Passive-Aggressive Learning and Control

Dimitar Ho, Nikolai Matni and John C. Doyle

***Abstract*— In this work, we investigate the problem of simultaneously learning and controlling a system subject to adversarial choices of disturbances and system parameters. We study the problem for a scalar system with l_∞-norm bounded disturbances and system parameters constrained to lie in a known bounded convex polytope. We present a controller that is globally stabilizing and gives continuously improving bounds on the worst case state deviation. The proposed controller simultaneously learns the system parameters and controls the system. The controller emerges naturally from an optimization problem, and balances exploration and exploitation in such a way that it is able to efficiently stabilize unstable and adversarial system dynamics. Specifically if the controller is faced with large uncertainty, the initial focus is on exploration, retrieving information about the system by applying state-feedback controllers with varying gains and signs. In a pre-specified bounded region around the origin, our control strategy can be seen as *passive* in the sense that it learns very little information. Only once the noise and/or system parameters act in an adversarial way, leading to the the state exiting the aforementioned region for more than one time-step, our proposed controller behaves *aggressively* in that it is guaranteed to learn enough about the system to subsequently robustly stabilize it. We end by demonstrating the efficiency of our methods via numerical simulations.**

I. Introduction

With the proliferation of big-data, and the success of machine-learning algorithms being applied to planning and control problems, there has been a renewed interest in combining and applying learning and control to continuous systems. Modern results build on the foundational ideas of adaptive control [1], [2], which we cannot hope to adequately survey here, but place an emphasis on finite-time, rather than asymptotic, guarantees of performance and stability.

To the best of our knowledge, recent results of this nature have focused on the stochastic setting wherein system parameters are unknown, and must be identified despite stochastic excitations to the system. By combining concentration results from high-dimensional statistics with techniques from robust and optimal control, regret and performance bounds can be obtained as a function of the number of data points seen by the controller. Notable examples include [3]–[7], which provide varying degrees of guarantees and practically applicable algorithms. However, as far as we are aware, no comparable results exist for the setting of bounded but adversarial process noise and parametric uncertainty.

Recently it has been shown that in the l_∞ bounded adversarial setting, solutions to the state-estimation problem [8], [9] and the robust control problem subject to quantization and delay in the control loop [10] admit particularly intuitive and appealing forms. This work shows that the same holds true for a joint learning and control problem.

Our main contribution is what we call a passive-aggressive learning and control algorithm that trades off between identifying the true system parameters and stabilizing the system. The defining feature of this controller is that unless the system parameters and noise act in an adversarial way, pushing the state sufficiently far away from the origin, it is content with passively observing the state evolution and updating its uncertainty set. However, when the system conspires to push the state sufficiently far from the origin, it aggressively learns the system parameters and applies control actions aimed at stabilizing the system.

The rest of the paper is organized as follows: in Section II we define the problem and the necessary notions of consistent parameter sets given a sequence of observations. In Section III we then show that in the case of "strongly stabilizable" initial parameter uncertainty sets, a simple static state-feedback policy is sufficient to guarantee robust stability for all possible choices of system realization. We then build on this result in Section IV to show that if the controller updates the set of feasible system parameters with each observation, the controller performance can be strictly improved. Finally, in Section V we consider the case of general initial uncertainty sets, and show that if a two-stage controller is applied, then the uncertainty set can eventually be reduced to one that is strongly stabilizable, allowing us to switch to the aforementioned control policies. We demonstrate the efficacy of our approach in Section VI, and end with conclusions and future work in Section VII.

II. System and Problem Definition

A. System Dynamics

We consider the scalar linear discrete-time system

$$x_{n+1} = ax_n + bu_n + w_n \tag{1}$$

$$|w_n| \leq \eta \quad \begin{bmatrix} a \\ b \end{bmatrix} \in \mathcal{P}_0 \tag{2}$$

with the state x_n, the control input u_n and the disturbance w_n. We assume that the disturbance w_n is l_∞ bounded by η, and that the state-space parameters a and b are unknown constants, but that are constrained to lie in a known bounded convex polytope $\mathcal{P}_0$.

We furthermore assume that

$$b \neq 0 \ \forall \begin{bmatrix} a \\ b \end{bmatrix} \in \mathcal{P}_0, \tag{3}$$

The authors are with the department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA ({dho,nmatni,doyle}@caltech.edu).

Thanks to funding from AFOSR and NSF and gifts from Huawei and Google.

i.e., that the system is controllable for all possible realizations of the unknown state-space parameters (a, b).

B. Consistent Sets

We denote by $x_{i:j}$ and $u_{i:j}$ the stacked vector of state values x_n and control inputs u_n, respectively, for $i \leq n \leq j$. It follows from the dynamics (1) that at time N, the following entry-wise inequality must hold for the true parameters (a, b).

$$\left| x_{1:N} - [x_{0:N-1}, u_{0:N-1}] \begin{bmatrix} a \\ b \end{bmatrix} \right| \leq \mathbb{1}\eta, \tag{4}$$

where $\mathbb{1}$ is the all ones vector of compatible dimension. This inequality therefore allows us to characterize the subset of the initial uncertainty set $\mathcal{P}_0$ that is consistent with the observed state and control input histories given the known bound η on the magnitude of the disturbance process w_n.

This motivates the following definition of the *consistent set* at time N:

$$\begin{aligned} &\mathcal{S}(x_{0:N}, u_{0:N-1}) \\ &:= \left\{ \begin{bmatrix} a \\ b \end{bmatrix} \middle| \left| x_{1:N} - [x_{0:N-1}, u_{0:N-1}] \begin{bmatrix} a \\ b \end{bmatrix} \right| \leq \mathbb{1}\eta \right\}. \end{aligned} \tag{5}$$

It then follows that given state and control histories $x_{0:N}$, $u_{0:N-1}$ and the initial uncertainty set P_0, we have that $[a, b]^T$ lies in the bounded convex polytope $P_0 \cap \mathcal{S}(x_{0:N}, u_{0:N-1})$.

