Implications of Regret on Stability of Linear Dynamical Systems [★]

Aren Karapetyan [*] **Anastasios Tsiamis** [*] **Efe C. Balta** [*]
Andrea Iannelli [**] **John Lygeros** [*]

[*] *Automatic Control Laboratory, ETH Zürich, Zürich 8092,
Switzerland (e-mails: {akarapetyan, atsiamis, ebalta,
jlygeros}@control.ee.ethz.ch)*
[**] *University of Stuttgart, Institute for Systems Theory and Automatic
Control, Stuttgart 70569, Germany
(e-mail: andrea.iannelli@ist.uni-stuttgart.de)*

Abstract: The setting of an agent making decisions under uncertainty and under dynamic constraints is common for the fields of optimal control, reinforcement learning, and recently also for online learning. In the online learning setting, the quality of an agent's decision is often quantified by the concept of regret, comparing the performance of the chosen decisions to the best possible ones in hindsight. While regret is a useful performance measure, when dynamical systems are concerned, it is important to also assess the stability of the closed-loop system for a chosen policy. In this work, we show that for linear state feedback policies and linear systems subject to adversarial disturbances, linear regret implies asymptotic stability in both time-varying and time-invariant settings. Conversely, we also show that bounded input bounded state stability and summability of the state transition matrices imply linear regret.

Keywords: Regret, Linear Control Systems, Time-Varying Systems, Optimal Control

1. INTRODUCTION

A number of real-world problems can be cast into the framework of agents making optimal decisions under uncertainty and/or adversarial disturbances. In this setting, the agent has an associated dynamical system and at each timestep suffers an *a priori* unknown cost that depends on its state and input. In the case of perfect knowledge of the dynamics and the future costs, this can be turned into an optimal control problem and solved with one of the plethora of available methods. As is often the case, however, the dynamics, disturbances, and/or future costs are either entirely unknown or only partially known. This is the setting, for example, in reinforcement learning and approximate dynamic programming (Bertsekas, 2015a).

As agents make decisions "on-the-go", there is a need to quantify and compare the performance of various algorithms. Considering a closed-loop system with a given, possibly time-varying policy, one can study its asymptotic stability as an asymptotic metric. Another approach is to look into the problem through the lens of online optimization. An important metric in the literature of the latter is the notion of regret. Given a policy μ, its regret $\mathcal{R}_T$, is defined as the difference between its accumulated cost over some time horizon T and that of some benchmark policy π. It is often desirable to have sublinear growth of $\mathcal{R}_T$

with respect to T, which will achieve average convergence to the benchmark in the limit.

A special case of the optimal control problem, the linear quadratic regulator (LQR) has been extensively studied in this context. The results in (Simchowitz and Foster, 2020) show that the certainty equivalence approach can synthesize stable linear state feedback controllers as long as the model estimate errors are small enough. In the same spirit of certainty equivalence, the algorithms proposed in (Cohen et al., 2018; Jedra and Proutiere, 2022) yield sublinear regret bounds by sequentially solving Riccati equations. The LQR problem with adversarial disturbances has been studied in (Yu et al., 2020; Zhang et al., 2021) in the presence of a prediction window, in (Karapetyan et al., 2022) to quantify the regret of robustness in the sense of $\mathcal{H}_\infty$ control, and in a regret-optimal setting in (Sabag et al., 2021) for unconstrained and in (Didier et al., 2022; Martin et al., 2022) for constrained cases. The setting with general convex and differentiable cost functions with adversarial disturbances has been studied for linear time-invariant (LTI) (Agarwal et al., 2019) and for linear time-varying (LTV) (Gradu et al., 2020) systems; algorithms that achieve sublinear regret bounds have been proposed for both classes of systems.

To study and analyze algorithms in both the control theoretic and online learning contexts, one needs to understand the relationship between regret and stability. While most online learning-inspired works seek sublinear regret in pursuit of suboptimality guarantees, it is unclear whether good performance in this aspect also implies stability. Conversely, it is not known what the absence of such

guarantees means even for linear dynamical systems. In a recent work (Nonhoff and Müller, 2022), the authors study the interconnection between bounded regret and closed-loop stability for non-linear systems in the absence of noise. In the current work, we consider linear systems subject to process noise and study the relationship between linear regret and stability. Our goal is to develop sufficient conditions under which regret guarantees of an algorithm imply stability of the closed-loop system and vice versa.

In particular, we consider generic, time-varying costs, LTV systems subject to adversarial disturbances with LTV state feedback policies, and also LTI systems as a special case. We study the connection between the regret of such policies and their closed-loop stability in the sense of bounded input-bounded state (BIBS) stability with respect to disturbances and asymptotic stability in the absence of disturbances. Under suitable assumptions, we show the following:

(a) For LTV systems and LTV state feedback policies linear regret implies asymptotic stability of the closed-loop system. Conversely, BIBS stability and absolute summability of the state transition matrices imply linear regret.

