2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

An Online Learning Analysis of Minimax Adaptive Control

Venkatraman Renganathan, Andrea Iannelli, and Anders Rantzer

Abstract— **We present an online learning analysis of minimax adaptive control for the case where the uncertainty includes a finite set of linear dynamical systems. Precisely, for each system inside the uncertainty set, we define the model-based regret by comparing the state and input trajectories from the minimax adaptive controller against that of an optimal controller in hindsight that knows the true dynamics. We then define the total regret as the worst case model-based regret with respect to all models in the considered uncertainty set. We study how the total regret accumulates over time and its effect on the adaptation mechanism employed by the controller. Moreover, we investigate the effect of the disturbance on the growth of the regret over time and draw connections between robustness of the controller and the associated regret rate.**

I. Introduction

The interplay between machine learning, system identification and adaptive control has unveiled a fertile area of research which has the potential to answer some of the standing research questions in the field of learning-based control. Recent advances in online learning techniques have provided new perspectives on the design of algorithms where unknown systems can be controlled by acquiring knowledge through repeated interactions with the unknown environment [1]. This has close connections with adaptive control [2] and in general with learning-based control techniques [3]. Minimax adaptive control is taken in this work as a prototypical example of the latter line of works to draw connections with regret, i.e. the performance metrics used in online learning. Design of minimax control for uncertain systems was investigated as early as in [4], [5]. Subsequently, the design of minimax adaptive control was investigated for scalar systems with unknown input matrix sign in [6], for finite sets of linear systems in [7], [8] and for the output feedback case in [9], respectively. There have been earlier works on robust adaptive control in [10], [11] where uncertainties in system dynamics were considered. Minimax adaptive control problems are generally challenging as obtaining exact ℓ_2-gain bounds as explained in [7], [12], [13] can be hard for multiple input multiple outputs systems with finite set of linear models and optimality can only be achieved if the exploration and exploitation trade-off is exactly captured.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 834142 (Scalable Control). V. Renganathan and A. Rantzer are with the Department of Automatic Control, Lund University, Sweden and are also members of the ELLIIT Strategic Research Area in Lund University. A. Iannelli is with the Institute of Systems Theory and Automatic Control, University of Stuttgart, Germany. Emails: (venkatraman.renganathan,anders.rantzer)@control.lth.se, andrea.iannelli@ist.uni-stuttgart.de.

The recent interest developed towards analyzing control algorithms for systems with unknown dynamics through the lens of regret analysis has the promise to enable a better understanding of this trade-off. There are quantities that are of interest but are unknown in advance to the online controller. We refer to such unknown entity as *Quantity of Interest (QI)*. Lack of knowledge about a QI determines an accumulated cost, with respect to a control designed with perfect knowledge, which denotes the notion of regret. For instance, the growth of expected regret in linear quadratic control was investigated in [14] when matrices (A, B) were unknown. Regret bounds have been investigated in [15] for adaptive control problems in stochastic setting. This paper proposes an online learning analysis of minimax adaptive control of linear-time invariant systems featuring adversarial disturbance and a priori knowledge of a finite set of systems. One of the distinctive novelty is a new definition of regret, suitable for this setting in which the QIs are both the system dynamics and the exogenous disturbance. From an online learning perspective, efficient adaptive control algorithms are characterized by limiting the growth of regret over time. To quantify the regret, we usually require an optimal control policy (policy regret) or a sequence of best control actions (dynamic regret) available in hindsight as in [16]. Here, we propose studying the policy regret associated with the minimax adaptive controller by comparing it against the standard $\mathcal{H}_\infty$ control which knows the dynamics.
Recently, [17] investigated the regret of robustness of an $\mathcal{H}_\infty$ controller (whose QI is just the adversarial disturbance) when compared to an oracle controller which has knowledge of the future disturbance trajectory. Also related is the work in [18], which investigated the regret analysis for the generic non-stochastic control problem and their system identification approach employed random inputs before controlling it using disturbance-based policy. Similarly, [19] studied online control with adversarial disturbances and proposed a disturbance action control policy based efficient algorithm to obtain nearly tight regret bounds. On the contrary, our work looks at nonlinear adaptive state feedback policy which *concurrently* controls the system under adversarial disturbance and implicitly learns the system dynamics. This gives rise to an interesting trade-off in the adversary strategy, whereby the worst-case disturbance is the one that delays the learning process of the controller while minimizing the energy spent (which is penalized in the total cost).

Contributions: We provide a detailed analysis for the minimax adaptive control algorithm proposed in [7] with

the aim to improve our understanding on the role of the adaptation mechanism and the adversary disturbance on the regret. Since an explicit expression for the optimal minimax adaptive controller is not known, we apply our analysis to the candidate sub-optimal minimax adaptive control algorithm[1] developed in [7], [8]. Specifically, the main contributions are:

1) Definition of the: *model-based regret* corresponding to a specific model in the uncertainty set characterizing the accumulated cost with respect to an optimal controller in hindsight which knows the true dynamics; *total regret* as the worst-case model-based regret corresponding to any model in the uncertainty set.
2) Construction of an adversarial disturbance policy which provably prevents the minimax adaptive controller from learning the true dynamics (Theorem 1).
3) Despite the possible difficulty in the identification of the true dynamics, we show that the minimax adaptive controller enjoys a sub-linear regret rate with respect to the best $\mathcal{H}_\infty$ controller in hindsight (Theorem 2).