For $N = 1$, the set $\mathcal{S}(x_{0:1}, u_0)$ reduces to a slice of thickness $2\eta/\sqrt{x_0^2 + u_0^2}$ with normal vectors $\pm[x_0, u_0]^T$ in parameter space (see Fig.1). This motivates the following recursive definition of the consistent set $\mathcal{S}(x_{0:N}, u_{0:N-1})$ at time N as the intersection of N such slices:

$$\mathcal{S}(x_{0:N}, u_{0:N-1}) = \bigcap_{i=0}^{N-1} \mathcal{S}(x_{i:i+1}, u_i). \tag{6}$$

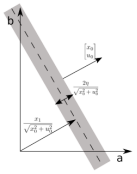


Fig. 1: Example of $S(x_{0:1}, u_0)$

C. Problem Statement

Our objective is to find the best causal control strategy $u_k(x_{0:k}, \mathcal{P}_0)$ that minimizes the worst-case state deviation $\|x_{1:\infty}\|_\infty$ despite adversarial noise $w_{0:\infty}$ and system parameter choices $[a, b] \in \mathcal{P}_0$.

Formally, we seek a solution to the following infinite-horizon min-max problem

$$V(x_0, \mathcal{P}_0) := \min_{u_{0:\infty}} Q^{1:\infty}(x_0, \mathcal{P}_0, u_{0:\infty}) \tag{7}$$

where we define $Q^{1:N}(x_0, \mathcal{P}_0, u_{0:N-1})$ as

$$\begin{aligned} Q^{1:N}(x_0, \mathcal{P}_0, u_{0:N-1}) := \max_{[\begin{smallmatrix} a \\ b \end{smallmatrix}] \in \mathcal{P}_0} \max_{\|w_{0:N-1}\|_\infty \leq \eta} & \|x_{1:N}\|_\infty \\ \text{s.t.} \quad & \text{dynamics (1).} \end{aligned}$$

Our approach is to begin with what we term *strongly stabilizable* initial uncertainty sets $\mathcal{P}_0$, that is to say initial uncertainty sets for which an appropriately chosen static state-feedback gain is guaranteed to be stabilizing for all realizations of the system parameters (a, b). We show that such a policy is optimal for optimization problem (7) restricted to static memoryless control policies. We then show that by adding an adaptive element to such a control policy, the performance can be further improved, and that an exploration/exploitation strategy naturally emerges. Finally, we tackle the case of general initial uncertainty sets and show that after an initial "passive" learning phase, the uncertainty set is eventually "aggressively" reduced to one that is strongly stabilizable, allowing for our previously derived adaptive strategy to be applied.

III. Robust Static State-Feedback for Strongly Stabilizable Uncertainty Sets

In this section, we consider a restriction of optimization problem (7) to static state-feedback control policies, i.e., we restrict $u_n = kx_n$ for all $n \geq 0$. The resulting optimization problem then reads as

$$V_{RSF}(x_0, \mathcal{P}_0) \tag{8}$$

$$\begin{aligned} = \min_k \max_{[\begin{smallmatrix} a \\ b \end{smallmatrix}] \in \mathcal{P}_0} \max_{\|w_{0:\infty}\|_\infty \leq \eta} & \|x_{1:\infty}\|_\infty \\ \text{s.t.} \quad & x_{n+1} = (a + bk)x_n + w_n, \\ & \forall n \geq 0. \end{aligned} \tag{9}$$

$$= \min_{u_{0:k} = kx_{0:k}} Q^{1:\infty}(x_0, \mathcal{P}_0, u_{0:\infty}) \tag{10}$$

$$=: \min_k Q^{1:\infty}_{RSF}(x_0, \mathcal{P}_0, k) \tag{11}$$

As we are restricting ourselves to static state-feedback polices, it follows that $V_{RSF}(x_0, \mathcal{P}_0)$ is an upper bound for $V_0(x_0, \mathcal{P}_0)$, i.e.:

$$V_0(x_0, \mathcal{P}_0) \leq V_{RSF}(x_0, \mathcal{P}_0). \tag{12}$$

We consider this simpler problem as it has several appealing properties. First, the optimal cost-to-go $V_{RSF}(x_0, \mathcal{P}_0)$ and the corresponding minimizing k^* can be solved for in closed form. Further it motivates the definition of *strongly stabilizble* initial uncertainty sets $\mathcal{P}_0$, which naturally captures how difficult an uncertain system can be to stabilize. To that end, we introduce the following measure of stabilizability for an uncertainty set $\mathcal{P}$.

Definition III.1. We use $\lambda(\mathcal{P})$ to denote the *stability margin* of the parameter set $\mathcal{P}$, and define it as

$$\lambda(\mathcal{P}) := \min_k \max_{[\begin{smallmatrix} a \\ b \end{smallmatrix}] \in \mathcal{P}} |a + bk|. \tag{13}$$

Furthermore, we call the corresponding minimizer

$$K(\mathcal{P}) = \text{argmin}_k \max_{[\begin{smallmatrix} a \\ b \end{smallmatrix}] \in \mathcal{P}_0} |a + bk| \tag{14}$$

the *gain* of the parameter set $\mathcal{P}_0$.

The stability margin of a set $\lambda(\mathcal{P})$ is a functional mapping sets to $\mathbb{R}_0^+$ and describes the smallest system eigenvalue

achievable by a constant state-feedback, assuming worst case parameter choice of $[a,b]^T \in \mathcal{P}$. It then follows that if for the initial uncertainty set $\mathcal{P}_0$, it holds that the stability margin satisfies $\lambda(\mathcal{P}_0) < 1$, then we can use $k = K(\mathcal{P}_0)$ as a state-feedback control law to stabilize the system for all parameters in $\mathcal{P}_0$. We will refer to such initial uncertainty sets $\mathcal{P}_0$ as *strongly stabilizable*. On the other hand, if $\lambda(\mathcal{P}_0) \geq 1$ then for any state-feedback controller $u_n = kx_n$ there will exist some $[a_0, b_0] \in \mathcal{P}_0$ that leads to an unstable closed-loop system, and therefore one cannot guarantee stability of the closed loop system for any static state-feedback control policy.

As the next lemma shows, when the initial uncertainty set is strongly stabilizable, the optimal solution to the static state-feedback control problem (11) is precisely given by the gain of the parameter set $\mathcal{P}_0$, and the cost-to-go is governed by its stability margin.

Lemma III.1. *If the initial uncertainty set $\mathcal{P}_0$ is strongly stabilizable (i.e., if $\lambda(\mathcal{P}_0) < 1$), then the cost-to-go $V_{RSF}(x_0, \mathcal{P}_0)$, as described in equation* (11)*, is minimized by the choosing $k = K(\mathcal{P}_0)$, and the optimal value is given by*

$$\max\{\lambda(\mathcal{P}_0)|x_0| + \eta, \frac{1}{1-\lambda(\mathcal{P}_0)}\eta\}. \tag{15}$$

Conversely, if the initial uncertainty set $\mathcal{P}_0$ is not strongly stabilizable (i.e., if $\lambda(\mathcal{P}_0) \geq 1$), then for any choice of k, it holds that $V_{RSF}(x_0, \mathcal{P}_0) = \infty$.