(b) For LTI systems and LTI state feedback policies linear regret is attained if and only if the closed-loop system is asymptotically stable.

An essential requirement for the results to hold is the "observability" of the state with respect to costs. We provide a counterexample of a notable case where the asymptotic stability implication of linear regret fails to hold when this assumption is violated. A numerical example to showcase the result for the LTI case is provided.

Notation: For a square matrix A the spectral radius and the spectral norm are denoted by $\rho(A)$, and $\|A\|$, respectively. For a vector $w \in \mathbb{R}^n$, $\|w\|$ denotes its Euclidean norm. $\mathbb{R}_+$ is the set of positive real numbers and $\mathbb{N}$ that of non-negative integers.

2. PROBLEM SETUP

2.1 Preliminaries

Consider discrete-time LTV systems of the form
$$x_{t+1} = A_t x_t + B_t u_t + w_t, \qquad (1)$$
evolving over non-negative times $t \in \mathbb{N}$, where $A_t \in \mathbb{R}^{n \times n}$, $B_t \in \mathbb{R}^{n \times m}$ are real matrices, and $x_t, w_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$ are, respectively, the state, disturbance and input signals at timestep t. The disturbances are treated as adversarial and are assumed to be norm bounded, that is, there exists $W \in \mathbb{R}_+$ such that $\|w_t\| \leq W$ for all t. The objective of the optimal control problem is to minimize the cost function

$$J_T(x_0, \boldsymbol{u}; \boldsymbol{w}) = \sum_{t=0}^{T} c_t(x_t, u_t), \qquad (2)$$

where $c_t(x_t, u_t) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the stage cost function at t and $\boldsymbol{u} := [u_0^\top \ldots u_{T-1}^\top]^\top$, $\boldsymbol{w} := [w_0^\top \ldots w_{T-1}^\top]^\top$.

Benchmark Controller: If the system matrices (A_t, B_t) in (1), the stage costs $\{c_t\}_{t=0}^T$ and the disturbance signal $\boldsymbol{w}$ are known at time 0, then the control input signal $\boldsymbol{u}^\star$

that minimises (2) can be obtained by solving the following optimization problem

$$\arg\min_{\boldsymbol{u}} \quad J_T(x_0, \boldsymbol{u}; \boldsymbol{w})$$
$$\text{subject to} \quad (1) \quad \forall\, 0 \leq t < T.$$

Note that $\boldsymbol{u}^\star$ is referred to as the optimal input signal in hindsight, obtained with full knowledge of the uncertainty.

Linear Policies: In this work, we consider linear time-varying policies of the form [1] $\mu_t(x_t) = -K_t x_t$ for some $K_t \in \mathbb{R}^{m \times n}$ for all $t \in \mathbb{N}$, where the matrices K_t are independent of the state and input history. This covers the class of offline linear policies discussed in (Goel and Hassibi, 2022). Moreover, we note that the considered policies also cover the case where the time-varying parameters of the stage costs are revealed sequentially, thus generating K_t online. For example, consider the LQR problem with time-varying cost matrices Q_t, R_t, that depend only on time t and are revealed in the form of predictions, studied in (Zhang et al., 2021). The future cost functions, disturbances, as well as the system matrices may be unknown to the policy.

For a given noise signal realisation $\boldsymbol{w}$ and cost functions $\{c_t\}_{t=0}^T$, we define the regret of a given policy μ to be
$$\mathcal{R}_T(\mu; \boldsymbol{w}) := J_T(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) - J_T(x_0, \boldsymbol{u}^\star; \boldsymbol{w}),$$
where $\boldsymbol{u}^\mu \in \mathbb{R}^{Tm}$ is the input signal generated by μ.

Whereas the regret provides intuition on the performance of the given policy with respect to the best possible cost, its implication for the stability of the associated closed-loop system is unclear. Though several works, e.g. in the stochastic setting (Dean et al., 2018; Cohen et al., 2018), also provide stability guarantees for the online policies, there is no clear relationship between the order of regret and the notion of stability; the results below attempt to address this shortcoming.

We start by defining linear regret to show the conditions under which it implies stability and vice versa.

Definition 2.1. *A policy μ is said to have linear regret if for given $W, X \in \mathbb{R}_+$, and class of stage costs [2], there exist $C_w, C_0 \in \mathbb{R}_+$ such that for all $x_0 \in \mathbb{R}^n$ with $\|x_0\| \leq X$, and for all admissible sequences $\{c_t\}_{t=0}^T$ and $\{w_t\}_{t=0}^{T-1}$*
$$\mathcal{R}_T(\mu; \boldsymbol{w}) \leq C_0 + C_w T, \quad \forall T > 0. \qquad (3)$$

In the rest of this section, we review some notions of stability and provide the assumptions required to show the points (a) and (b).