The rest of the paper is organised as follows. The problem formulation is discussed in §II. The online learning analysis is performed in §III, and some of its features are further elucidated through numerical simulation in §IV. Finally, the main findings of the paper are summarized in §V.

Notation and Preliminaries. The cardinality of the set A is denoted by $|A|$. The set of real numbers, integers and the natural numbers are denoted by $\mathbb{R}, \mathbb{Z}$, and $\mathbb{N}$ respectively. For a matrix $A \in \mathbb{R}^{n\times n}$, we denote its transpose and its trace by $A^\top$ and $\mathbf{Tr}(A)$ respectively. We denote by $\mathbb{S}^n$, the set of symmetric matrices in $\mathbb{R}^{n\times n}$. For $A \in \mathbb{S}^n$, we write $A \succ 0$ and $A \succeq 0$ to say that A is positive definite and positive semi-definite, respectively. An identity matrix of dimension n is denoted by I_n. Given $x \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times n}$, the notations $\|x\|_A^2$ and $\|B\|_A^2$ mean $x^\top Ax$ and $\mathbf{Tr}\left(B^\top AB\right)$ respectively. A signal $\{x_k\}$ is said to be in ℓ_2 space if it has finite energy meaning that $\sum_{k=0}^{\infty} x_k^2 < \infty$. For any time $T \in \mathbb{N}$, if the truncation of a signal $\{x_k\}$ to the interval $[0, T]$ lies in the ℓ_2 space, then the signal is said to be lying in the extended ℓ_2 space denoted by ℓ_{2e}.

II. Problem Formulation Using Minimax Adaptive Control

In this section we introduce the minimax adaptive control subject of our investigations through online learning.

A. Minimax adaptive control with finite set of linear systems

Consider the following discrete-time linear system

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad k \in \mathbb{N}, \tag{1}$$

where $x_k \in \mathbb{R}^n$ and $u_k \in \mathbb{R}^m$ denote the system states and control inputs, respectively, and the additive disturbance $w_k \in \mathbb{R}^n$ is assumed to be adversarial. The true system matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are unknown but assumed to belong to a set $\mathcal{M}$ with $|\mathcal{M}| = \mathcal{F} \in \mathbb{N}$ defined such that $M_i := (A_i, B_i) \in \mathcal{M}, i = 1, \ldots, \mathcal{F}$, where all pairs are assumed throughout to be stabilizable. For instance, control of a discrete-time linearized inverted pendulum dynamics falls under the above setting when the pendulum length is uncertain. Generally, minimax adaptive control approach can be a suitable design solution when multiple systems who do not share common Lyapunov function need to be controlled by a single controller. Let us denote by Π the set of admissible control policies such that

$$u_k = \pi_k\left(x_0, x_1, \ldots, x_k, u_0, \ldots, u_{k-1}\right), \quad \pi_k \in \Pi. \tag{2}$$

An optimal adaptive control policy should interact with the system in order to extract information about the unknown system matrices A, B while also guaranteeing good performance and robustness to the adversarial disturbance. This can be achieved by optimizing the following minimax cost

$$\inf_{\pi\in\Pi} \underbrace{\sup_{w,A,B} \sum_{k=0}^{\infty} \left(c(x_k^\pi, u_k^\pi, Q, R) - \gamma^2 \|w_k\|^2\right)}_{J_\pi(x_0,\gamma)}. \tag{3}$$

where $c(x^\pi, u^\pi, Q, R) := \|x^\pi\|_Q^2 + \|u^\pi\|_R^2$ for given penalty matrices $Q \succ 0, R \succ 0$; x^π denotes the evolution of the state of (1) starting from x_0 under the control input u^π from the policy π; and $\gamma > 0$ quantifies the desired level of robustness to the external disturbance (higher γ resulting in weaker robustness requirements). The optimal minimax control policy $\pi^\dagger$ and the associated cost are given by

$$\pi^\dagger := \operatorname*{argmin}_{\pi\in\Pi} J_\pi(x_0, \gamma), \quad J^\dagger(x_0, \gamma) := J_{\pi^\dagger}(x_0, \gamma) \tag{4}$$

and the resulting disturbance attenuation level achieved by the control policy $\pi^\dagger$ from disturbance to the regulated output $\zeta := \begin{bmatrix} x^\top & u^\top \end{bmatrix}^\top$ is denoted by $\gamma^\dagger$ and is defined as

$$\gamma^\dagger := \sqrt{\sup_{w^\dagger \neq 0} \frac{\sum_{k=0}^{\infty} c(x_k^{\pi^\dagger}, u_k^{\pi^\dagger}, Q, R)}{\sum_{k=0}^{\infty} \left\|w_k^\dagger\right\|^2}}. \tag{5}$$

This formulation provides a family of minimax control policy parameterized by γ, which are guaranteed to exist $\forall \gamma > \gamma^\dagger$. We cast the problem as a zero-sum dynamic game with the control policy π being the minimizing player and the adversaries (w, A, B) being the maximizing players [7]. The solution boils down to solving a minimax dynamic programming problem, which is intractable in most cases. An approximate (i.e. sub-optimal) solution has been recently proposed in [7], [8], and this will be the subject of this study. The following lemma summarizes the main result of [7], i.e. an explicit expression for an adaptive controller satisfying a pre-specified ℓ_2-gain bound from disturbance to error.