Proof. Notice that $x_{n+1} = (a+bk)x_n + w_n$ implies

$$x_n = (a+bk)^N x_0 + \sum_{i=0}^{N-1} (a+bk)^{n-1-i} w_i \tag{16}$$

and

$$\begin{aligned} \max_{\|w_k\|_\infty \leq \eta} \quad & |x_N| \\ s.t. \quad & x_{n+1} = (a+bk)x_n + w_n \end{aligned} \tag{17}$$

$$= |a+bk|^N |x_0| + \frac{|a+bk|^N - 1}{|a+bk| - 1}\eta \tag{18}$$

Therefore,

$$\begin{aligned} \max_{\|w_n\|_\infty \leq \eta} \quad & \|x_{1:\infty}\|_\infty \\ \text{s.t.} \quad & x_{n+1} = (a+bk)x_n + w_n, \ \forall n \geq 0 \end{aligned} \tag{19}$$

$$\begin{aligned} = \max_{\substack{N \in \mathbb{N} \\ N \geq 1}} \max_{\|w_n\|_\infty \leq \eta} \quad & |x_N| \\ \text{s.t.} \quad & x_{n+1} = (a+bk)x_n + w_n, \ \forall n \geq 0 \end{aligned} \tag{20}$$

$$= \max_{\substack{N \in \mathbb{N} \\ N \geq 1}} |a+bk|^N |x_0| + \frac{|a+bk|^N - 1}{|a+bk| - 1}\eta \tag{21}$$

$$=: C(|a+bk|, |x0|, \eta) \tag{22}$$

So $V_{RSF}(x_0, \mathcal{P}_0)$ can be written as

$$V_{RSF}(x_0, \mathcal{P}_0) = \min_k \max_{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathcal{P}_0} C(|a+bk|, |x_0|, \eta) \tag{23}$$

No since $C(|a+bk|, |x_0|, \eta)$ is monotonic in $|a+bk|$, we obtain

$$V_{RSF}(x_0, \mathcal{P}_0) = C(\lambda(\mathcal{P}_0), |x_0|, \eta) \tag{24}$$

$$= \begin{cases} \infty & \text{if } \lambda(\mathcal{P}_0) \geq 1 \\ \max\limits_{\substack{n \in \mathbb{N} \\ n \geq 1}} \lambda(\mathcal{P}_0)^n |x_0| + \frac{1-\lambda(\mathcal{P}_0)^n}{1-\lambda(\mathcal{P}_0)}\eta & \text{if } \lambda(\mathcal{P}_0) < 1 \end{cases} \tag{25}$$

with $k^* = K(\mathcal{P}_0)$ as the minimizer state-feedback gain.

Finally, the cost achieved for strongly stabilizable initial uncertainty sets is seen to be monotonic increasing (decreasing) in n if $(1-\lambda(\mathcal{P}_0)|x_0| > (<)\eta$, and hence is maximized at either $n = 1$ or $n = \infty$, from which the cost in equation (15) follows. □

With this result, we see that for strongly stabilizable initial uncertainty sets $\mathcal{P}_0$, we can pick a static state-feedback controller with $k = K(\mathcal{P}_0)$ that is guaranteed to robustly stabilize the system for any system realization $(a, b) \in \mathcal{P}_0$, and the achieved optimal cost (15) provides a finite upper bound to $V(x_0, \mathcal{P}_0)$. This result also shows that no static feedback gain can be guaranteed to stabilize the system if $\mathcal{P}_0$ is not strongly stabilizable, i.e. $\lambda(\mathcal{P}_0) \geq 1$.

Although this robust feedback controller is guaranteed to stabilize the system, it does so in an inefficient way. In particular, it does not update its control policy to reflect the fact that with each observation, more information about the underlying true parameters is revealed. As previously discussed, the observations $x_{0:N}, u_{0:N-1}$ allow us to reduce the space of consistent parameters $[a,b]^T$ to be $P_0 \cap S(x_{0:N}, u_{0:N-1})$. In what follows we improve upon the static state-feedback policy results of this section, and ultimately show that learning is necessary to compute stabilizing controllers for general initial uncertainty sets $\mathcal{P}_0$.

IV. Robust Adaptive State-Feedback for Strongly Stabilizable Uncertainty Sets

Keeping our focus on strongly stabilizable initial uncertainty sets, we propose two controllers that strictly outperform the static state-feedback policy $k = K(\mathcal{P}_0)$ defined in the previous section. The following adaptive schemes, which we call *weakly adaptive RSF* and *strongly adaptive RSF*, simultaneously learn the system dynamics while controlling the system. The latter algorithm decides at every time-step n between a control action that reduces $|x_{n+1}|$ and an exploratory control action that leads to more information about the system parameters. Our key result is a decomposition theorem that exploits the fact that the control policy at time n is allowed to be a function of all past state-measurements $x_{0:n}$. In what follows we let $a \vee b := \max\{a, b\}$ to help simplify notation.

A. Weakly Adaptive Robust State-Feedback Controller

The robust state-feedback controller $u_{RSF}(x)$ at time step n is $K(\mathcal{P}_0)x_n$ and is not using the information of recent observations. A rather simple yet significantly better strategy is to apply $K(\mathcal{P}_n)x_n$ at time-step n. We will refer to

this controller as the *weakly adaptive robust state-feedback controller* $u^{WRSF}(x_{0:n})$:

$$u^{WRSF}(x_{0:n}) = K(\mathcal{P}_n)x_n \tag{26}$$

B. Strongly Adaptive Robust State-Feedback

Definition IV.1. Define the one-step reachable set $\mathcal{R}(x_0, u_0, \mathcal{P}_0)$ as the set of possible x_1 given the initial condition x_0, the initial uncertainty set $\mathcal{P}_0$ and the initial disturbance w_0, i.e.

$$\mathcal{R}_\eta(x_0, u_0, \mathcal{P}_0) \tag{27}$$

$$= \left\{ ax_0 + bu_0(x_0) + w_0 | \begin{bmatrix} a \\ b \end{bmatrix} \in \mathcal{P}_0,\ |w_0| \le \eta \right\} \tag{28}$$

