Proceedings of Machine Learning Research vol 211:1–14, 2023 5th Annual Conference on Learning for Dynamics and Control

# Online switching control with stability and regret guarantees

**Yingying Li** YINGLI2@CALTECH.EDU
**James A. Preiss** JAPREISS@CALTECH.EDU
**Na Li** NALI@SEAS.HARVARD.EDU
**Yiheng Lin** YIHENGL@CALTECH.EDU
**Adam Wierman** ADAMW@CALTECH.EDU
**Jeff Shamma** JSHAMMA@ILLINOIS.EDU

## Abstract

This paper considers online switching control with a finite candidate controller pool, an unknown dynamical system, and unknown cost functions. The candidate controllers can be unstabilizing policies. We only require at least one candidate controller to satisfy certain stability properties, but we do not know which one is stabilizing. We design an online algorithm that guarantees finite-gain stability throughout the duration of its execution. We also provide a sublinear policy regret guarantee compared with the optimal stabilizing candidate controller. Lastly, we numerically test our algorithm on quadrotor planar flights and compare it with a classical switching control algorithm, falsification-based switching, and a classical multi-armed bandit algorithm, Exp3 with batches.

## 1. Introduction

This paper considers an online switching control problem with a finite pool of candidate controllers $\{\pi_1, \ldots, \pi_N\}$, an unknown nonlinear system $x_{t+1} = f(x_t, u_t, w_t)$ with process noises $w_t$, and unknown (time-varying) cost functions $c_t(x_t, u_t)$. Notice that some candidate controllers can be unstabilizing policies. We only require at least one candidate controller to be stabilizing, but we may not know which one(s) are the stabilizing controllers.[1] We consider a single-trajectory setting, where the online switching control algorithm (also called a 'supervisor' in the literature, Hespanha et al., 2003) implements a candidate controller at each stage *without* resetting the system state. Our goal is to design an online algorithm that both stabilizes the system and optimizes the total cost among the candidate controllers.

Online switching control enjoys a long history of research, see e.g., (Hespanha et al., 2003; Stefanovic and Safonov, 2008; Al-Shyoukh and Shamma, 2009; Patil et al., 2021; Preiss and Sukhatme, 2021), and wide applications, e.g., power systems (Meng et al., 2016; Dragičević et al., 2013), healthcare (Bin et al., 2021; Marchetti et al., 2008), autonomous vehicles (Aguiar and Hespanha, 2004), Internet of Things (Zolanvari et al., 2019), etc. Online switching control is particularly useful in complex scenarios, such as when the problem has non-continuous uncertainties like unknown system orders (Liu and Yang, 2017) and hybrid systems (Garcia et al., 2013); when multiple control designs are used, for example, comparing model predictive control and PID control (Nikoofard et al., 2014); and when the controller updates are computationally demanding in real-time (Zhou and Doyle, 1998).

In the online switching control literature, most papers focus on system stabilization, and various approaches have been proposed, e.g., estimation-based supervisory control (Hespanha et al., 2003),

1. A switching control problem with at least one stabilizing candidate controller is sometimes called a 'feasible' problem in the literature (Sajjanshetty and Safonov, 2018; Stefanovic and Safonov, 2008).

performance-based falsification (Sajjanshetty and Safonov, 2018; Rosa et al., 2011; Al-Shyoukh and Shamma, 2009; Stefanovic and Safonov, 2008), multi-model adaptive control (Shahab and Miller, 2021; Kuipers and Ioannou, 2010), and others. As for the optimality analysis, most papers either only analyze convergence/asymptotic optimality, e.g., (Shahab and Miller, 2021; Kuipers and Ioannou, 2010), or discuss the optimality with respect to a cost function designed for stability purposes, instead of a cost function from "nature", e.g., (Al-Shyoukh and Shamma, 2009; Sajjanshetty and Safonov, 2018; Stefanovic and Safonov, 2008). Hence, the non-asymptotic optimality on the actual cost $\sum_t c_t(x_t, u_t)$ is largely under-explored for online switching control.

