block column is $(F^{\infty\top})^{j-i+1}P^\infty - (F^\top)^{j-i+1}P$. It follows that

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} c_i &\leq 2c \|\Delta w\| \left(\|P^\infty\| \frac{\lambda^\infty}{1-\lambda^\infty} + \|P\| \frac{\lambda}{1-\lambda} \right) \\ &\leq 4c \|\Delta w\| \|\bar{P}\| \left(\frac{\bar{\lambda}}{1-\bar{\lambda}} \right), \end{aligned}$$

where we have used the properties of block Toeplitz matrices. We can similarly get the same bound for $\sum_{i=0}^{T-1} b_i$. With a similar reformulation for f_i and g_i

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} f_i &\leq \|\Delta w\|^2 \|H\| \|P\|^2 \frac{c^2}{(1-\lambda)^2}, \\ \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} g_i &\leq 2 \|\Delta w\| \|H\| \|P\|^2 \frac{c^2}{(1-\lambda)^2}. \end{aligned}$$

Summing the terms and setting

$$\begin{aligned} k_2 &= 2 \|\bar{P}\| + \|H\| \|P\|^2 \frac{c^2}{(1-\lambda)^2} \\ k_1 &= 4 \|\bar{P}\| + 4c \|\bar{P}\| (2+X) \left(\frac{\bar{\lambda}}{1-\bar{\lambda}} \right) \\ &\quad + 2 \|H\| \|P\|^2 \frac{c^2}{(1-\lambda)^2}, \end{aligned} \quad (26)$$

completes the proof. \square

We note that the constraint of the noise signal having a unit energy is without loss of generality and the same result can also be attained by modifying the set (12).

2) *Finite Horizon*: A similar bound is obtained for the dynamic regret of the finite horizon controller.

Theorem 3.2: The \mathcal{H}_∞ controller, that solves the disturbance attenuation problem (4) for a horizon length T , attains dynamic regret

$$\mathcal{R}_T(\mathcal{H}_\infty, w) \leq k'_1 \|\Delta w\| + k'_2 \|\Delta w\|^2$$

for all initial states in (12) and constants $k'_1, k'_2 \in \mathbb{R}^+$, given below in (27).

The proof follows closely the structure for the infinite horizon case. With the same induction hypothesis it is shown that there exist $P_i^\infty \in \mathbb{R}^{n \times n}$, $v_i^\infty \in \mathbb{R}^n$ and $q_i^\infty \in \mathbb{R}$, such that at time step i , $0 \leq i < T$, $J_i^T(u^\infty, w, x_i) = x_i^\top P_i^\infty x_i + x_i^\top v_i^\infty + q_i^\infty$. The optimal offline controller for the finite horizon case has a similar structure to (20) and $J_i^T(u^*, w, x_i) = x_i^\top P_i x_i + x_i^\top v_i + q_i \forall 0 \leq i < T$ for some $P_i \in \mathbb{R}^{n \times n}$, $v_i \in \mathbb{R}^n$ and $q_i \in \mathbb{R}$ [6], [11].

The regret of the finite horizon \mathcal{H}_∞ controller is then the difference of the two extended quadratic functions,

$$\mathcal{R}_T(\mathcal{H}_\infty, w) = x_0^\top (P_0^\infty - P_0) x_0 + x_0^\top (v_0^\infty - v_0) + q_0^\infty - q_0.$$

Substituting the expressions for the coefficients, the above can be written in terms of the *disturbance-reality* gap. Using the same argument of equal costs for the worst-case disturbance signal, k'_1 and k'_2 are defined as follows to complete

the proof

$$\begin{aligned} k'_2 &= \|\bar{P}'\| \left(2 + \bar{\tau}^2 \|\bar{H}\| \|\bar{P}'\| \frac{(1-\bar{\eta}^T)^2}{(1-\bar{\eta})^2} \right) \\ k'_1 &= 2 \|\bar{P}'\| \left(2 + 2\bar{\tau}\bar{\eta}(2+X) \left(\frac{1-\bar{\eta}^T}{1-\bar{\eta}} \right) \right. \\ &\quad \left. + \bar{\tau}^2 \|\bar{H}\| \|\bar{P}'\| \frac{(1-\bar{\eta}^T)^2}{(1-\bar{\eta})^2} \right), \end{aligned} \quad (27)$$

where $\bar{P}' \succeq P_i, P_i^\infty$ and $\bar{H} \succeq B(R+B^\top P_i B)^{-1} B^\top \forall 0 \leq i \leq T$. The constants $\bar{\tau} := \sqrt{\frac{\lambda_{\max}(\bar{P}')}{\lambda_{\min}(Q)}}$ and $\bar{\eta} := \sqrt{1 - \frac{1}{\bar{\tau}^2}} < 1$. The full proof can be found in the extended version of the paper [27].