The LTV system in (1) takes the following closed-loop form for considered policies $\mu_t(x_t) = -K_t x_t$
$$x_{t+1} = F_t x_t + w_t, \qquad (4)$$
where $F_t = A_t - B_t K_t$. We refer to (4) as the closed-loop system for the considered policy; for questions of asymptotic or exponential stability, we consider the unforced version of (4) with $w_t = 0$ for all t (Ludyk, 2013).

Let $\Phi(t, t_0) \in \mathbb{R}^{n \times n}$ denote the state transition matrix of the free system $x_{t+1} = F_t x_t$ at time t, starting at some

---

[1] When the context is clear, we drop the subscript t and/or the argument when referring to the linear policy $\mu_t(x_t)$.

[2] For example, the one defined later in Assumption 2.7.

initial time t_0. In the following, we provide the definitions of asymptotic, exponential, and BIBS stability, specialized for LTV systems (Callier and Desoer, 2012; Ludyk, 2013).

Definition 2.2. *(Asymptotic Stability). The LTV system (4) is asymptotically stable if and only if for all $x_0 \in \mathbb{R}^n$ and for all $t_0 \in \mathbb{N}$*

(1) There exists $b \in \mathbb{R}_+$, such that $\|\Phi(t, t_0)x_0\| \leq b$ for all $t \geq t_0$,

(2) $\lim_{t \to \infty} \Phi(t, t_0)x_0 = \mathbf{0}$.

Definition 2.3. *(BIBS stability). The LTV system (4) is BIBS stable if and only if*

$$\sup_{t>0}\left[\sum_{k=1}^{t}\|\Phi(t,k)\|\right] < \infty.$$

Definition 2.4. *(Exponential Stability). The LTV system (4) is exponentially stable if and only if there exist $\delta \in [0, 1)$ and $d \in \mathbb{R}_+$, such that for all $t_0 \in \mathbb{N}$*

$$\|\Phi(t, t_0)\| \leq d\delta^{t-t_0} \quad \forall t \geq t_0.$$

We make the following assumptions on the dynamics.
Assumption 2.5. *(System Dynamics).*

i. There exists a policy $\overline{\mu}_t(x_t) = -\overline{K}_t x_t$, such that the closed-loop system (4) with $F_t = \overline{F}_t := A_t - B_t \overline{K}_t$ is exponentially stable.

ii. The state transition matrix for (4) satisfies either of the following,
(a) $\lim_{T \to \infty} \|\Phi(T, 0)\| \in \mathbb{R} \cup \{\infty\}$

(b) $0 < \liminf_{T \to \infty} \|\Phi(T, 0)\| < \limsup_{T \to \infty} \|\Phi(T, 0)\| < \infty.$

iii. The considered linear state feedback policies are such that the system matrix F_t is full rank for all $t \in \mathbb{N}$.

We note that Assumption 2.5.i. is required for performance guarantees of the benchmark policy and corresponds to the stabilizability in the LTI case. Assumption 2.5.ii. excludes chaotic LTV systems in the sense of Li-Yorke (Shi and Chen, 2009) and is always satisfied for LTI systems. It can be relaxed, but we introduce it here to keep the discussion simple. Assumption 2.5.iii. is standard in the discrete-time case to avoid deadbeat-type responses that imply a lack of uniqueness of the solutions backward in time.

Assumption 2.5.iii. leads to the following lemma (Callier and Desoer, 2012).
Lemma 2.6. *Assume that Assumption 2.5.iii. holds, then the LTV system (4) is asymptotically stable if and only if*

$$\lim_{T \to \infty} \|\Phi(T, 0)\| = 0. \tag{5}$$

We also introduce the following assumption for the costs.
Assumption 2.7. *(Stage Costs). There exist positive $\overline{M}, \underline{M} \in \mathbb{R}_+$ and $\overline{s}, \underline{s} \geq 1$, such that[3] for all $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ and $t \in \mathbb{N}$,*

$$\underline{M}\|x\|^{\underline{s}} \leq c_t(x, u) \leq \overline{M}(\|x\|^{\overline{s}} + \|u\|^{\overline{s}}).$$

The lower bound in the above assumption, which we refer to as the "observability" of the state with respect to cost, excludes cases when unstable states can be "hidden"

---
[3] We consider throughout $\underline{s} = \overline{s} = 2$ without loss of generality.

in the cost. This condition is common in the literature (Bertsekas, 2015b) and is also referred to as positive definiteness or detectability of the system with respect to stage costs (Postoyan et al., 2016).

2.2 Relation to Tracking Problems

For completeness, we also point out a connection between the stability discussion below and tracking problems. Consider the problem of tracking an unknown, bounded, time-varying reference signal $r_t \in \mathbb{R}^n$ given system dynamics (1), where the agent has access to r_t at timestep t, but not before. We can then write the state evolution of the tracking error $\tilde{x}_t := x_t - r_t$ as

$$\tilde{x}_{t+1} = A\tilde{x}_t + Bu_t + \underbrace{w_t - r_{t+1} + Ar_t}_{\nu_t},$$

where $\nu_t \in \mathbb{R}^n$ can be considered as the disturbance of the modified system; note that r_{t+1} and w_t are unknown at time t. If the assumptions above are satisfied the tracking problem can be treated as the regulation under the adversarial noise problem in Section 2.1.