In contrast, there is rich literature in the online learning area that aims to optimize non-asymptotic performance/regret with respect to the actual cost functions (Auer et al., 2002; Arora et al., 2012). Since online switching control is closely related to online learning, especially multi-armed bandit (MAB) with memory (each candidate controller is an arm, and the current cost depends on the controllers used previously), it is tempting to leverage MAB-with-memory algorithms for online switching control (Lin et al., 2022). However, with unstabilizing candidate controllers, our problem does not satisfy the uniform bounded costs in the MAB literature (Auer et al., 2002; Arora et al., 2012; Lin et al., 2022). Further, MAB algorithms may cause *unstable* systems when some candidate controllers are unstabilizing (see, e.g., Figure 1).

Therefore, a natural question arises: *Can we design an online switching control algorithm with both stability and non-asymptotic optimality/regret guarantees on the true cost functions?*

**Contributions.** We design an online switching control algorithm Exp3-ISS by integrating an MAB algorithm Exp3 with a stability certification rule. Exp3-ISS deactivates controllers that fail the stability certification, then switches to other controllers that have not been deactivated.

Theoretically, our Exp3-ISS guarantees finite-gain stability and a sublinear policy regret when compared with the optimal stabilizing candidate controller. We prove a regret bound that scales as $\tilde{O}(N^{1/3}T^{2/3}) + \exp(O(\mathbb{M}))$, where $T$ is the horizon length, $N$ is the number of candidate controllers, and $\mathbb{M}$ is the number of candidate controllers without the desirable stability properties. Notice that $\tilde{O}(N^{1/3}T^{2/3})$ is the optimal regret for MAB with memory (Dekel et al., 2014), which suggests the optimality for our online switching control problem due to its close relation to MAB with memory. The regret $\exp(O(\mathbb{M}))$ is intuitive if candidate controllers are black boxes, in which case we must try each candidate controller at least once to determine its performance, and trying $\mathbb{M}$ unstabilizing controllers consecutively may result in exponentially large states and regrets.

Numerically, we test Exp3-ISS on quadrotor planar flight simulations and compare it with Exp3 and the falsification-based switching algorithm in (Al-Shyoukh and Shamma, 2009).

**Related work.** *Online switching control* has been studied under different names, e.g., supervisory control (Hespanha et al., 2003), logic-based switching control (Aguiar and Hespanha, 2007), and multi-model adaptive control (Kuipers and Ioannou, 2010). There are two major types of switching rules: model-estimation-based rules (Hespanha et al., 2003) and performance-based rules that do not estimate models (Al-Shyoukh and Shamma, 2009). This paper belongs to the second type.

Our stability certification is inspired by Rosa et al. (2011) and Al-Shyoukh and Shamma (2009) but is slightly different because our certification is checked at every stage, while the certification in Rosa et al. (2011); Al-Shyoukh and Shamma (2009) is only checked every $\Delta_T$ stages, where $\Delta_T$ is determined by their algorithm. The combination of a stability certification and a performance-optimization algorithm was also discussed in Rosa et al. (2011), but without optimality guarantees.

There are other stability certificates, e.g., control Lyapunov functions (Brunke et al., 2022).

*Online control and online learning.* Online control and its connection with online learning (with memory) have attracted a lot of attention recently (Wang and Boyd, 2009; Lin et al., 2022; Li et al., 2021a; Kakade et al., 2020; Boffi et al., 2021; Li et al., 2019, 2021b). Most papers consider linear systems, but there is a growing interest in nonlinear systems (Kakade et al., 2020; Boffi et al., 2021; Lin et al., 2022). This work is mostly related to (Lin et al., 2022; Arora et al., 2012; Dekel et al., 2014). However, these papers all assume uniform bounded cost functions, which corresponds to all candidate controllers being stabilizing in our case. One major contribution of this paper is to guarantee stability via a novel online control design despite unstabilizing candidate controllers.