We then have that the following decomposition theorem holds:

Theorem IV.2. *For the cost-to-go function V as defined in optimization problem* (7)*, it holds that*

$$\begin{aligned} &V(x_0, \mathcal{P}_0) \\ =&\min_{u_0} \big[Q^{1:1}(x_0, \mathcal{P}_0, u_0) \vee \\ &\max_{x_1 \in \mathcal{R}(x_0, \mathcal{P}_0, u_0)} V\left(x_1, \mathcal{P}_0 \cap \mathcal{S}(x_{0:1}, u_0)\right) \Big] \end{aligned}$$

Proof. First notice that we can write

$$\begin{aligned} &V\left(x_0, \mathcal{P}_0\right) \\ =&\min_{u_{0:\infty}} Q^{1:\infty}(x_0, \mathcal{P}_0, u_{0:\infty}) \\ =&\min_{u_{0:\infty}} \max_{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathcal{P}_0} \max_{\|w_{0:\infty}\|_\infty \le \eta} \|x_{1:\infty}\|_\infty \ \text{ s.t. dynamics (1)} \\ =&\min_{u_0} \Bigg[\max_{x_1 \in \mathcal{R}(x_0, u_0, \mathcal{P}_0)} \min_{u_{1:\infty}} \\ &\max_{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathcal{P}_0 \cap \mathcal{S}(x_{0:1}, u_0)} \max_{\|w_{1:\infty}\|_\infty \le \eta} |x_1| \vee \|x_{2:\infty}\|_\infty \Bigg] \\ &\text{s.t. dynamics (1)}, \end{aligned}$$

where the last equality follows from the fact that the state x_1 is known to the sequence of control actions $u_{1:\infty}$, and that $\|x_{1:\infty}\|_\infty = |x_1| \vee \|x_{2:\infty}\|_\infty$. This last line can further be rewritten as

$$\begin{aligned} &V\left(x_0, \mathcal{P}_0\right) \\ =&\min_{u_0} \Bigg[\left(\max_{x_1 \in \mathcal{R}(x_0, u_0, \mathcal{P}_0)} |x_1| \right) \vee \Bigg(\max_{x_1 \in \mathcal{R}(x_0, u_0, \mathcal{P}_0)} \\ &\dots \min_{u_{1:\infty}} \max_{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \in \mathcal{P}_0 \cap \mathcal{S}(x_{0:1}, u_0)} \max_{\|w_{1:\infty}\|_\infty \le \eta} \|x_{2:\infty}\|_\infty \Bigg) \Bigg] \\ &\text{s.t. dynamics (1)} \\ =&\min_{u_0} \big[Q^{1:1}(x_0, \mathcal{P}_0, u_0) \vee \\ &\max_{x_1 \in \mathcal{R}(x_0, u_0, \mathcal{P}_0)} V(x_1, \mathcal{P}_0 \cap \mathcal{S}(x_{0:1}, u_0), u_{1:\infty}) \big] \end{aligned}$$

where the first equality follows from the fact that $\max_x f(x) \vee g(x) = (\max_x f(x)) \vee (\max_x g(x))$, and the second from the definition of the cost-to-go function V, as defined in (7), and from the identity $\max_{x_1 \in \mathcal{R}(x_0, u_0, \mathcal{P}_0)} |x_1| = Q^{1:1}(x_0, \mathcal{P}_0, u_0(x_0))$. □

Theorem IV.2 sheds light on the structure of the optimal control policy. Specifically, let $\mathcal{P}_i := \mathcal{P}_0 \cap \mathcal{S}(x_{0:i}, u_{0:i-1})$; then the optimal control action at time i is given by[1]

$$u_i(x_{0:i}) = \text{argmin}_u \max_{x_{i+1} \in \mathcal{R}(x_i, \mathcal{P}_i, u)} |x_{i+1}| \vee \tag{29}$$

$$V\left(x_{i+1}, \mathcal{P}_i \cap \mathcal{S}(x_{i:i+1}, u)\right). \tag{30}$$

In particular, we see that the control action u_i is a function of both the state history $x_{0:i}$ and the updated uncertainty set $\mathcal{P}_i$, and naturally results in a trade-off between exploration and exploitation. If the first term dominates the cost function, this indicates that the controller is in an exploitation mode, using its gathered information on the uncertainty set to minimize state deviation. In contrast, if the second term dominates, this can be interpreted as an exploration action aimed at reducing the effects of parametric uncertainty on future state deviations.

Unfortunately, Eq. (29) do not provide a practical means of computing an optimal controller. Nevertheless, we can approximate (29) by using V_{RSF} as an upper bound for the cost-to-go function in (29). In particular, we suggest using the solution to the following optimization problem as the control policy at time i:

$$u_i^{SRSF}(x_i, \mathcal{P}_i) \tag{31}$$

$$= \text{argmin}_u \max_{x_{i+1} \in \mathcal{R}(x_i, \mathcal{P}_i, u)}$$

$$\dots |x_{i+1}| \vee V_{RSF}\left(x_{i+1}, \mathcal{P}_i \cap \mathcal{S}(x_{i:i+1}, u)\right) \tag{32}$$

$$V_i^{SRSF}(x_i, \mathcal{P}_i) \tag{33}$$

$$= \min_u \max_{x_{i+1} \in \mathcal{R}(x_i, \mathcal{P}_i, u)}$$

$$\dots |x_{i+1}| \vee V_{RSF}\left(x_{i+1}, \mathcal{P}_i \cap \mathcal{S}(x_{i:i+1}, u)\right) \tag{34}$$

$$\tag{35}$$

Thus to apply $\hat{u}^{SRSF}$ requires the solution of a scalar min-max problem at every time-step. As the next theorem shows, if the initial uncertainty set is strongly stabilizable, then this SRSF policy is stabilizing and performs at least as well as the static state-feedback policy described in the previous section for large initial conditions $x(0)$ – empirically however we see a strict and dramatic improvement in performance, and future work will look to close this gap between theory and practice.

C. Performance Bounds Comparison

Here we show that under suitably adversarial choices of control action and system parameters, both the weakly and strongly ARSF policies outperform the static memoryless policy defined in the previous section.