Remark: One could use the same technique to obtain similar results for time-varying system matrices in (1) and cost functions (2). We forego this extra generality in the interest of consistent notation with the infinite horizon case.

IV. CE OPTIMISTIC CONTROLLER

In this section, the certainty equivalent optimistic controller that has an inaccurate prediction $\bar{w} \in \mathcal{L}_2$ of the disturbance signal and acts optimally with respect to it is considered. It solves the optimization problem (2) subject to $x_{t+1} = Ax_t + Bu_t + \bar{w}_t$. For simplicity only the infinite horizon case is considered, however, the results for the finite horizon are derived analogously. The infinite horizon CE controller is the same as in (20), only with feedback on \bar{w}

$$u_t^{\text{CE}} = -Kx_t - (R + B^\top P B)^{-1} B^\top \sum_{i=0}^{\infty} F^{\top i} P \bar{w}_{t+i}. \quad (28)$$

Its regret is shown to be proportional to $\|\Delta \bar{w}\|^2$, where $\Delta \bar{w} := \bar{w} - w$, is the error vector between the predicted and the observed true disturbance signals, or the *prediction error*. The result is formulated in the following proposition.

Proposition 4.1: The CE optimistic controller (28) attains dynamic regret that is upper bounded by

$$\mathcal{R}(\text{CE}, w) \leq \|\Delta \bar{w}\|^2 \|H\| \|P\|^2 \frac{c^2}{(1-\lambda)^2}.$$

Proof: The induction hypothesis that for all timesteps the cost-to-go is an extended quadratic function is still valid. Following the same technique as in (18), regret is given as

$$\begin{aligned} \mathcal{R}(\text{CE}, w) &= J(u^{\text{CE}}, w; x_0) - J(u^*, w, x_0) = \bar{q}_0 - q_0 \\ &= \sum_{i=0}^{\infty} \left(\sum_{j=0}^{\infty} (F^\top)^j P \Delta \bar{w}_{i+j} \right)^\top H \left(\sum_{j=0}^{\infty} (F^\top)^j P \Delta \bar{w}_{i+j} \right). \end{aligned}$$

Similar to the proof of the \mathcal{H}_∞ controller, the required upper bound is then achieved. \square

Comparing the above with the regret for the infinite horizon \mathcal{H}_∞ controller we note that in addition to the dependence on $\|\Delta w\|$, the regret bound for the \mathcal{H}_∞ also has additional terms in the coefficient of $\|\Delta w\|^2$. Thus, given equal $\|\Delta w\|$ and $\|\Delta \bar{w}\|$, \mathcal{H}_∞ has a strictly higher regret upper bound compared to the certainty equivalent controller. The additional terms in the upper bound of \mathcal{H}_∞ regret are

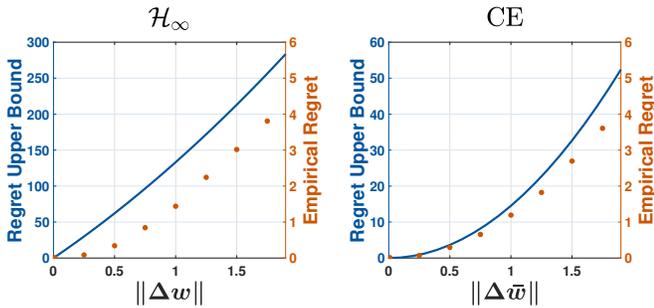


Fig. 1. The theoretical regret upper bounds and worst-case simulated regret for the \mathcal{H}_∞ and CE controllers.

due to the “sub-optimal” state feedback gain on the state. This results in a mismatch between P^∞ and P , as well as v_0^∞ and v_0 , the coefficients of the initial state; this is not the case for the CE controller. This makes the *gap* explicitly dependent on the initial state leading to additional terms in the regret bound.

V. NUMERICAL EXAMPLE

A simple system, also considered in [23], with $A = 1$, $B = 1$ and cost matrices $Q = 1$, $R = 1$ is controlled using both the \mathcal{H}_∞ and the CE finite horizon controllers with $T = 100$. For each fixed $\|\Delta w\|$ and $\|\Delta \bar{w}\|$, a number of random noise signals are generated. The system evolution is then simulated starting from an initial condition $x_0 = 4 \in X_s$. The parameter $\bar{\gamma}$ is found for this initial state using a trial-and-error method as described in [23]. The dynamic regret of both controllers is then calculated and the highest regret (for each $\|\Delta w\|$ and $\|\Delta \bar{w}\|$) is plotted in Figure 1 along with the upper bounds obtained in this work. It is inferred from the plots that the analytic bounds capture the order of the empirically calculated worst-case regret. Preliminary numerical tests show that they can become tighter for certain adversarial noise signals.