Many online control and learning-based control papers assume to know a stabilizing policy beforehand (Lin et al., 2022; Agarwal et al., 2019; Fazel et al., 2018; Li et al., 2021a), which can be restrictive in certain applications. There is a growing interest on online (learning-based) control without prior knowledge of a stabilizing policy. This paper contributes to this area since we do not know which candidate controller is stabilizing. Besides, our result is related with Chen and Hazan (2021), which consider online linear control and provide a regret bound of $\tilde{O}(\text{poly}(d)T^{2/3}) + \exp(\text{poly}(d))$, where $d$ is the system dimension. Notice that Chen and Hazan (2021) only consider linear policies so their regret can depend on the system dimension, while our problem considers black-box controllers without restrictions or knowledge of controller structures for nonlinear systems, so our regret bound depends on the number of unstabilizing candidate controllers. It is an interesting future direction to study how to leverage controller structures in online nonlinear control to generate regret bounds that also depend on the system dimension instead of the number of controllers. Yu et al. (2022) also considers learning to stabilize a linear system but provides no regret guarantees.

*Reinforcement learning.* This work is also related to model-free reinforcement learning, especially zeroth-order policy gradient for control, which also updates policies based on observed cost performance (Fazel et al., 2018; Malik et al., 2019; Li et al., 2021c). The major difference is that we consider a finite policy pool while policy gradient considers a continuous policy pool. Further, under proper conditions, policy gradient can guarantee every selected controller updates with small enough gradient steps to be stabilizing, while our problem allows quick updates of controllers at a cost of potential encounters with unstabilizing policies.

**Notations.** $\|\cdot\|$ refers to the Euclidean norm.

## 2. Problem formulation

This paper focuses on an online supervisory/switching control problem. We consider an unknown nonlinear dynamical system $x_{t+1} = f(x_t, u_t, w_t)$ and unknown time-varying cost functions $c_t(x_t, u_t)$, with state $x_t \in \mathbb{R}^n$, action $u_t \in \mathbb{R}^m$, and process noise $w_t \in \mathbb{R}^n$. We consider a *bandit* setting, i.e., we can only observe the value of $c_t(x_t, u_t)$ after observing $x_t$ and implementing $u_t$ at stage $t$. The process noise $w_t$ is bounded by a known set $\mathcal{W} = \{w : \|w\|_2 \leq w_{\max}\}$ and can be obliviously adversarial, i.e., $w_t$ does not depend on the history states and actions. We consider a finite pool of candidate controllers

$$\{i \in \mathcal{P}_0 = \{1, \ldots, N\} : u_t = \pi_i(x_t)\}. \tag{1}$$

Some candidate controllers may not stabilize the system, and we do not know which controllers stabilize the system. Further, we treat the candidate controllers as black boxes in this paper and do not assume knowledge of their explicit forms, which is convenient for complex controllers, e.g., when the controllers are represented by neural networks. It is left as future work to consider candidate controllers with known structures.

Our goal is to design an online algorithm $\mathcal{A}$ that selects a candidate controller $I_t \in \mathcal{P}_0$ at each stage $t$ in order to both *stabilize* the system and *optimize* the total cost $J_T(\mathcal{A})$ defined below.

$$J_T(\mathcal{A}) = \sum_{t=0}^{T} c_t(x_t(\mathcal{A}), u_t(\mathcal{A})), \quad \text{where } u_t(\mathcal{A}) = \pi_{I_t}(x_t(\mathcal{A})).$$

In the supervisory control literature, this online algorithm is often called a "supervisor" (Hespanha, 2001; Hespanha et al., 2003; Tsao and Safonov, 2001; Al-Shyoukh and Shamma, 2009). We now formally introduce our assumptions and our performance metric, policy regret.

**1) Assumptions on the candidate controllers.** In our problem, we do not need all the candidate controllers to be stabilizing controllers. In fact, we only require at least one of them to satisfy desirable stability properties, which are formally introduced below.

Firstly, we consider input-to-state stability (ISS), which is commonly used in nonlinear systems with process noises $w_t$ (Sontag, 2008). Further, for the purpose of non-asymptotic analysis, we consider exponential-ISS (E-ISS) below (see e.g., Shi et al. (2021); Kolathaya et al. (2018)).