3. MAIN RESULTS

In this section, we derive our main results for LTV and LTI systems. We start with the following motivating example of discounted LQR as a particular case when an unstable system attains linear regret if Assumption 2.7 is violated.
Example 3.1. *Consider the dynamics (1), with $A_t = A$, $B_t = B$ for all $t \in \mathbb{N}$, and the following cost function*

$$J_T^d(x_0, \boldsymbol{u}; \boldsymbol{w}) = x_T^\top P_\alpha x_T + \sum_{i=0}^{T-1} \alpha^i \left(x_i^\top Q x_i + u_i^\top R u_i\right), \tag{6}$$

where $\alpha \in (0, 1)$ is a discount factor, $Q \in \mathbb{R}^{n \times n}$, $R \in \mathbb{R}^{m \times m}$ are positive definite matrices, and $P_\alpha \in \mathbb{R}^{n \times n}$ satisfies the following equation

$$P_\alpha = Q + \alpha A^\top P_\alpha A - \alpha^2 A^\top P_\alpha B (R + \alpha B^\top P_\alpha B)^{-1} B^\top P_\alpha A.$$

The above is the discrete algebraic Riccati equation (DARE) for the modified system $(\sqrt{\alpha}A, B)$ with cost matrices Q and $\frac{R}{\alpha}$. The stage costs do not satisfy the condition in Assumption 2.7 as there exists no uniform lower bound. Consider now the certainty equivalent controller that minimises the cost (6) assuming $w_t = \mathbf{0} \quad \forall t \in \mathbb{N}$. The optimal policy μ_t in this setting is then a linear state feedback, given by $\mu_t = -K_\alpha x_t$ (Bertsekas, 2015a), where

$$K_\alpha = -\alpha(R + \alpha B^\top P_\alpha B)^{-1} B^\top P_\alpha A.$$

We make the hypothesis that the value function (cost-to-go) at time step t associated with this policy is given by

$$V_t(x_t) = \alpha^t \left[x_t^\top P_\alpha x_t + v_t^\top x_t + q_t\right],$$

where $v_t \in \mathbb{R}^n$, $q_t \in \mathbb{R}$. At timestep T, it is easily verified with $v_T = \mathbf{0}$, $q_T = 0$. Assuming the hypothesis is true for $t + 1$, the cost-to-go at t is

$$V_t(x_t) = \alpha^t[x_t^\top(Q + K_\alpha^\top R K_\alpha)x_t + \alpha x_t^\top F_\alpha^\top P_\alpha F_\alpha x_t$$

$$+ 2\alpha x_t^\top \left(F_\alpha^\top P_\alpha w_t + \frac{1}{2}v_{t+1}\right)$$

$$+ \alpha \left(w_t^\top P_\alpha w_t + w_t^\top v_{t+1} + q_{t+1}\right)].$$

One can verify that the induction hypothesis is verified only if P_α satisfies the DARE for the modified system, and

$$v_t = 2\alpha F_\alpha \left(P_\alpha w_t + \frac{1}{2} v_{t+1} \right)$$

$$q_t = \alpha \left(w_t^\top P_\alpha w_t + w_t^\top v_{t+1} + q_{t+1} \right).$$

The cost under this linear state feedback policy can then be attained by applying the preceding recursion repeatedly

$$J_T^d(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) = x_0^\top P_\alpha x_0 + x_0^\top v_0 + q_0,$$

where

$$v_0 = 2 \sum_{k=0}^{T-1} \alpha^{k+1} \left(F_\alpha^\top \right)^{k+1} P_\alpha w_k,$$

$$q_0 = \sum_{k=0}^{T-1} \alpha^{k+1} \left(w_k^\top P_\alpha w_k + \sum_{i=1}^{T-1} \left(\alpha^i \left(F_\alpha^\top \right)^i P_\alpha w_i \right) \right),$$

and $F_\alpha := A - BK_\alpha$. It is known that for α small enough the closed-loop system matrix F_α may in fact be unstable (Postoyan et al., 2016). Consider the set,

$$\Gamma := \{ \alpha \in (0,1) \mid \alpha \| F_\alpha \| < 1, \rho(F_\alpha) > 1 \}.$$

Then for some $\bar{\alpha} \in \Gamma$, and for $\sigma := \bar{\alpha} \| F_\alpha \|$

$$J_T^d(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \leq \|x_0\|^2 \|P_{\bar{\alpha}}\| + 2\|x_0\| \|P_{\bar{\alpha}}\| W \frac{\sigma(1 - \sigma^T)}{1 - \sigma}$$

$$+ W^2 \|P_{\bar{\alpha}}\| \frac{\bar{\alpha}(1 - \bar{\alpha}^T)}{1 - \bar{\alpha}} + 2\|P_{\bar{\alpha}}\| \frac{\sigma}{1 - \sigma} WT.$$

It is then evident that there exist $C_w, C_0 \in \mathbb{R}_+$, such that

$$J_T^d(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \leq C_0 + C_w T.$$

Moreover, such C_0 and C_w exist for all costs of the form (6). Hence, for all such $\{c_t\}_{t=0}^T$, the regret will also attain the same bound. Thus, while the considered policy achieves linear regret, the closed-loop with the system is unstable.