Lemma 1. *Given a compact set of linear models $\mathcal{M}$, and positive definite penalty matrices $Q \in \mathbb{R}^{n\times n}, R \in \mathbb{R}^{m\times m}$, suppose that there exists $K_1, \ldots, K_{\mathcal{F}} \in \mathbb{R}^{m\times n}$ and matrices $P_{ij} \in \mathbb{R}^{n\times n}$ with $0 \prec P_{ij} = P_{ji} \prec \gamma^2 I$ such that*

$$\|x\|_{P_{il}}^2 \geq \|x\|_Q^2 + \|K_l x\|_R^2 - \gamma^2 \left\|(\bar{A}_{il} - \bar{A}_{jl})x/2\right\|^2 + \left\|(\bar{A}_{il} + \bar{A}_{jl})x/2\right\|_{(P_{ij}^{-1} - \gamma^{-2} I)^{-1}}^2, \tag{6}$$

[1]The distinction between the optimal and the sub-optimal minimax adaptive control policies will be made clear at appropriate places.

where $\bar{A}_{il} = A_i - B_i K_l$ denotes the closed loop system matrix for $x \in \mathbb{R}^n$ with $i, j, l \in \{1, \dots, \mathcal{F}\}$. Then, the bound $J_{\bar{\pi}}(x_0, \gamma) \leq \max_{i,j} \|x_0\|^2_{P_{ij}}$ is valid for the minimax adaptive control policy $\bar{\pi}$ defined by

$$u_k = -K_{l_k} x_k, \quad \textit{where}, \tag{7a}$$

$$l_k := \operatorname*{argmin}_{i \in \{1,\dots,\mathcal{F}\}} \underbrace{\sum_{\tau=0}^{k-1} \|x_{\tau+1} - A_i x_\tau - B_i u_\tau\|^2}_{:=\alpha_i}. \tag{7b}$$

The controller defined in (7a), and denoted by $\bar{\pi}$ in the reminder, is sub-optimal compared to $\pi^\dagger$ (4), i.e. the associated ℓ_2 gain is $\bar{\gamma} > \gamma^\dagger$. Further, the cost $J_{\bar{\pi}}(x_0, \gamma)$ is finite as long as $\gamma > \bar{\gamma}$. The control input (7a) is nonlinear as it depends on all the past history, an approach based on least squares estimation from [5].

B. Known Dynamics Case: Standard $\mathcal{H}_\infty$ Control

When the system matrices A, B are known, problem (3) reduces to the standard $\mathcal{H}_\infty$ control. That is, a control input $u = Kx, K \in \mathbb{R}^{m \times n}$ is sought such that it minimizes the $\mathcal{H}_\infty$ norm of the closed loop system from d to ζ

$$T_{d \to \zeta}[K](z) := \begin{bmatrix} I \\ K \end{bmatrix} (zI - A - BK)^{-1}. \tag{8}$$

where $T_{d \to \zeta}$ is related to the cost function in (3) by appropriate choice of matrices Q, R. Using this observation, we define for every system model $M_i := (A_i, B_i) \in \mathcal{M}$, the associated $\mathcal{H}_\infty$ control policy $\pi_i^\star \in \Pi$, which can be found by solving the coupled Riccati equations below [20]

$$\mathbf{M}_i = Q + A_i^\top \mathbf{M}_i \Lambda_i^{-1} A_i, \quad \mathbf{M}_i \prec (\gamma_i^\star)^2 I, \tag{9}$$

$$\Lambda_i = I + \left(B_i R^{-1} B_i^\top - (\gamma_i^\star)^{-2} I \right) \mathbf{M}_i. \tag{10}$$

The dynamic game has an unique saddle point solution

$$u_k^{\pi_i^\star} = \pi_i^\star(x_k) = -K_i^\star x_k, \quad \text{and} \tag{11}$$

$$w_k^{\psi_i^\star} = \psi_i^\star(x_k) = L_i^\star x_k, \tag{12}$$

where $K_i^\star = R^{-1} B_i^\top \mathbf{M}_i \Lambda_i^{-1} A_i$ and $L_i^\star = (\gamma_i^\star)^{-2} \mathbf{M}_i \Lambda_i^{-1} A_i$. Here, $\psi_i^\star$ denotes the worst case adversarial disturbance policy and it is, like $\pi_i^\star$, a linear function of x_k. The quantity

$$\gamma_i^\star := \sqrt{\sup_{w^{\psi_i^\star} \neq 0} \frac{\sum_{k=0}^\infty c\left(x_k^{\pi_i^\star}, u_k^{\pi_i^\star}, Q, R\right)}{\sum_{k=0}^\infty \left\| w_k^{\psi_i^\star} \right\|_2^2}} \tag{13}$$

denotes the corresponding worst-case ℓ_2 gain from the disturbance to the regulated output for the model $M_i, i \in \mathcal{M}$.