**Definition 1 (E-ISS)** *A controller $\pi$ is called exponential-ISS (E-ISS) with parameters $(\kappa, \rho, \beta)$ if, for any $x_0 \in \mathbb{R}^n$ and $\|w_t\|_2 \leq w_{\max}$ for all $t \geq 0$, the trajectory $x_{t+1} = f(x_t, \pi(x_t), w_t)$ satisfies $\|x_t\|_2 \leq \kappa\rho^t \|x_0\|_2 + \beta w_{\max}$.*[2]

In addition, we consider incremental stability ($\delta$-S), which is commonly adopted to rigorously quantify the dependence of the current states on the history (see e.g., Angeli (2002); Rüffer et al. (2013)). For the purpose of non-asymptotic analysis, we consider exponentially decaying dependence, i.e., incremental exponential stability ($\delta$-ES).

**Definition 2 ($\delta$-ES)** *A controller $\pi$ is called incrementally exponentially stable ($\delta$-ES) with parameters $(\kappa, \rho)$ if we have $\|x_t - y_t\|_2 \leq \kappa\rho^t \|x_0 - y_0\|_2$ for two trajectories $x_{t+1} = f(x_t, \pi(x_t), w_t)$ and $y_{t+1} = f(y_t, \pi(y_t), w_t)$ with any $x_0, y_0 \in \mathbb{R}^n$ and any $\|w_s\|_2 \leq w_{\max}$, $s \leq t-1$.*

**Assumption 1 (On candidate controllers)** *There exists at least one candidate controller $\pi_k$ for $k \in \mathcal{P}_0$ to satisfy Definitions 1 and 2 with parameters $(\kappa, \rho, \beta)$, which are known a priori.*[3] *Further, $\pi_i(x)$ for all $i \in \mathcal{P}_0$ are $L_\pi$-Lipschitz continuous. We define $\bar{\pi}_0$ as $\max_{i \in \mathcal{P}_0} \|\pi_i(0)\| \leq \bar{\pi}_0$.*

Notice that there are several important controller designs that satisfy Definitions 1 and 2. For example, it is straightforward to verify that stabilizing linear controllers on linear systems satisfy Definitions 1 and 2. Similarly, feedback linearization controllers on nonlinear systems also satisfy the two definitions above because the resulting closed-loop system is linear. Furthermore, Definitions 1 and 2 can be implied by exponentially incremental ISS (E$\delta$-ISS), which is commonly adopted in the online nonlinear control literature (Boffi et al., 2021; Tsukamoto et al., 2021). Besides, one can design the controller based on one stability property and verify the other stability, e.g., min-norm policy by an E-ISS control Lyapunov function can also satisfy $\delta$-ES in some cases (see (Li et al., 2023)).

The candidate controllers can be constructed by e.g., (i) domain knowledge of potentially well-performing policies, (ii) different control designs with a finite list of possible policy parameters

2. Strictly speaking, this is a relaxed version of E-ISS since we do not require exponentially decaying dependence on history disturbances as in (Shi et al., 2021).
3. For simplicity, we assume Definition 1 and 2 share the same $\kappa$, $\rho$, but our results can still hold for different parameters.

associated with each control design, (iii) listing a finite set of possible system dynamics $\mathcal{D}$ and designing controllers for this set, (iv) a combination of the methods above, etc. (see e.g., (Hespanha et al., 2003) for more discussions). For method (iii), if the true system belongs to $\mathcal{D}$ and if the controllers designed for each possible system satisfy the desirable stability properties and the Lipschitz continuity when the corresponding system is the true system, then Assumption 1 is satisfied. In practice, when the true system does not belong to $\mathcal{D}$ but is close to $\mathcal{D}$, and if the control design enjoys some robustness, our algorithm can still generate desirable numerical performance as shown in Section 5. Assumption 1 is mostly needed for theoretical analysis (see Remarks 5-6 in Section 3 for more discussions).

Lastly, Assumption 1 assumes to know the parameters $(\kappa, \beta, \rho)$ a priori, which is for simplicity and was similarly assumed in the online linear control literature (Agarwal et al., 2019; Minasyan et al., 2021). Remark 6 briefly discusses how to address the case with unknown parameters.

**2) Performance metric.** We measure the optimality performance of our online algorithm by policy regret, which compares with the optimal policy that satisfies Definitions 1 and 2.