This example shows how the violation of Assumption 2.7 can make regret hide the instability of the system.

3.1 Linear Time-Varying Systems

The following theorem establishes the regret-stability relation for the time-varying policy $\mu_t(x_t) = -K_t x_t$.

Theorem 3.2. *Assume that Assumption 2.7 holds. Given the cost function (2) and LTV system (1), consider a linear time-varying state feedback policy $\mu_t(x_t) = -K_t x_t$. If the closed-loop system (4) under this policy is BIBS stable, and if there exists $D \in \mathbb{R}_+$ such that*

$$\lim_{T \to \infty} \sum_{t=0}^T \|\Phi(t, 0)\| \leq D, \qquad (7)$$

then the policy μ attains linear regret.

Proof. Using the condition in Assumption 2.7 and defining $M := \overline{M}(1 + \max_{0 \leq t < T} \|K_t\|^2)$

$$J_T(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) = \sum_{t=0}^T c_t(x_t, K_t x_t) \leq M \sum_{t=0}^T \|x_t\|^2$$

$$= M \sum_{t=0}^T \left\| \Phi(t, 0) x_0 + \sum_{k=1}^t \Phi(t, k) w_{k-1} \right\|^2$$

$$\leq 2M \|x_0\|^2 \sum_{t=0}^T \|\Phi(t, 0)\|^2 + 2MW^2 \sum_{t=0}^T \sum_{k=1}^t \|\Phi(t, k)\|^2$$

$$\leq 2M\overline{D}X^2 + 2M\overline{H}W^2 T,$$

where $\overline{D}, \overline{H} \in \mathbb{R}_+$ are such that

$$\lim_{T \to \infty} \sum_{t=0}^T \|\Phi(t, 0)\|^2 \leq \overline{D} < \infty, \qquad (8)$$

$$\lim_{t \to \infty} \sum_{k=1}^t \|\Phi(t, k)\|^2 \leq \overline{H} < \infty. \qquad (9)$$

The limit in (8) exists due to the condition in (7) and the fact that the series contains only non-negative terms. The limit in (9) exists due to the BIBS assumption and the series of non-negative terms only. Since $J_T(x_0, \boldsymbol{u}; \boldsymbol{w}) \geq 0$ for all $\boldsymbol{u}$, the regret will necessarily attain the same bound with $C_w = 2M\overline{H}W^2$ and $C_0 = 2M\overline{D}X^2$. $\qquad \square$

We note that given Assumption 2.5.iii. holds, the condition in (7) is stronger than asymptotic stability. In fact, if (7) is satisfied, then Assumption 2.5.iii. implies asymptotic stability of the closed-loop system.

The following theorem establishes the implication of linear regret on the stability of the closed-loop system with linear time-varying state feedback policies.

Theorem 3.3. *Assume that Assumptions 2.5 and 2.7 hold. Given the cost function (2) and LTV system (1), if a linear time-varying state feedback policy $\mu_t(x_t) = -K_t x_t$ attains linear regret, then the closed-loop system (4) under this policy is asymptotically stable.*

Proof. Given Assumption 2.5.i. holds, there exists a policy $\overline{\mu}_t = -\overline{K}_t x_t$ such that the closed-loop system (4) under this policy is exponentially stable. It can be shown that this implies BIBS stability (Ludyk, 2013), and the bound in (7) is always satisfied. Specifically, from the definition of exponential stability, it can be inferred that taking $t_0 = 0$ and $D = \frac{d}{1-\delta}$ the bound in (7) is achieved. Therefore, from Theorem 3.2 there exist some $C_w^\star, C_0^\star \in \mathbb{R}_+$, such that

$$J_T(x_0, \boldsymbol{u}^\star; \boldsymbol{w}) \leq J_T(x_0, \boldsymbol{u}^{\overline{\mu}}; \boldsymbol{w}) \leq C_0^\star + C_w^\star T,$$

where the first inequality follows from the definition of $\boldsymbol{u}^\star$. Then, linear regret of any policy μ is equivalent to

$$J_T(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \leq \overline{C}_0 + \overline{C}_w T, \qquad (10)$$

where $\overline{C}_w = (C_w^\star + C_w)$ and $\overline{C}_0 = (C_0^\star + C_0)$. Assume, for the sake of contradiction, that there exists a policy $\mu_t(x_t) = -K_t x_t$, that attains the bound in (10) but the associated closed-loop system (4) is not asymptotically stable. This is equivalent (Lemma 2.6 and Assumption 2.5.ii.) to assuming that either the limit in (5) does not exist or it is greater than zero, or equivalently

$$\limsup_{T \to \infty} \|\Phi(T, 0)\| \geq \liminf_{T \to \infty} \|\Phi(T, 0)\| > 0. \qquad (11)$$