III. Regret of Minimax Adaptive Control

Regret analysis compares the performance of an online algorithm that takes decisions in the presence of uncertainty with respect to a clairvoyant policy with hindsight knowledge. For this reason, it is used here in order to better understand the performance achieved when controlling the system (1) using minimax adaptive control algorithm.

Definition 1. *Regret of an online control algorithm $\mathcal{A}$ operating in the presence of uncertainty is defined as the additional cost incurred by the algorithm $\mathcal{A}$ in comparison to an optimal controller in hindsight that operates by knowing the uncertainty.*

We choose here the $\mathcal{H}_\infty$ controller associated with the true system as the optimal policy in hindsight. Note that $\forall i \in 1, \dots, \mathcal{F}$, $J_{\pi_i^\star}(x_0, \gamma) < J^\dagger(x_0, \gamma)$ as the minimax adaptive control policy $\pi^\dagger$ can never do better than the $\mathcal{H}_\infty$ policy $\pi_i^\star$ of the corresponding true system. It is possible to use a different policy other than the $\mathcal{H}_\infty$ policy for the comparison. One could compare against a control policy that solves the linear quadratic problem with known disturbance but the true $(A, B) \in \mathcal{M}$ being unknown. However, to the best of our knowledge, there is no *causal* solution for the optimal control policy to that problem. In principle, the optimal policy in hindsight should know apriori about any of the QIs that minimax does not know and also have a closed form causal solution.

The study of the minimax adaptive control problem through online learning is divided in three steps: investigation of adversarial disturbance strategies that can lead to performance deterioration of the policy $\bar{\pi}$; definition of suitable notions of regret for this problem; investigation of the regret properties of the policy $\bar{\pi}$. We do not advocate the regret as a metric to measure the robustness of a control policy. Rather, we suggest to use the regret as a tool to identify areas of improvement of an online control policy by comparing it against multiple optimal policies in hindsight. Regret analysis could also give insights for the online control design to foresee and counteract against several possible strategies of adversaries trying to worsen its performance. One such possible strategy of an adversary with respect to the policy $\bar{\pi}$ is illustrated below.

A. Adversarial disturbance strategies for minimax control

The key adaptive mechanism of policy $\bar{\pi}$ in (7) can be interpreted as an implicit identification of the underlying plant. It is then natural to ask whether this is provably able to eventually converge to the correct estimate for the system. The following theorem gives a negative answer by constructing an adversarial disturbance strategy preventing the controller from optimally controlling the true system.

Theorem 1. *Given a compact set of models $\mathcal{M}$ with $|\mathcal{M}| = \mathcal{F}$ including the true model of the system (1), consider the policy $\bar{\pi}$ given by (7). Let $j \in \{1, \dots, \mathcal{F}\}$ denote the index of the true model unknown to the policy $\bar{\pi}$. Then, $\forall k \in \mathbb{N}$, $\exists \theta_{f,k} \in \mathbb{R}$, $f = 1, \dots, \mathcal{F}$ such that the disturbance given by*

$$w_k = \sum_{f=1}^{\mathcal{F}} \theta_{f,k} (A_f x_k + B_f u_k), \tag{14}$$

lets $\bar{\pi}$ to determine a minimizer $l_k \neq j$ in (7b).

Proof. Recall from (7b) that when $i = j$, we simply get $\alpha_j = \sum_{\tau=0}^{k-1} \|w_\tau\|^2$. For other cases when $i \neq j$, we expand

α_i using the w_k given by (14) to get

$$\alpha_i = \sum_{\tau=0}^{k-1} \left\| v_\tau^{(i)} + \sum_{f=1, f\neq j}^{\mathcal{F}} \theta_{f,k}(A_f x_k + B_f u_k) \right\|^2, \quad (15)$$

with $v_\tau^{(i)} = (\theta_{j,k}A_j - A_i)x_\tau + (\theta_{j,k}B_j - B_i)u_\tau$. Then, the disturbance can let the controller choose $l_k = i$ as per (7b) deviating from the true value of j through the appropriate selection of the constants $\{\theta_{f,k}\}_{f=1}^{\mathcal{F}}$ such that $\alpha_i < \alpha_j$. One simple choice would be to choose $\theta_{j,k} = -1, \theta_{i,k} = 1$ and $\{\theta_{f,k}\}_{f=1, f\neq i, f\neq j}^{\mathcal{F}} = 0$ at time k such that $\alpha_i = 0$ in (15). Such a disturbance strategy would let the controller choose $l_k = i$ rather than j. Note that the adversary has the freedom to make $\alpha_i = 0$ for its own choice of $i \in \{1, \ldots, \mathcal{F}\}, i \neq j$ at any time step k using $\{\theta_{f,k}\}_{f=1}^{\mathcal{F}}$. □

Remarks: Disturbances with smaller magnitudes maximise the cost given in (3). Though, the disturbance given by (14) can make the learning hard for the controller, it need not have a smaller magnitude for a given $\gamma > 0$ and $\{\theta_{f,k}\}_{f=1}^{\mathcal{F}}$, and hence it may *not* lead to the worse cost. Further, for certain range of $\{\theta_{f,k}\}_{f=1}^{\mathcal{F}}$, the associated closed loop system may turn out to be unstable. The negative result formulated in Theorem 1 justifies further analysis on the sub-optimality faced by the minimax adaptive controller, which is studied in the next sections through the concept of regret.