VI. CONCLUSIONS

The \mathcal{H}_∞ algorithm is considered in the context of dynamic regret to characterize its extra cost due to planning for the worst-case disturbance realization. The upper bound of this regret is shown to scale with the norm of the *gap* between the worst-case predicted by the \mathcal{H}_∞ controller and the true one. This result is compared with the CE optimistic controller with erroneous predictions. While both controllers have similar regret upper bound structures, for equal *disturbance-reality gap* and *prediction error* norms, \mathcal{H}_∞ 's regret attains a strictly higher upper bound. A numerical example is presented to show that the order of the worst-case simulated regret is captured by the theoretical bounds. A possible direction for further research, is the consideration of the case where no saddle point exists for the \mathcal{H}_∞ problem. The results can also provide insights on the development of algorithms that estimate the future disturbances online.

Acknowledgements: The authors thank Efe Balta for fruitful discussions on the topic.

REFERENCES

- [1] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic estimation and control: a unified approach to H_2 and H_∞ theories*. SIAM, 1999.
- [2] K. Zhou and J. C. Doyle, *Essentials of robust control*, vol. 104. Prentice hall Upper Saddle River, NJ, 1998.
- [3] M. Green and D. J. Limebeer, *Linear robust control*. Courier Corporation, 2012.
- [4] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006.
- [5] E. Hazan, S. Kakade, and K. Singh, “The nonstochastic control problem,” in *Algorithmic Learning Theory*, pp. 408–421, PMLR, 2020.
- [6] G. Goel and B. Hassibi, “The power of linear controllers in LQR control,” *arXiv preprint arXiv:2002.02574*, 2020.
- [7] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, “Regret bounds for robust adaptive control of the linear quadratic regulator,” *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [8] M. Simchowitz and D. Foster, “Naive exploration is optimal for online LQR,” in *International Conference on Machine Learning*, pp. 8937–8948, PMLR, 2020.
- [9] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, “Online control with adversarial disturbances,” in *International Conference on Machine Learning*, pp. 111–119, PMLR, 2019.
- [10] D. Foster and M. Simchowitz, “Logarithmic regret for adversarial online control,” in *International Conference on Machine Learning*, pp. 3211–3221, PMLR, 2020.
- [11] C. Yu, G. Shi, S.-J. Chung, Y. Yue, and A. Wierman, “The power of predictions in online control,” *arXiv preprint arXiv:2006.07569*, 2020.
- [12] R. Zhang, Y. Li, and N. Li, “On the regret analysis of online LQR control with predictions,” in *2021 American Control Conference (ACC)*, pp. 697–703, IEEE, 2021.
- [13] D. Muthirayan, D. Kalathil, and P. P. Khargonekar, “Online robust control of linear dynamical systems with prediction,” *arXiv preprint arXiv:2111.15063*, 2021.
- [14] G. Goel and B. Hassibi, “Regret-optimal control in dynamic environments,” *arXiv preprint arXiv:2010.10473*, 2020.
- [15] O. Sabag, G. Goel, S. Lale, and B. Hassibi, “Regret-optimal controller for the full-information problem,” in *2021 American Control Conference (ACC)*, pp. 4777–4782, IEEE, 2021.
- [16] A. Martin, L. Furieri, F. Dörfler, J. Lygeros, and G. F. Trecate, “Safe control with minimal regret,” *arXiv preprint arXiv:2203.00358*, 2022.
- [17] A. Didier, J. Sieber, and M. N. Zeilinger, “A system level approach to regret optimal control,” *arXiv preprint arXiv:2202.13763*, 2022.
- [18] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.
- [19] D. Muthirayan, J. Yuan, D. Kalathil, and P. P. Khargonekar, “Online learning for receding horizon control with provable regret guarantees,” *arXiv preprint arXiv:2111.15041*, 2021.
- [20] E. Hazan, “Introduction to online convex optimization,” *arXiv preprint arXiv:1909.05207*, 2019.
- [21] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.
- [22] D. Limebeer, M. Green, and D. Walker, “Discrete-time \mathcal{H}_∞ control,” in *Proceedings of the 28th IEEE Conference on Decision and Control*, pp. 392–396, IEEE, 1989.
- [23] G. Didinsky and T. Başar, “Design of minimax controllers for linear systems with non-zero initial states under specified information structures,” *International Journal of Robust and Nonlinear Control*, vol. 2, no. 1, pp. 1–30, 1992.
- [24] E. Mageirou, “Values and strategies for infinite time linear quadratic games,” *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 547–550, 1976.
- [25] J. Willems, “Least squares stationary optimal control and the algebraic riccati equation,” *IEEE Transactions on automatic control*, vol. 16, no. 6, pp. 621–634, 1971.
- [26] P. Lancaster and H. K. Farahat, “Norms on direct sums and tensor products,” *mathematics of computation*, vol. 26, no. 118, pp. 401–414, 1972.
- [27] A. Karapetyan, A. Iannelli, and J. Lygeros, “On the regret of \mathcal{H}_∞ control,” *arXiv preprint arxiv.2203.16237*, 2022.