Using the lower bound in Assumption 2.7, we then have

$$J_T(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) = \sum_{t=0}^T c_t(x_t, K_t x_t) \geq \underline{M} \sum_{t=0}^T \|x_t\|^2$$

$$= \underline{M} \sum_{t=0}^T \| \Phi(t, 0) x_0 + \sum_{k=1}^t \Phi(t, k) w_{k-1} \|^2.$$

Consider first the case when $\liminf_{T \to \infty} \|\Phi(T, 0)\| = \infty$. Then, necessarily there exists a $\bar{x}_0 \neq \boldsymbol{0}$, such, that $\liminf_{T \to \infty} \|\Phi(T, 0)\bar{x}_0\| = \infty$. For $x_0 = \bar{x}_0$ and $w_t = \boldsymbol{0}$ for all $t \in \mathbb{N}$, one can take the time-averaged limit of the above

$$\liminf_{T \to \infty} \frac{1}{T} J_T(\bar{x}_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \geq \liminf_{T \to \infty} \frac{M}{T} \sum_{t=0}^{T} \|\Phi(t,0)\bar{x}_0\|^2$$

$$\geq \liminf_{T \to \infty} \underline{M} \|\Phi(T,0)\bar{x}_0\|^2 = \infty,$$

where the last inequality follows from the Stolz-Césaro theorem (Choudary and Niculescu, 2014). This leads to a contradiction with the upper bound in (10)

$$\liminf_{T \to \infty} \frac{1}{T} J_T(\bar{x}_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \leq \overline{C}_w.$$

Given Assumption 2.5.ii., the only other remaining case is when the upper and lower limits in (11) are finite (whether equal or not) and non-zero. For this case, consider $x_0 = \mathbf{0}$ and $w_{k-1} = C\Phi(k,0)\overline{w}_0$, where

$$C = \min_{k \geq 0} \left[\frac{W}{\|\Phi(k,0)\overline{w}_0\|} \right],$$

and the vector $\overline{w}_0 \in \mathbb{R}^n$ is non-zero. Since $\limsup_{T \to \infty} \|\Phi(T,0)\|$ is finite, such a $C \in \mathbb{R}_+$ exists. Then

$$J_T(x_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \geq \underline{M}C^2 \sum_{t=0}^{T} \|\sum_{k=1}^{t} \Phi(t,k)\Phi(k,0)\overline{w}_0\|^2$$

$$= \underline{M}C^2 \sum_{t=1}^{T} \|\sum_{k=1}^{t} \Phi(t,0)\overline{w}_0\|^2 = \underline{M}C^2 \sum_{t=1}^{T} t^2 \|\Phi(t,0)\overline{w}_0\|^2.$$

By taking the time-averaged limits of both sides of the inequality and applying the Stolz-Césaro theorem again

$$\liminf_{T \to \infty} \frac{1}{T} J_T(\bar{x}_0, \boldsymbol{u}^\mu; \boldsymbol{w}) \geq \underline{M}C^2 \liminf_{T \to \infty} T^2 \|\Phi(T,0)\overline{w}_0\|^2.$$

The lower bound becomes unbounded, contradicting (10) and completing the proof by contraposition. $\square$

Thus, under suitable conditions, linear regret guarantees asymptotic stability. However, as shown in Theorem 3.2 and also pointed out in (Nonhoff and Müller, 2022) for a similar setting, the converse is not always the case, i.e. asymptotic stability alone does not imply linear regret.

Remark 3.1. *The results can be generalized to affine policies of the form $\mu(x_t) = -K_t x_t + d_t$, for some $d_t \in \mathbb{R}^m$, given that d_t are bounded for all $t \in \mathbb{N}$.*

Remark 3.2. *Note that inferring stability from regret requires not only stabilizability (Assumption 2.5.i.) but also certain conditions on the benchmark. Indeed to guarantee (10), it is necessary that a stabilizing policy exists in the feasible space of the benchmark. This holds here since the benchmark is the non-causal controller $\boldsymbol{u}^\star$ (dynamic regret), but weaker benchmarks that are constrained to specific policy classes can also be considered (policy regret).*

3.2 Linear Time-Invariant Systems

In this section, we consider the optimal control problem of minimizing the cost (2) subject to the LTI dynamics

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{12}$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$. The following establishes the relation between linear regret and stability in this setting.