B. Regret Definitions

Note that each model $(A_i, B_i) \in \mathcal{M}$ suffers different regret when compared against the optimal $\mathcal{H}_\infty$ controller in hindsight. Hence, we quantify the regret of each model in the set $\mathcal{M}$ in the following definition.

Definition 2. *Given a model $M_i := (A_i, B_i) \in \mathcal{M}, i \in \{1, \ldots, \mathcal{F}\}$, we define the model-based regret of the minimax adaptive control policy $\pi^\dagger \in \Pi$ with respect to the optimal control policy $\pi_i^\star$ for $\gamma \geq \gamma^\dagger > \gamma_i^\star$ and time $T \in \mathbb{N}$ as*

$$\mathcal{R}(\pi^\dagger, \pi_i^\star, T) = \sup_{w\in\ell_{2e}} \sum_{k=0}^{T} d_k\left(\pi^\dagger, \pi_i^\star\right), \quad \textit{where}, \quad (16)$$

$$d_k\left(\pi^\dagger, \pi_i^\star\right) := \left\| x_k^{\pi^\dagger} - x_k^{\pi_i^\star} \right\|_Q^2 + \left\| u_k^{\pi^\dagger} - u_k^{\pi_i^\star} \right\|_R^2.$$

Any disturbance that is not in the ℓ_{2e} space will result in diverging states. We note that the choice of regret metric is not conventional, as the standard approach would be to define it as difference of costs, that is,

$$\bar{\mathcal{R}}(\pi^\dagger, \pi_i^\star, T) = \sup_{w\in\ell_{2e}} \sum_{k=0}^{T} \bar{d}_k\left(\pi^\dagger, \pi_i^\star\right), \quad (17)$$

$$\bar{d}_k\left(\pi^\dagger, \pi_i^\star\right) := c\left(x_k^{\pi^\dagger}, u_k^{\pi^\dagger}, Q, R\right) - c\left(x_k^{\pi_i^\star}, u_k^{\pi_i^\star}, Q, R\right)$$

While (17) captures how close the systems controlled by the minimax adaptive controller and the optimal $\mathcal{H}_\infty$ controller in hindsight are in terms of the performance, it does not provide information on how close the two state and inputs trajectories are. Further, (17) cannot account for the direction of the control input being applied to the system. For these reasons, we propose to use (16) as the definition of model-based regret in this work. Note that the model-based regret in (16) is a function of the chosen level of robustness γ because this parameter affects the two policies $\pi^\dagger$ and $\pi_i^\star$ (this dependence is omitted for the sake of clarity). The regret is defined for $\gamma \geq \gamma^\dagger > \gamma_i^\star$ to ensure that a fair comparison is made between the resulting trajectories from controllers that share the same level of disturbance attenuation capabilities. To compute (16), we need to characterize the trajectories of the system $x_k^{\pi^\dagger}$ and $x_k^{\pi_i^\star}$ given by (1) under the same sequence of adversarial disturbance inputs affecting the system using the control policies $\pi^\dagger$ and $\pi_i^\star$ respectively. This naturally leads us to investigate what would be the worst-case model-based regret corresponding to any arbitrary model $M_i \in \mathcal{M}$, i.e., the total regret.

Definition 3. *The total regret of the minimax adaptive controller is defined as*

$$\mathfrak{R}(\pi^\dagger, T) := \max_{i\in\{1,\ldots,\mathcal{F}\}} \mathcal{R}(\pi^\dagger, \pi_i^\star, T). \quad (18)$$

While comparing policies, it is important to compare their disturbance attenuation levels too. Sub-optimality gap indicates a room for improvement in terms of the robustness. Since, minimax adaptive controller can never match the $\mathcal{H}_\infty$ controller, the difference in their disturbance attenuation level is referred as the *model-based sub-optimality gap*.

Definition 4. *Given a model $(A_i, B_i) \in \mathcal{M}, i \in \{1, \ldots, \mathcal{F}\}$, the model-based sub-optimality gap of the minimax adaptive control policy $\pi^\dagger$ is defined as*

$$\mathcal{O}(\pi^\dagger, \pi_i^\star) := \gamma^\dagger - \gamma_i^\star. \quad (19)$$

The model-based sub-optimality gap satisfies by definition $\mathcal{O}(\pi^\dagger, \pi_i^\star) \geq 0$ and characterizes how the lack of knowledge about the QIs results in a worst disturbance attentuation level of the minimax adaptive controller (or reduction in robust performance). In a similar spirit to the definition of total regret, we define below the minimal and the maximal sub-optimality gaps, which are by definition both non-negative.