Theorem IV.3 (RSF vs. WRSF)**.** *Let x_n^{RSF} and x_n^{WRSF} be sequences generated from running u^{RSF} and u^{WRSF}*

[1]This follows from Theorem IV.2 by a simple induction argument which is omitted in the interest of space.

in closed loop with the same initial condition x_0. Then the sequences are bounded by

$$\left|x_n^{RSF}\right| \leq B_n^{RSF} \tag{36}$$
$$\left|x_n^{WRSF}\right| \leq B_n^{WRSF} \tag{37}$$

with the bounds B_n^{RSF} and B_n^{WRSF} defined as

$$B_n^{RSF} = \lambda(\mathcal{P}_0)^n \left|x_0\right| + \sum_{i=0}^{n-1} \lambda(\mathcal{P}_0)^i \eta \tag{38}$$

$$B_n^{WRSF}(\mathcal{P}_{0:n-1}) = \prod_{j=1}^{n} \lambda(\mathcal{P}_{j-1}) \left|x_0\right| + \sum_{i=0}^{n-1} \prod_{j=0}^{i} \lambda(\mathcal{P}_{j-1})^i \eta \tag{39}$$

$$\lambda(\mathcal{P}_{-1}) := 1 \tag{40}$$

Furthermore, $B_n^{WRSF} \leq B_n^{RSF}$ i.e. u^{WRSF} has a tighter performance bound.

Proof. Rolling out the dynamics

$$x_n = \prod_{j=1}^{n}(a+bk_{j-1})x_0 + \sum_{i=0}^{n-1}\prod_{j=0}^{i}(a+bk_j)w_i \tag{41}$$

we can upperbound $|x_n|$ by

$$|x_n| \leq \prod_{j=1}^{n}|(a+bk_{j-1})|x_0 + \sum_{i=0}^{n-1}\prod_{j=0}^{i}|(a+bk_j)|\eta \tag{42}$$

Now if we apply u^{RSF} and u^{SRSF} we can guarantee $|a+bk_i| \leq \lambda(\mathcal{P}_0)$ and $|a+bk_i| \leq \lambda(\mathcal{P}_i)$ respectively. This gives the defined bounds B_n^{WRSF} and B_n^{RSF}. Furthermore, since $\lambda(\mathcal{P}_{i+1}) \leq \lambda(\mathcal{P}_i) \leq \cdots \leq \lambda(\mathcal{P}_0) < 1$ we show $B_n^{WRSF} \leq B_n^{RSF}$. □

Theorem IV.4 (SRSF vs. WRSF)**.** *Let x_n^{WRSF} and x_n^{SRSF} be sequences generated from running u^{WRSF} and u^{SRSF} in closed loop with the same initial condition x_0. Furthermore, assume that both sequences obtain the same uncertainty sets $\mathcal{P}_n$, then x_n^{SRSF} and x_n^{WRSF} are upper bounded by*

$$|x_n^{SRSF}| \leq V^{SRSF}(x_{n-1}^{SRSF}, \mathcal{P}_{n-1}) \tag{43}$$
$$\left|x_n^{WRSF}\right| \leq B_n^{WRSF}(\mathcal{P}_{0:n-1}) \tag{44}$$

with B_n^{WRSF} defined as in (IV.3)*. Then $|x_n^{SRSF}| \leq B_n^{WRSF}(\mathcal{P}_{0:n-1})$, i.e. u^{SRSF} has no worse performance bound than u^{WRSF}.*

Proof. We prove our result by induction. Trivially, $|x_0^{SRSF}| \leq |x_0^{WRSF}| = B_0^{WRSF}$ Now for arbitrary j assume that $|x_j^{SRSF}| \leq B_j^{WRSF}$, then

$$\begin{aligned}|x_{j+1}^{SRSF}| &\leq V^{SRSF}(x_j^{SRSF}, \mathcal{P}_j) \leq V^{RSF}(x_j^{SRSF}, \mathcal{P}_j)\dots \\ &\leq V^{RSF}(B_j^{WRSF}(\mathcal{P}_{0:j-1}), \mathcal{P}_j) = B_{j+1}^{WRSF}(\mathcal{P}_{0:j})\end{aligned}$$

The second inequality follows from V^{WRSF} being a relaxation of V^{SRSF}, the third inequality follows from V^{WRSF} being increasing in the magnitude of the initial condition. The final equality follows by inspecting the definition of B_j^{WRSF} and V^{RSF}. □

Finally the following corollary, which will be needed in the proof of our main result in the next section, is immediate from the previous results.

Corollary IV.1. *If $|x(0)| \geq \eta/(1-\lambda(\mathcal{P}_0))$, $\lambda(\mathcal{P}_0) < 1$ and we apply $u_n^{SRSF}(x_n, \mathcal{P}_n)$ as a control law, then $|x_n|$ will be bounded as*

$$|x_n| \leq \eta/(1-\lambda(\mathcal{P}_0)) + \lambda(\mathcal{P}_0)^n \left|x(0)\right|$$

V. Passive Aggressive Feedback Controller

In this section we introduce a control policy that is applicable to initial uncertainty sets that are not strongly stabilizable. The controller evolves according to two stages: at first, a "passive-aggressive" feedback controller is deployed that is use as long as the consistent set $\mathcal{P}_0 \cap \mathcal{S}(x_{0:N}, u_{0:N-1})$ is not strongly stabilizable. We show that this control policy is guaranteed to shrink the initial uncertainty set to one that is strongly stabilizable once the state becomes sufficiently large. Once the uncertainty set has been reduced to a strongly stabilizable one, the controller switches to the ARSF strategy described in the previous section and drives the state to the origin.

Definition V.1. Let $\mathcal{P}_n$ be the remaining uncertainty in the system parameters after observing $x_{0:n}$ and $u_{0:n-1}$. Specifically

$$\mathcal{P}_{n+1} = \mathcal{P}_n \cap \mathcal{S}\left(x_{n:n+1}, u_n\right) \tag{45}$$

where we set $\mathcal{P}_0$ to be the initial uncertainty set.

Definition V.2. Define $k_{max}(\mathcal{P}_i)$ as the maximum deadbeat controller gain among the parameters in $\mathcal{P}_i$:

$$k_{max}(\mathcal{P}_i) := \max_{\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right] \in \mathcal{P}_i} \left|-\frac{a}{b}\right| \tag{46}$$

We begin with an intermediate result that shows if the system state is sufficiently large, then the stability margin of the uncertainty set can be reduced by an amount governed by the noise bound η and the size of the state itself. In this way, there is a notion of signal-to-noise that comes into play in the ability to learn an uncertainty set.