**Definition 3 (Policy regret)** *We define* $\mathrm{PolicyRegret}(\mathcal{A}) = \mathbb{E}_{(I_t)_{t\geq 0}} J_T(\mathcal{A}) - \min_{i\in\mathcal{B}} J_T(\pi_i)$, *where the expectation is over the potentially random controller selection* $I_t$ *generated by algorithm* $\mathcal{A}$ *and* $\mathcal{B} = \{i \in \mathcal{P}_0 \mid \pi_i$ *satisfies Definitions 1 and 2 with parameters* $(\kappa, \rho, \beta).\}$.

In addition, we adopt the finite-gain stability, which is a commonly used stability measure for nonlinear systems with process noise (Sastry, 2013).

**Definition 4 (Finite-gain stability)** *For any* $1 \leq p \leq +\infty$, *a system* $x_{t+1} = f(x_t, w_t)$ *is called finite-gain* $l_p$ *stable if there exists* $0 \leq M_1, M_2 < +\infty$ *for any* $x_0, T$ *and any* $w_t \in \mathcal{W}$ *such that*

$$(\textstyle\sum_{t=0}^T \|x_t\|_2^p)^{1/p} \leq M_1(\sum_{t=0}^T \|w_t\|_2^p)^{1/p} + M_2.$$

**3) Assumptions on the dynamics and costs.** We consider Lipschitz continuous nonlinear dynamics with 0 as the equilibrium point below.

**Assumption 2 (On dynamics)** $f$ *is* $L_f$*-Lipschitz continous with respect to* $(x, u, w)$, *i.e., for any* $x, u, w, x', u', w' \in \mathbb{R}^n$ *(*$w$ *can be in the bounded region), i.e.,* $|f(x,u,w) - f(x',u',w')| \leq L_f(\|x - x'\| + \|u - u'\| + \|w - w'\|)$. *Further,* $f(0,0,0) = 0$.

We consider locally Lipschitz continuous cost functions below, which include quadratic tracking cost $(x_t - \hat{x}_t)^\top Q(x_t - \hat{x}_t) + (u_t - \hat{u}_t)^\top R(u_t - \hat{u}_t)$ with bounded $\{\hat{x}_t, \hat{u}_t\}$ as special cases.

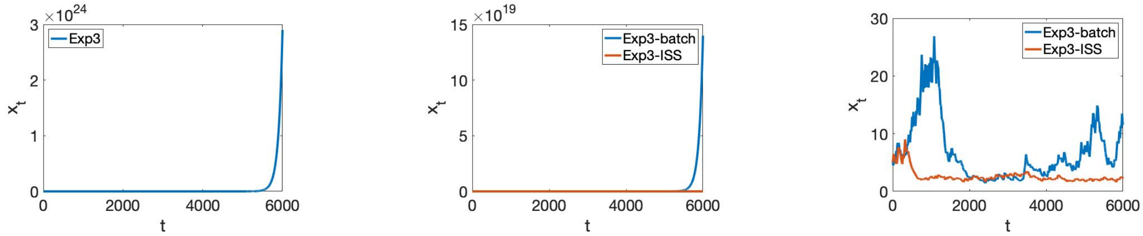
**Assumption 3 (On cost functions)** *There exists* $L_{c1}, L_{c2}$ *such that* $c_t(x,u)$ *satisfies the following inequality for any* $t$, $x, x'$, $u, u'$: $|c_t(x,u) - c_t(x',u')| \leq (L_{c1}(\max(\|x\|, \|x'\|) + \max(\|u\|, \|u'\|)) + L_{c2})(\|x - x'\| + \|u - u'\|)$. *Further, for all* $c_t(x,u)$, *there exists* $c_0 \geq 0$ *such that* $0 \leq c_t(0,0) \leq c_0$.

For the rest of this paper, we consider $\kappa \geq 1$, $\beta \geq 1$, $L_f \geq 1$, $L_\pi \geq 1$ for analytical simplicity.[4]

## 3. Algorithm design

In this section, we introduce our online algorithm for selecting candidate controllers from a controller pool that may contain unstable controllers.