Definition 5. *The minimal sub-optimality gap and the maximal sub-optimality gap of the minimax adaptive control policy $\pi^\dagger$ are respectively defined as*

$$\underline{\mathcal{O}}(\pi^\dagger) := \gamma^\dagger - \max_{i\in\{1,\ldots,\mathcal{F}\}} \gamma_i^\star, \quad \textit{and} \quad (20)$$

$$\overline{\mathcal{O}}(\pi^\dagger) := \gamma^\dagger - \min_{i\in\{1,\ldots,\mathcal{F}\}} \gamma_i^\star. \quad (21)$$

C. Study of Minimax Adaptive Control Regret

The following theorem establishes the asymptotic behaviour of the total regret associated with the minimax adaptive control policy $\mathfrak{R}(\bar{\pi}^\dagger, T)$.

Theorem 2. *Consider the uncertain linear dynamical system given by (1) with the uncertainty described by $\mathcal{M}$. If the disturbance signal is in ℓ_2 space, then the associated total*

regret (18) is sub-linear, i.e.

$$\lim_{T\to\infty} \frac{\mathfrak{R}(\bar{\pi}^\dagger, T)}{T} = 0. \tag{22}$$

Proof. Recall that both minimax adaptive control policy $\bar{\pi}$ given by (7a) and $\mathcal{H}_\infty$ control policy given by (11) are stabilising (with exponential decay of states and controls) for any adversarial disturbance in ℓ_2 space. That is, given any disturbance signal w_k in ℓ_2 space for plant model $i \in \{1, \dots, \mathcal{F}\}$, we have

$$\lim_{k\to\infty} \left\| x_k^{\bar{\pi}^\dagger} \right\|_Q^2 = 0, \quad \text{and} \quad \lim_{k\to\infty} \left\| x_k^{\pi_i^\star} \right\|_Q^2 = 0. \tag{23}$$

Then, this means that $\lim_{k\to\infty} \left\| x_k^{\bar{\pi}^\dagger} - x_k^{\pi_i^\star} \right\|_Q^2 = 0$. Since at any time $k \in \mathbb{N}$ both minimax adaptive control input $u_k^{\bar{\pi}^\dagger}$ given by (7a) and the $\mathcal{H}_\infty$ control input $u_k^{\pi_i^\star}$ given by (11) are functions of the states $x_k^{\bar{\pi}^\dagger}$ and $x_k^{\pi_i^\star}$ respectively that decay to zero asymptotically, we infer that

$$\lim_{k\to\infty} \left\| u_k^{\bar{\pi}^\dagger} \right\|_R^2 = 0, \quad \text{and} \quad \lim_{k\to\infty} \left\| u_k^{\pi_i^\star} \right\|_R^2 = 0. \tag{24}$$

Then, this means that $\lim_{k\to\infty} \left\| u_k^{\bar{\pi}^\dagger} - u_k^{\pi_i^\star} \right\|_R^2 = 0$. Therefore, the difference term decays as well to zero meaning that $\lim_{k\to\infty} d_k(\bar{\pi}^\dagger, \pi_i^\star) = 0$. Hence, the result follows. □

An insight gathered from the proof is that stability of the policy implies certain regret properties. This has connections with recent findings in [21] which studied the relationship between stability and regret for disturbances in ℓ_∞ space.

IV. Numerical Simulation

In this section, we exemplify our analysis using a linear dynamical system with a model uncertainty consisting of four different linear models.

A. Problem Setup

We consider the following numerical example of a linear dynamical system with four possible models. The state and control penalty matrices were $Q = I_3, R = 1$ and $T = 50$. We simulated the system using the minimax adaptive controller and the $\mathcal{H}_\infty$ controller available in hindsight separately when the pair (A_2, B_2) (corresponds to $j = 2$ as per Theorem 1) was the true model.

$$\begin{aligned}
A_1 &= \begin{bmatrix} 1.908 & 0.853 & 0.633 \\ 0.853 & 0.142 & 0.645 \\ 0.633 & 0.645 & 0.018 \end{bmatrix}, A_2 = \begin{bmatrix} 0.060 & 0.335 & 0.809 \\ 0.335 & 0.017 & 1.507 \\ 0.809 & 1.507 & 0.873 \end{bmatrix}, \\
A_3 &= \begin{bmatrix} 0.182 & 1.435 & 0.730 \\ 1.435 & 1.714 & 1.183 \\ 0.730 & 1.183 & 0.452 \end{bmatrix}, A_4 = \begin{bmatrix} 0.922 & 0.800 & 1.350 \\ 0.800 & 1.431 & 1.462 \\ 1.350 & 1.462 & 0.786 \end{bmatrix}, \\
B_1 &= \begin{bmatrix} 1.830 & 1.285 & 0.002 \end{bmatrix}^\top, B_2 = \begin{bmatrix} 0.873 & 0.098 & 0.099 \end{bmatrix}^\top, \\
B_3 &= \begin{bmatrix} 1.073 & 1.524 & 0.695 \end{bmatrix}^\top, B_4 = \begin{bmatrix} 0.358 & 1.266 & 1.248 \end{bmatrix}^\top.
\end{aligned} \tag{25}$$