Theorem 3.4. *Assume that Assumptions 2.5.i., and 2.7 hold. Given the cost function (2) and LTI system (12), a linear stationary state feedback policy $\mu(x_t) = -Kx_t$ attains linear regret if and only if the closed-loop system $x_{t+1} = Fx_t$ under this policy is asymptotically stable.*

Proof. To show that the required regret bound is achieved for an asymptotically stabilizing stationary policy, consider the state at timestep $0 \leq t < T$, given by

$$x_t = F^t x_0 + \sum_{k=0}^{t-1} F^k w_{t-k-1},$$

taking $w_{-1} = \mathbf{0}$. Using the cost bounds in Assumption 2.7 and denoting $M := \overline{M}(1 + \|K\|^2)$

$$\sum_{t=0}^{T} c_t(x_t, Kx_t) \leq \sum_{t=0}^{T} \overline{M} \left(\|x_t\|^2 + \|K\|^2 \|x_t\|^2 \right)$$

$$= M \sum_{t=0}^{T} \|F^t x_0 + \sum_{k=0}^{t-1} F^k w_{t-k-1}\|^2$$

$$\leq 2Mg^2 \sum_{t=0}^{T} \left(\varepsilon^{2t} \|x_0\|^2 + W^2 \left(\frac{1-\varepsilon^t}{1-\varepsilon} \right)^2 \right)$$

$$\leq \frac{2Mg^2}{1-\varepsilon^2} X^2 + \frac{2Mg^2}{(1-\varepsilon)^2} W^2 T,$$

where we used the fact that for $\rho(F) < 1$ there exists a $g \in \mathbb{R}_+$, such that $\|F^k\| \leq g\varepsilon^k$ for all $k > 0$, where $\varepsilon := \frac{1+\rho(F)}{2} \in (0,1)$ (Horn and Johnson, 2012). Since costs are non-negative, regret attains the same bound.

To prove the reverse statement, assume, for the sake of contradiction, that there exists a matrix K' such that $\rho(A - BK') \geq 1$ attaining linear regret. From the previous analysis, any stabilizing state feedback matrix K attains a cost that scales linearly with the time horizon. Moreover, such a matrix exists as for LTI systems Assumption 2.5.i. corresponds to the stabilizability of the pair (A, B). Then, using the same arguments as in the proof of Theorem 3.3, there exist $\overline{C}_w, \overline{C}_0 \in \mathbb{R}_+$ such that

$$\sum_{t=0}^{T} c_t(x_t, K'x_t) \leq \overline{C}_0 + \overline{C}_w T. \tag{13}$$

From Assumption 2.7 it is true that

$$\sum_{t=0}^{T} c_t(x_t, K'x_t) \geq \underline{M} \sum_{t=0}^{T} \|F'^t x_0 + \sum_{k=0}^{t-1} F'^k w_{t-k-1}\|^2.$$

Since the result should hold for any initial state and any disturbance within the defined Euclidian ball, consider $x_0 = \mathbf{0}$ and $w_t = \overline{w}$ $\forall t \in \mathbb{N}$ such that $\|\overline{w}\| = W$ and $F'\overline{w} = \rho(F')\overline{w}$. We then have

$$\underline{M} \sum_{t=0}^{T} \|\sum_{k=0}^{t-1} F'^k w_{t-k-1}\|^2 = \underline{M} \sum_{t=0}^{T} \|\sum_{k=0}^{t-1} F'^k \overline{w}\|^2$$

$$= \underline{M} \sum_{t=0}^{T} \|\sum_{k=0}^{t-1} \rho(F')^k \overline{w}\|^2 \geq \underline{M} \sum_{t=0}^{T} t^2 W^2 \geq \underline{M}W^2 \frac{T^2 + T}{2},$$

leading to a contradiction with (13). $\square$

In contrast to the LTV case, in the LTI setting, asymptotic and exponential stability are equivalent, leading to an equivalency between linear regret and stability.

4. NUMERICAL EXAMPLE

To visualize the necessary and sufficient condition in Theorem 3.4, a simple two-dimensional system with single input
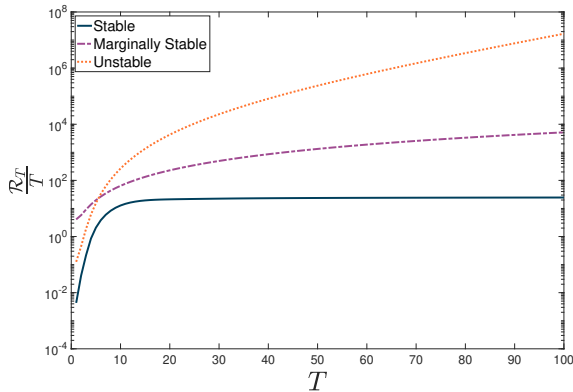
Fig. 1. Time-averaged regret, $\frac{\mathcal{R}_T}{T}$ for a LTI system on a semilog scale. The stable system can be upper bounded by a constant while the marginally stable and unstable ones scale with an order of $log(T)$.

is considered. In particular, for $A = [1 \quad 1; 0 \quad 1]$ and $B = [1; 0.5]$, three LTI state feedback controllers are considered, $K_1 = [0.2 \quad 0.4]$, $K_2 = [0 \quad 1]$, and $K_3 = [-0.02 \quad 0.5]$. These produce respectively, stable, marginally stable, and unstable closed-loop systems. The cost function is taken to be quadratic with $Q = [1.5 \quad 0; 0 \quad 1.5]$ and $R = 1$ as the state and input weighting matrices, respectively. The disturbance w_t for all $0 \le t < T$ is taken to be the normalized eigenvector of the closed-loop system matrix corresponding to the largest eigenvalue, to approximate the worst-case regret for the given policy. The time-averaged regret for each of the controllers is calculated for a time horizon ranging from 1 to 100 and is plotted in a logarithmic scale in Figure 1. Specifically, the average regret of the stable controller can be upper bounded by a constant, that of the marginally stable controller scales with T and for the unstable one with a higher order of T.