Theorem V.3 (Passive-Aggressive Learning)**.** *Let $\mathcal{P}_0$ be the initial (not necessarily strongly stabilizable) uncertainty set and fix positive constants $\lambda^*, p > 0$ satisfying $\lambda^* > \frac{1}{p}$. Consider the following control strategy at time-step n:*

$$u_n = k_n x_n \tag{47}$$

where

$$k_n = -\text{sign}(k_{n-1})\frac{k_{max}(\mathcal{P}_n)}{\lambda^* p - 1},\ k_{-1} = -1;$$

Then if there exists some $n_0 \geq 1$ such that $\min\{|x_{n_0-1}|, |x_{n_0-2}|\} \geq p\eta$, it holds that $\lambda(\mathcal{P}_{n_0}) \leq \lambda^*$.

Proof. Define n_{-1}, n_{-2} to be n_0-1 and n_0-2 respectively. Then,

$$\begin{aligned}\lambda(\mathcal{P}_{n_0}) &= \lambda\left(\mathcal{P}_{n_{-1}} \cap \mathcal{S}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}})\right) \\ &\leq \lambda\left(\mathcal{P}_{n_{-1}} \cap \mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}})\right) \\ &\leq k_{max}\left(\mathcal{P}_{n_{-1}}\right) \Delta_b\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \dots \\ &\quad \dots \Delta_a\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \end{aligned} \tag{48}$$

where $\mathcal{B}(\dots)$ represents the smallest outer-bounding box set of the $\mathcal{S}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}})$. Furthermore, $\Delta_b\mathcal{B}$ and $\Delta_a\mathcal{B}$ are the maximum uncertainty of parameters a and b in the set $\mathcal{B}$ as discussed in the appendix (I). Finally we note that the last inequality follows from the discussion on stability margins of boxed uncertainties in (I). Now, as $k_{n_{-2}}$ and $k_{n_{-1}}$ have opposite sign by construction, we obtain from app.(I):

$$\begin{aligned}&\Delta_b\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \\ &= \left(\frac{\eta}{|x_{n_{-2}}|} + \frac{\eta}{|x_{n_{-1}}|}\right)\frac{1}{|k_{n_{-1}}| + |k_{n_{-2}}|}\end{aligned} \tag{49}$$

$$\begin{aligned}&\Delta_a\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \\ &= \left(\frac{|k_{n_{-1}}|\eta}{|x_{n_{-2}}|} + \frac{|k_{n_{-2}}|\eta}{|x_{n_{-1}}|}\right)\frac{1}{|k_{n_{-1}}| + |k_{n_{-2}}|}\end{aligned} \tag{50}$$

Now, since by assumption we have that $\min\{|x_{n_{-1}}|, |x_{n_{-2}}|\} \geq p\eta$, we can upper bound equations (49), (50) by

$$\Delta_b\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \leq \frac{1}{p}\frac{2}{|k_{n_{-1}}| + |k_{n_{-2}}|} \tag{51}$$

$$\Delta_a\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \leq \frac{1}{p} \tag{52}$$

Furthermore, notice that $k_{max}(\mathcal{P}_{n_{-2}}) \geq k_{max}(\mathcal{P}_{n_{-1}})$ so we have

$$|k_{n_{-1}}| + |k_{n_{-2}}| \geq \frac{2k_{max}(\mathcal{P}_{n_{-1}})}{\lambda^* p - 1} \tag{53}$$

$$\Leftrightarrow \qquad \frac{2}{|k_{n_{-1}}| + |k_{n_{-2}}|} \leq \frac{\lambda^* p - 1}{k_{max}(\mathcal{P}_{n_{-1}})}, \tag{54}$$

which lets us further upper-bound (51) by

$$\Delta_b\mathcal{B}(x_{n_{-2}:n_0}, u_{n_{-2}:n_{-1}}) \leq \frac{1}{k_{max}(\mathcal{P}_{n_{-1}})}\left(\lambda^* - \frac{1}{p}\right) \tag{55}$$

Finally, plugging the bounds (55) and (52) into equation (48) gives us the desired result:

$$\begin{aligned}\lambda(\mathcal{P}_{n_0}) &\leq \frac{k_{max}(\mathcal{P}_{n_{-1}})}{k_{max}(\mathcal{P}_{n_{-1}})}\left(\lambda^* - \frac{1}{p}\right) + \frac{1}{p} \\ &\leq \lambda^*\end{aligned}$$

□

With this ability to learn the uncertainty set, we now show how a two-stage controller can lead to a stabilizing (in the BIBO sense) adaptive controller.

Theorem V.4 (Passive-Aggressive Learning and Control)**.** *Let $\mathcal{P}_0$ be an initial (not necessarily strongly stabilizable) uncertainty set, and fix positive constants $\lambda^*, p > 0$, s.t. $1 - \frac{1}{p} > \lambda^* > \frac{1}{p}$ fixed constants. Consider the following feedback control for time-step n:*

$$u_n = \begin{cases} k_n x_n & \text{If } \lambda(\mathcal{P}_n) > \lambda^* \\ \hat{u}_n^{ARSF}(x_n, \mathcal{P}_n) & \text{If } \lambda(\mathcal{P}_n) \leq \lambda^* \end{cases}$$

where

$$k_n = -\text{sign}(k_{n-1})\frac{k_{max}(\mathcal{P}_n)}{\lambda^* p - 1},\ k_{-1} = -1;$$

Then there exists at most one time-step n_0, s.t. $|x_{n_0}| > |x_{n_0-1}| > |x_{n_0-2}| > p\eta$. Furthermore, if such n_0 exists, then $\lambda(\mathcal{P}_n) \leq \lambda^*$, $\forall n$.

Proof. First, notice that if for some n^*, it holds that $|x_{n^*}| \geq p\eta$ and $\lambda(\mathcal{P}_{n^*}) \leq \lambda^*$, then by Thm.(IV.1), we have that $\forall n \geq n^*$:

$$|x_n| \leq \eta/(1-\lambda^*) + \lambda^{*(n-n^*)}\,||x(0)| - \eta/(1-\lambda^*)|$$

and since

$$1 - \frac{1}{p} > \lambda^* \Leftrightarrow 1 - \lambda^* > \frac{1}{p} \Leftrightarrow \frac{1}{1-\lambda^*}\eta \leq p\eta \tag{56}$$

we see that there cannot be a $n_0 \geq n^*$ s.t. $|x_{n_0}| > |x_{n_0-1}| > |x_{n_0-2}| > p\eta$.