Three different disturbances constructions were used

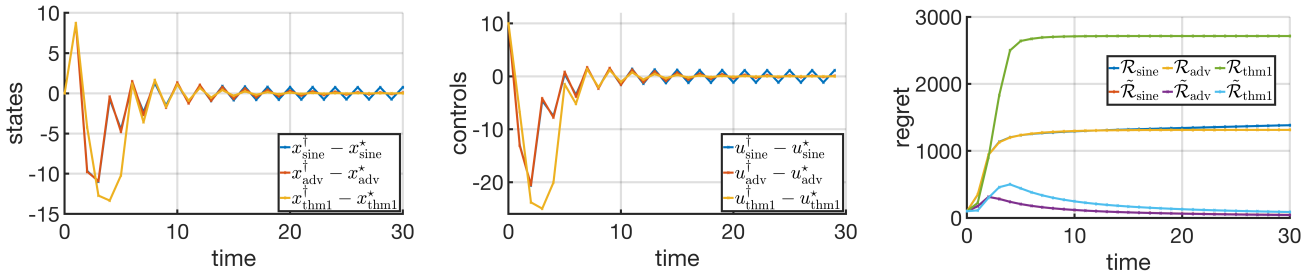
1) worst case disturbance signal obtained from the dynamic game based $\mathcal{H}_\infty$ approach given by (12).
2) sinusoidal disturbance with unit amplitude and its frequency being selected as the frequency where the $\mathcal{H}_\infty$ norm of $T_{d\to\zeta}[K](z)$ given by (8) was maximum.
3) the disturbance given by (14) used in the proof of Theorem 1 with $i = 3$ and tuned so that the controller always choose the optimal controller for (A_3, B_3).

The gains for the sub-optimal minimax adaptive controller were calculated using the method from [8] which is an improved version of Theorem 3 in [7] and we used the Yalmip toolbox with the MOSEK solver to solve the associated convex optimization problem with linear matrix inequality constraints. The code corresponding to the figures given in the paper is made publicly available at https://github.com/venkatramanrenganathan/minimaxadaptivecontrolregret.

B. Results & Discussions

The system dynamics with $\{(A_i, B_i)\}_{i=1}^4$ given by (25) was solved for the minimax adaptive control policy $\bar{\pi}$ to get $\bar{\gamma}^\dagger = 31.0086$. Then, the corresponding $\mathcal{H}_\infty$ controller was obtained using $\gamma = \bar{\gamma}^\dagger$. The optimal ℓ_2 gains, namely $\{\gamma_i^\star\}_{i=1}^4$ corresponding to the optimal $\mathcal{H}_\infty$ controller for the plants $\{(A_i, B_i)\}_{i=1}^4$ were $1.266, 4.544, 2.913, 2.298$ respectively. The model-based sub-optimality gaps were found using (19) as $\mathcal{O}(\bar{\pi}^\dagger, \pi_1^\star) = 29.7426$, $\mathcal{O}(\bar{\pi}^\dagger, \pi_2^\star) = 26.4646$, $\mathcal{O}(\bar{\pi}^\dagger, \pi_3^\star) = 28.0956$, and $\mathcal{O}(\bar{\pi}^\dagger, \pi_4^\star) = 28.7106$. Further, the minimal and maximal sub-optimality gaps were $\underline{\mathcal{O}}(\bar{\pi}^\dagger) = 26.4646$ and $\overline{\mathcal{O}}(\bar{\pi}^\dagger) = 29.7426$ respectively.

The results of simulating system with minimax adaptive controller and $\mathcal{H}_\infty$ controller with three different disturbance strategies are shown in Figure 1. The sub-figures 1(a), and 1(b) depict the difference of states and inputs respectively from the minimax adaptive controller and the $\mathcal{H}_\infty$ controller and precisely these are the main contributing factors of regret as per (18). The regret quantities $\mathcal{R}(\bar{\pi}^\dagger, \pi_2^\star, T)$ and $\frac{\mathcal{R}(\bar{\pi}^\dagger, \pi_2^\star, T)}{T}$ are abbreviated as $\mathcal{R}$ and $\tilde{\mathcal{R}}$ respectively are plotted in the sub-figure 1(c). When adversarial disturbance constructed from the worst case disturbance policy given by (12) was used, the associated regret was bounded as seen in sub-figure 1(c). Moreover, the shown results fulfils (22) as both the control policies were stabilising and the disturbance was regulated to zero as it was a linear function of the system states (as per (12)) which decayed in exponential time. When a sinusoidal disturbance with unit amplitude was employed with its frequency being selected as the frequency where the $\mathcal{H}_\infty$ norm of $T_{d\to\zeta}[K](z)$ given by (8) was maximum (for (A_2, B_2) this happens at π rad/s), the regret was not bounded anymore as the sinusoidal disturbance does not belong to the ℓ_2 space. However, the ratio namely $\mathcal{R}(\bar{\pi}^\dagger, \pi_2^\star, T)/T$ went to zero asymptotically as the terms contributing to the regret namely the differences of states and controls remained small and did grow slower than linearly. When the disturbance given by (14) was used, it ensured that the l_k value chosen by the minimax adaptive control policy $\bar{\pi}^\dagger$ according to (7b) was never equal to $j = 2, \forall k \in [0, T]$. Even though the ratio $\mathcal{R}(\bar{\pi}^\dagger, \pi_2^\star, T)/T$ went to zero asymptotically as disturbance was still a function of exponentially