4. This is without loss of generality because, if $\kappa < 1$ as an example, we can define $\kappa' = \max(\kappa, 1)$.




(a) $K_i \in [-1, -0.3, 1]$ (b) $K_i \in [-1, 0, 1]$ (c) $K_i \in [-1, -0.3, 1]$

Figure 1: Examples where Exp3 in (Auer et al., 2002) and Exp3-batch in Arora et al. (2012); Lin et al. (2022) fail to stabilize the system, in comparison to our Exp3-ISS, which stabilizes the system. Consider a system $x_{t+1} = x_t + 0.01u_t + w_t$ with $w_t$ i.i.d. generated from Uniform$[-0.3, 0.7]$. Consider candidate controllers $u_t = K_i x_t$, where $K_i$ are specified in the subfigure captions.

This problem is closely related to multi-armed bandit (MAB) with memory, by viewing each candidate controller as one arm and noticing that the cost of the current controller depends on the history of the controllers. Thus, it is tempting to apply MAB (with memory) algorithms to our problem, such as Exp3 (Auer et al., 2002) and Exp3-batch (Lin et al., 2022; Arora et al., 2012). However, it is easy to construct examples where Exp3(-batch) fails in this setting.

**Example 1 (When Exp3(-batch) fails.)** In Figure 1, we view each controller as an arm and implement Exp3 (Auer et al., 2002) and Exp3-batch (Lin et al., 2022; Arora et al., 2012), where Exp3-batch is a classical method for MAB with memory.

Figure 1(a) shows that Exp3 fails to stabilize the system even when a majority of candidate controllers are stabilizing, which is expected due to the memory-dependence of our problem. However, even with batches, Exp3 may still perform poorly, as shown in Figure 1(b-c). First, when a majority of candidate controllers do not enjoy desirable stability properties (which is exponential stability in this case), Figure 1(b) shows Exp3-batch can result in an exponential growth of states. This is because Exp3-batch is only guaranteed to work under *bounded costs* and *short memory*. However, unstabilizing candidate controllers' costs are *unbounded*, when the unstabilizing candidate controllers already steered the state $x_t$ to be very large, the stabilizing controller will also generate a large cost when implemented at stages $t, \ldots, t + \tau - 1$. In other words, the problem has *long memory* under large states. Consequently, Exp-batch may fail when there are many unstabilizing candidate controllers. Second, even when the number of unstabilizing candidate controllers is small, Exp3-batch may still perform poorly, as shown in Figure 1(c), where Exp3-batch generates large spikes in the state trajectory. This is due to explorations of unstabilizing candidates and is not rare because the cost of an unstabilizing candidate controller in one batch may not be forbiddingly large when it starts from a small initial state of this batch thanks to the stabilizing policies implemented previously. In conclusion, only adding batches to Exp3 is not enough to provide desirable stability performance for online switching control.

To handle the unstable candidate controllers, we design Exp3-ISS in Algorithm 1. In particular, we utilize Definition 1 to construct an ISS stability certificate (see Line 6 of Algorithm 1). We de-activate the controllers that fail the certificate (Line 6-8), update the controller selection probabilities $p_{j+1}$ for the active controller pool $\mathcal{P}_{j+1}$ (Line 10-12), and select a controller from the active controller pool at the start of each batch (Line 3). In this way, Exp3-ISS can stabilize the system, which is reflected in Figure 1(b-c) and will be formally proved in Theorem 9.