5. CONCLUSIONS

In this work, we studied the interconnection of the notion of regret coming from online optimization and the control theoretic concept of stability. Given a linear state feedback policy that attains linear regret, and certain upper and lower bounds on the objective stage costs, we show that the closed-loop system is necessarily asymptotically stable, both for the time-varying and time-invariant cases. The converse result also holds given that the closed-loop system is BIBS stable and has absolute summable norms of its state transition matrices. The results can be used to directly prove the stability of algorithms with regret guarantees and vice versa. This work can be a stepping stone for the consideration of adaptive policies, under which the considered setting is no longer linear; this will allow the analysis of a wider range of online algorithms.

REFERENCES

Agarwal, N., Hazan, E., and Singh, K. (2019). Logarithmic regret for online control. *Advances in Neural Information Processing Systems*, 32.

Bertsekas, D.P. (2015a). Dynamic programming and optimal control 4th edition, volume ii. *Athena Scientific*.

Bertsekas, D.P. (2015b). Value and policy iterations in optimal control and adaptive dynamic programming. *IEEE transactions on neural networks and learning systems*, 28(3), 500–509.

Callier, F.M. and Desoer, C.A. (2012). *Linear system theory*. Springer Science & Business Media.

Choudary, A.D.R. and Niculescu, C.P. (2014). *Real analysis on intervals*. Springer.

Cohen, A., Hasidim, A., Koren, T., Lazic, N., Mansour, Y., and Talwar, K. (2018). Online linear quadratic control. In *International Conference on Machine Learning*, 1029–1038. PMLR.

Dean, S., Mania, H., Matni, N., Recht, B., and Tu, S. (2018). Regret bounds for robust adaptive control of the linear quadratic regulator. *Advances in Neural Information Processing Systems*, 31.

Didier, A., Sieber, J., and Zeilinger, M.N. (2022). A system level approach to regret optimal control. *IEEE Control Systems Letters*.

Goel, G. and Hassibi, B. (2022). The power of linear controllers in lqr control. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, 6652–6657. IEEE.

Gradu, P., Hazan, E., and Minasyan, E. (2020). Adaptive regret for control of time-varying dynamics. *arXiv preprint arXiv:2007.04393*.

Horn, R.A. and Johnson, C.R. (2012). *Matrix analysis*. Cambridge university press.

Jedra, Y. and Proutiere, A. (2022). Minimal expected regret in linear quadratic control. In *International Conference on Artificial Intelligence and Statistics*, 10234–10321. PMLR.

Karapetyan, A., Iannelli, A., and Lygeros, J. (2022). On the regret of $\mathcal{H}_\infty$ control. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, 6181–6186. IEEE.

Ludyk, G. (2013). *Stability of time-variant discrete-time systems*, volume 5. Springer-Verlag.

Martin, A., Furieri, L., Dörfler, F., Lygeros, J., and Ferrari-Trecate, G. (2022). Safe control with minimal regret. In *Learning for Dynamics and Control Conference*, 726–738. PMLR.

Nonhoff, M. and Müller, M.A. (2022). On the relation between dynamic regret and closed-loop stability. *arXiv preprint arXiv:2209.05964*.

Postoyan, R., Buşoniu, L., Nešić, D., and Daafouz, J. (2016). Stability analysis of discrete-time infinite-horizon optimal control with discounted cost. *IEEE Transactions on Automatic Control*, 62(6), 2736–2749.

Sabag, O., Goel, G., Lale, S., and Hassibi, B. (2021). Regret-optimal full-information control. *arXiv preprint arXiv:2105.01244*.

Shi, Y. and Chen, G. (2009). Chaos of time-varying discrete dynamical systems. *Journal of Difference Equations and applications*, 15(5), 429–449.

Simchowitz, M. and Foster, D. (2020). Naive exploration is optimal for online LQR. In *International Conference on Machine Learning*, 8937–8948. PMLR.

Yu, C., Shi, G., Chung, S.J., Yue, Y., and Wierman, A. (2020). The power of predictions in online control. *Advances in Neural Information Processing Systems*, 33, 1994–2004.

Zhang, R., Li, Y., and Li, N. (2021). On the regret analysis of online LQR control with predictions. In *2021 American Control Conference (ACC)*, 697–703. IEEE.