(a) Minimax & $\mathcal{H}_\infty$ States: $x_k^{\bar{\pi}^\dagger}$ (b) Minimax & $\mathcal{H}_\infty$ Controls: $x_k^{\pi_2^\star}$ (c) Regret scaling vs time

Fig. 1: Simulation results with states and controls from the minimax adaptive controller and the $\mathcal{H}_\infty$ controller are plotted here along with the corresponding regret scaling over time. Only the first state of the system (25) is plotted for the demonstration purpose. Note that the quantities $\mathcal{R}(\bar{\pi}^\dagger, \pi_2^\star, T)$ and $\frac{\mathcal{R}(\bar{\pi}^\dagger, \pi_2^\star, T)}{T}$ are abbreviated as $\mathcal{R}$ and $\tilde{\mathcal{R}}$ respectively. The text in the subscript of quantities in all sub-plots denotes the type of disturbance being used.

decaying states, it can be observed that the disturbance signal had a larger magnitude than the one from (12). This shows that the disturbance that hardens the learning process need not necessarily worsen the performance as measured by (3) because it results in a decrease of the total cost.

V. Conclusion

An online learning-inspired analysis for a recently proposed solution for a class of minimax adaptive control problems has been presented. Model-based regret and total regret for the minimax adaptive control policy were defined by comparing the state and input trajectories against that of the optimal $\mathcal{H}_\infty$ controller in hindsight (i.e. having knowledge of the true system dynamics). One of the highlights of the analysis is that the total regret is sub-linear for exogenous disturbances in the ℓ_2 space, confirming links between system theoretic properties and regret for control systems. Future research will seek to characterize transient properties of regret, and their connections with the exploration-exploitation trade-off inherently captured by the minimax adaptive control algorithms. Starting from the definitions of regret proposed here, designing regret-optimal adaptive controllers that lower the conservatism of minimax solutions is also an important research question lying ahead.

Acknowledgment

The authors are thankful to Daniel Cederberg at Linköping University for providing us with the code and to Olle Kjellqvist at Lund University for his insightful comments.

References

[1] E. Hazan and K. Singh, "Introduction to online nonstochastic control," in *arXiv 2211.09619*, 2022.
[2] K. J. Astrom and B. Wittenmark, "On self tuning regulators," *Automatica*, vol. 9, no. 2, pp. 185–199, 1973.
[3] M. Benosman, "Model-based vs data-driven adaptive control: an overview," *International Journal of Adaptive Control and Signal Processing*, vol. 32, no. 5, pp. 753–776, 2018.
[4] D. Salmon, "Minimax controller design," *IEEE Transactions on Automatic Control*, vol. 13, no. 4, pp. 369–376, 1968.
[5] G. Didinsky and T. Basar, "Minimax adaptive control of uncertain plants," in *Proceedings of 1994 33rd IEEE Conference on Decision and Control*, vol. 3. IEEE, 1994, pp. 2839–2844.
[6] A. Rantzer, "Minimax adaptive control for state matrix with unknown sign," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 58–62, 2020.
[7] ——, "Minimax adaptive control for a finite set of linear systems," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 893–904.
[8] D. Cederberg, A. Hansson, and A. Rantzer, "Synthesis of minimax adaptive controller for a finite set of linear systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 1380–1384.
[9] O. Kjellqvist and A. Rantzer, "Minimax adaptive estimation for finite sets of linear systems," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 260–265.
[10] D. Chichka and J. L. Speyer, "An adaptive controller based on disturbance attenuation," *IEEE transactions on automatic control*, vol. 40, no. 7, pp. 1220–1233, 1995.
[11] J. Yoneyama, J. L. Speyer, and C. H. Dillon, "Robust adaptive control for linear systems with unknown parameters," *Automatica*, vol. 33, no. 10, pp. 1909–1916, 1997.
[12] M. French and S. Trenn, "l_p gain bounds for switched adaptive controllers," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 2865–2870.
[13] G. Vinnicombe, "Examples and counterexamples in finite l2-gain adaptive control," in *Leuven: Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, 2004.
[14] Y. Jedra and A. Proutiere, "Minimal expected regret in linear quadratic control," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 10 234–10 321.
[15] N. M. Boffi, S. Tu, and J.-J. E. Slotine, "Regret bounds for adaptive nonlinear control," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 471–483.
[16] G. Goel and B. Hassibi, "Regret-optimal control in dynamic environments," *arXiv preprint arXiv:2010.10473*, 2020.
[17] A. Karapetyan, A. Iannelli, and J. Lygeros, "On the regret of H_∞ control," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 6181–6186.
[18] E. Hazan, S. Kakade, and K. Singh, "The nonstochastic control problem," in *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, vol. 117. PMLR, 08 Feb–11 Feb 2020, pp. 408–421.
[19] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, "Online control with adversarial disturbances," in *International Conference on Machine Learning*. PMLR, 2019, pp. 111–119.
[20] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.
[21] A. Karapetyan, A. Tsiamis, E. C. Balta, A. Iannelli, and J. Lygeros, "Implications of regret on stability of linear dynamical systems," *IFAC-PapersOnLine*, 2023, 22nd IFAC World Congress.