**Algorithm 1** Exp3-ISS

1: **Input:** $(\eta_j)_{j\geq 0}$ where $\eta_j$ is non-increasing. $\tau$. $\kappa, \rho, \beta$. $\tilde{G}_{-1}(i) = 0$ for any $i \in \mathcal{P}_0$. A uniform distribution $p_0$ defined on $\mathcal{P}_0$. $t_0 = 0$.
2: **for** Batch $j = 0, 1, 2, \ldots,$ **do**
3: Initialize $\mathcal{P}_{j+1} = \mathcal{P}_j$. Select $I_j$ from distribution $p_j$. Terminate the algorithm if $\mathcal{P}_j$ is empty.
4: **for** $t = t_j, \ldots, \min(t_j + \tau - 1, T)$ **do**
5: Implement $\pi_{I_j}$, observe $x_{t+1}$.
6: **if** $\|x_{t+1}\|_2 > \kappa\rho^{t+1-t_j}\|x_{t_j}\|_2 + \beta w_{\max}$ **then**
7: Set $\mathcal{P}_{j+1} = \mathcal{P}_j - \{I_j\}$.
8: **Break**
9: Let $t_{j+1} = t + 1$.
10: Let $g_j(I_j; I_{j-1:0}) = \frac{1}{\tau}\sum_{j=t_j}^{t_{j+1}-1} c_t(x_t, u_t)$ and $\tilde{g}_j(i; I_{j:0}) = \frac{g_j(i;I_{j-1:0})}{p_j(i)}\mathbb{1}_{(I_j=i)}$ for $i \in \mathcal{P}_{j+1}$.
11: Let $\tilde{G}_j(i; I_{j:0}) = \tilde{G}_{j-1}(i; I_{j:0}) + \tilde{g}_j(i; I_{j:0})$ for all $i \in \mathcal{P}_{j+1}$.
12: Define

$$p_{j+1}(i) = \frac{\exp\Big(-\eta_j \tilde{G}_j(i; I_{j:0})\Big)}{\sum_{k\in\mathcal{P}_{j+1}} \exp\Big(-\eta_j \tilde{G}_j(k; I_{j:0})\Big)}, \quad \forall i \in \mathcal{P}_{j+1}.$$

**Remark 5** *Notice that Algorithm 1 can be implemented as long as there exists an E-ISS candidate controller. We do not need the controller to also satisfy $\delta$-ES for implementation and for finite-gain stability in Theorem 9. This can be helpful in practice when $\delta$-ES is difficult to satisfy or verify.*

*Though Assumption 1 requires global stability properties for theoretical analysis, since our Exp3-ISS can guarantee $x_t$ to stay in a relatively small region, local stability properties within this region are already enough for successful implementation of our algorithms. This greatly extends the applicability of our algorithm and is reflected in our numerical experiments in Section 5.*

**Remark 6** *If none of the controllers in $\mathcal{P}_0$ is E-ISS, Algorithm 1 may terminate (Line 3) during implementation since it may de-activate all the controllers. If some controllers in $\mathcal{P}_0$ are E-ISS, theoretically, we can select large enough $\kappa, \beta$ and $\rho$ close to 1 to ensure at least some controllers can pass the ISS-stability certificate in Line 6 of Algorithm 1, thus avoiding early termination of the algorithm. In practice, we can also start with reasonably large $\kappa, \beta, \rho$. If all the controllers are de-activated under the current parameters, we can increase the parameters by, e.g., $\kappa \leftarrow \kappa + \Delta\kappa, \beta \leftarrow \beta + \Delta\beta$ and $\rho \leftarrow \frac{1+\rho}{2}$, then re-start Algorithm 1. If there exists an E-ISS controller in $\mathcal{P}_0$, Algorithm 1 can still guarantee stability since there will only be finite times of parameter updates. In practice, if we do not know whether there exists an E-ISS candidate controller, we can adopt additional termination rules, e.g., terminate the algorithm if the updated $\kappa, \beta, \rho$ exceed certain thresholds.*

## 4. Theoretical results

In this section, we discuss our main results, which provide stability and regret bounds for our online algorithm. For ease of reference, we introduce two useful notations below. First, we define $\mathbb{M}$ as the number of candidate controllers that do not satisfy Definition 1.

**Definition 7** *Define $\mathcal{B}_0$ as the set of controllers that do not satisfy Definition 1 under the $\kappa, \beta, \rho$ used in Algorithm 1. Let $\mathbb{M}$ denote the number of controllers in $\mathcal{B}_0$. Notice that $\mathcal{B}_0 \subseteq \mathcal{B}^c$.*

Second, we let $J$ denote the number of batches in Algorithm 1 for $T$ stages.[5] It is shown in our online supplementary material (Li et al., 2023) that $J$ is upper bounded by the following:

**Lemma 8 (Number of batches)** *In horizon $T$, the number of batches satisfies $J \leq \lceil \frac{T-\mathbb{M}}{\tau} \rceil + \mathbb{M}$.*

We are now ready to present our stability results.