Now let us assume that there exists n_0 such that $|x_{n_0}| > |x_{n_0-1}| > |x_{n_0-2}| > p\eta$. Then our previous discussion implies that $\lambda(\mathcal{P}_{n_0-1}) > \lambda^*$ and $\lambda(\mathcal{P}_{n_0-2}) > \lambda^*$ has to hold. This implies that for $n = n_0-1$ and $n = n_0-2$ the learning feedback controller (47) is being applied. Furthermore, since $|x_{n_0-1}| > |x_{n_0-2}| > p\eta$, we learn aggressively and by Thm.V.3, we obtain $\lambda(\mathcal{P}_{n_0}) \leq \lambda^*$. Finally, referring to our previous discussion, this implies that there does not exist a $n_1 \geq n_0$ s.t. $|x_{n_1}| > |x_{n_1-1}| > |x_{n_1-2}| > p\eta$, since we start applying u^{ARSF} and begin driving the state into the region $|x| \leq p\eta$. □

The "passive-aggressive" nomenclature is chosen because the policy is such that if the system parameters and process noise are stabilizing (i.e., keep the state close to the origin), then we only focus on learning the uncertainty set via the learning control policy (47) – this is the passive phase of the control policy. In particular, if for the specific process noise and system parameter realizations no such n_0 exists, then by definition we are guaranteed to have $|x_n| \leq C\eta$ for all n, where C is a constant depending only on p, λ^* and $k_{max}(\mathcal{P}_n)$ – this follows directly from noting that the state can only exceed $p\eta$ in size for at most one time step at a time. In contrast, when the noise is such that the state is pushed sufficiently far from the origin, we are able to aggressively decrease the stability margin of the uncertainty set and switch to an ARSF policy. The result is a stabilizing control scheme that is applicable to arbitrary initial uncertainty sets $\mathcal{P}_0$. Future work will explore two parallel directions: (a) adaptive selections of p and λ^* and performance bounds for the suggested approach, and (b) the injection of artificial noise into the learning step to guarantee that the state eventually exceeds $p\eta$ in magnitude.

VI. Simulation Results

In the following we show how the passive-aggressive controller performs under different scenarios. Recall, that our plant is modeled as

$$x_{n+1} = ax_n + bu_n + w_n \tag{57}$$

$$|w_n| \leq \eta \quad \begin{bmatrix} a \\ b \end{bmatrix} \in \mathcal{P}_0 \tag{58}$$

We set the initial uncertainty set to be

$$\mathcal{P}_0 = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} \middle| -3 \leq a \leq 3,\ 0.1 \leq b \leq 3 \right\} \tag{59}$$

and we pick the true parameters of the system to be

$$a_0 = 2 \qquad\qquad b_0 = 0.5. \tag{60}$$

The u^{SRSF} controller is parametrized with $p = 10$ and $\lambda^* = 0.5$. Notice that this initial uncertainty set is not strongly stabilizable. In what follows, we apply the controller described in Theorem V.4 for different noise and system parameter realizations: fixed system parameters and adversarial/random noise, adversarial system parameters and noise, and fixed parameters with no noise.

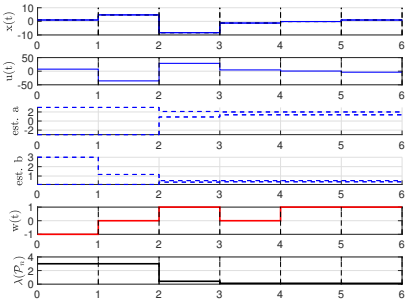
Each of the figures (**??**)-(4) show sequences of the state x_n and control action u_n and the maximum and minimum feasible a and b of the current polytope $\mathcal{P}_n$. The right subfigures overlay the area of all polytopes $\mathcal{P}_n$ and display how the uncertainty polytopes $\mathcal{P}_n$ shrink with each iteration. The shade of the polytopes becomes lighter with increasing n.

For the simulations with adversarial noise, we choose $w_k = \text{sign}(ax_k + bu_k)$ which is easily seen to be the solution to the inner maximization problem in equation (IV.2) with the surrogate function V_{RSF} to determine the adversarial noise and system parameters.

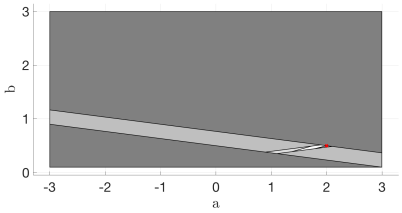
Notice that $\lambda(\mathcal{P}_n)$ decreases monotonically, and in presence of noise the controller learns to stabilize the system within two time-steps. It is also worth noticing that in presence of no noise, the controller still chooses to perturb the system on purpose to gather information, i.e. an exploration phase naturally emerges to better identify the system parameters before a robustly stabilizing control policy is applied.

VII. Conclusions and Future Work

In this paper we defined and analyzed the passive-aggressive learning and control strategy for scalar systems with bounded but adversarial process noise and parametric uncertainty. We showed that for strongly stabilizable initial uncertainty sets, sharp bounds on the state-deviation can be obtained using an ARSF control policy. We then extended these results to the general setting by proposing a two-stage controller: the first stage seeks to passively learn the system so long as the state remains sufficiently close to the origin. However, if the process and system noise are such that the state is pushed sufficiently far from the origin, the controller is able to aggressively reduce the uncertainty set to one that is strongly stabilizable, thus allowing for either the weakly or strongly ARSF policies to be applied. Future work will look to actively inject noise into the passive stage of the aforementioned two-stage control policy to expedite the learning process, as well as characterize sharp regret bounds on the proposed policy. Of additional interest is the extension of the proposed methods to the vector valued setting.



(a)



(b)

Fig. 2: $x_0 = 1$, $a_0 = 2$, $b_0 = 0.5$, $\eta = 1$, $\lambda^* = 0.5$, random noise

Appendix I
Stability Margin Bounds

A. Box-shaped Uncertainty Sets

Lemma I.1. *Let $\mathcal{B}$ be a controllable boxed uncertainty*

$$\mathcal{B} = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} \middle| \begin{array}{c} l_a \leq a \leq u_a \\ 0 < l_b \leq b \leq u_b \end{array} \right\}$$

and define $\Delta_b\mathcal{B}$, $\Delta_a\mathcal{B}$, a_{av}, b_{av}, k_{av} as

$$a_{av} = (u_a + l_a)/2 \qquad b_{av} = (u_b + l_b)/2$$
$$\Delta_a\mathcal{B} = (u_a - l_a)/2 \qquad \Delta_b\mathcal{B} = (u_b - l_b)/2$$
$$k_{av} = -\frac{a_{av}}{b_{av}}$$

where a_{av}, b_{av} is the average system of the box and k_{av} the corresponding deadbeat feedback. Then the stability margin of $\mathcal{B}$ can be computed as

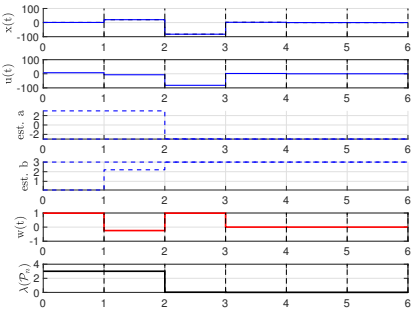
$$\lambda(\mathcal{B}) = |k_{av}|\Delta_b\mathcal{B} + \Delta_a\mathcal{B}$$

and can be interpreted geometrically as shown in Fig. (**??**).

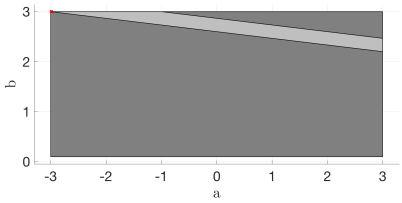
Proof. The proof is omitted but follows by solving the following optimization problem

$$\min_k \max_{\substack{l_a \leq a \leq u_a \\ l_b \leq b \leq u_b}} |a + bk|$$

□

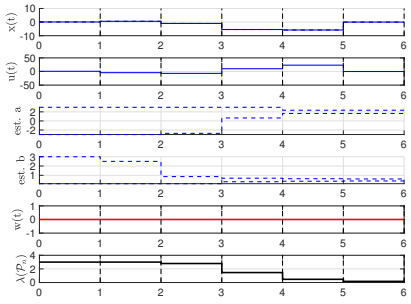


(a)

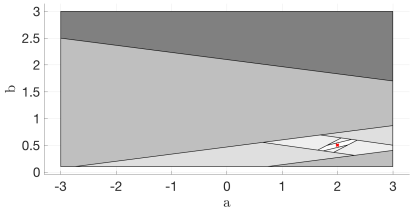


(b)

Fig. 3: $x_0 = 1$, $a_0 = 3$, $b_0 = 3$, $\eta = 1$, $\lambda^* = 0.5$, adversarial noise and system



(a)



(b)

Fig. 4: $x_0 = 0.1$, $a_0 = 2$, $b_0 = 0.5$, $\eta = 1$, $\lambda^* = 0.5$, no noise

B. Approximation for $S(x_{i:i+2}, u_{i:i+1})$ in Passive-Aggressive Learning

Consider applying $u_1 = k_1 x_1$ and $u_2 = -k_2 x_2$ as a feedback controller with $k_1 > 0$, $k_2 > 0$. Then the resulting uncertainty set $S(x_{i:i+2}, u_{i:i+1})$ resembles a parallelogram as shown in Fig.(5). An approximation of the stability margin $\lambda(S(x_{i:i+2}, u_{i:i+1}))$ is the stability margin of its outer-bounding box, i.e. $\lambda(\mathcal{B}(x_{i:i+2}, u_{i:i+1}))$. Using the notation in Fig.(5) and Lem.(I.1), the approximation can be computed as

$$\begin{aligned}&\lambda(\mathcal{B}(x_{i:i+2}, u_{i:i+1}))\\ =&k_{av}(\mathcal{B}(x_{i:i+2}, u_{i:i+1}))(x+y) + k_1 x + k_2 y\end{aligned}$$

From simple geometry, notice that the shaded area A can be computed in three ways:

$$A = h_1\sqrt{1+k_1^2}x = 2\eta\frac{x}{|x_1|} \tag{61}$$

$$= h_2\sqrt{1+k_2^2}y = 2\eta\frac{y}{|x_2|} \tag{62}$$

$$= (x+y)(k_1 x + k_2 y) - k_1 x^2 - k_2 y^2 \tag{63}$$

We can use these equations to solve for x and y and finally obtain:

$$\begin{aligned}&\lambda(\mathcal{B}(x_{i:i+2}, u_{i:i+1}))\\ =&k_{av}\left(\mathcal{B}(\dots)\right)\frac{\frac{\eta}{|x_2|}+\frac{\eta}{|x_1|}}{k_2+k_1} + \frac{k_2\frac{\eta}{|x_1|}+k_1\frac{\eta}{|x_2|}}{k_1+k_2}\end{aligned} \tag{64}$$

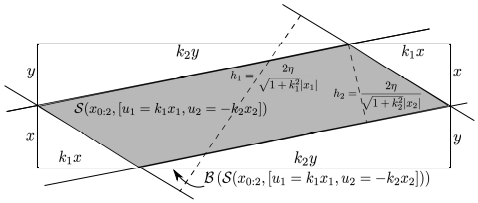


Fig. 5: Example uncertainty set after two timesteps of the passive-aggressive learner

REFERENCES

[1] A. Astolfi, D. Karagiannis, and R. Ortega, *Nonlinear and adaptive control with applications*. Springer Science & Business Media, 2007.

[2] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2013.

[3] A. Rantzer, "Concentration bounds for single parameter adaptive control," *IEEE 2018 American Control Conference, Submitted to.*, 2017.

[4] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of linear quadratic regulator," *Working draft*, 2017.

[5] C.-N. Fiechter, "Pac adaptive control of linear systems," in *Proceedings of the tenth annual conference on Computational learning theory*. ACM, 1997, pp. 72–80.

[6] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Online least squares estimation with self-normalized processes: An application to bandit problems," *arXiv preprint arXiv:1102.2670*, 2011.

[7] Y. Abbasi-Yadkori and C. Szepesvári, "Regret bounds for the adaptive control of linear quadratic systems," in *Proceedings of the 24th Annual Conference on Learning Theory*, 2011, pp. 1–26.

[8] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, 2007.

[9] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Transactions on automatic control*, vol. 58, no. 6, pp. 1497–1510, 2013.

[10] Y. Nakahira, N. Matni, and J. C. Doyle, "Hard limits on robust control over delayed and quantized communication channels with applications to sensorimotor control," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 7522–7529.