**Theorem 9 (Finite-gain stability)** *When $\tau \geq \frac{\log(2\sqrt{2}\kappa)}{-\log\rho}$, Algorithm 1 is finite-gain $l_1$ stable:*

$$\sum_{t=0}^{T} \|x_t\| \leq \beta w_{\max}(T+\alpha_1 J) + \alpha_2 (L_f(1+L_\pi)\kappa)^{\mathbb{M}} \|x_0\| + \alpha_3 (L_f(1+L_\pi)\kappa)^{\mathbb{M}} (\beta w_{\max} + \bar{\pi}_0),$$

*where $\alpha_1 = \frac{\kappa}{1-\rho}\frac{1}{1-\kappa\rho^\tau}$, $\alpha_2 = \alpha_1 \frac{L_f(1+L_\pi)\kappa}{L_f(1+L_\pi)\kappa-1}$, $\alpha_3 = \alpha_2(\frac{L_f(1+L_\pi)\kappa}{1-\kappa\rho^\tau} + L_f(2+L_\pi))$. Similarly, Algorithm 1 also achieves finite gain $l_2$ stability:*

$$\sum_{t=0}^{T} \|x_t\|^2 = O((L_f(1+L_\pi)\kappa)^{2\mathbb{M}} \|x_0\|^2 + \beta^2 w_{\max}^2 (T+J) + (L_f(1+L_\pi)\kappa)^{2\mathbb{M}} (\beta^2 w_{\max}^2 + \bar{\pi}_0^2)).$$

Theorem 9 indicates that Algorithm 1 can guarantee bounded states despite unstabilizing controllers in the initial controller pool, which is in contrast with (batch-based) Exp3.

The bound in Theorem 9 scales as $O((L_f(1+L_\pi)\kappa)^{\mathbb{M}} + T)$. The exponential dependence on $\mathbb{M}$ can be intuitively explained as follows: since the candidate controllers are black boxes, we must try each controller in $\mathcal{P}_0 - \mathbb{B}_0$ at least once to de-activate them. This may result in exponential growth if we try the controllers in $\mathcal{P}_0 - \mathbb{B}_0$ consecutively and these controllers are unstable. Further, since $\mathbb{M}$ does not depend on the horizon $T$, the dependence of $\frac{1}{T}\sum_{t=0}^{T}\|x_t\|$ on $\mathbb{M}$ will diminish for large enough $T$. It is future work to consider non-black-box candidate controllers and leverage the controller structures to reduce the exponential term.

More specifically, when the number of batches $J = o(T)$, and when $T$ goes to infinity, the average $l_1$ norm of the state converges to $\frac{1}{T}\sum_{t=0}^{T}\|x_t\| \to \beta w_{\max}$. Notice that this is the same state bound achieved by implementing an E-ISS stabilizing controller defined in Definition 1 from the beginning. This suggests that, in the long run, our algorithm can almost recover the performance of the E-ISS stabilizing controllers despite testing unstabilizing controllers at the beginning.

Next, we provide a regret guarantee for our algorithm.

**Theorem 10 (Policy regret bound)** *When $\tau \geq \frac{\log(2\sqrt{2}\kappa)}{-\log\rho}$ and $\eta_j = \eta$, Exp3-ISS's regret satisfies*

$$\text{PolicyRegret} \leq \alpha_4 \eta N T + (\alpha_5 \eta N \gamma^{4\mathbb{M}} + \alpha_6 \gamma^{2\mathbb{M}}) \text{poly}(\|x_0\|, \bar{\pi}_0) + \tau \log N/\eta + \alpha_7 J$$

*where $\gamma = L_f(1+L_\pi)\kappa$, $\alpha_4, \ldots, \alpha_7$ are polynomials of $L_f, L_{c1}, L_{c2}, c_0, L_\pi, \kappa, \beta w_{\max}, \frac{1}{1-\rho}, \frac{1}{1-2^{3/4}\kappa\rho^\tau}$.*

**Corollary 11 (Regret bound order)** *Let $\eta = O(\frac{1}{N^{2/3}T^{1/3}})$ and $\tau = \max(T^{1/3}N^{-1/3}, \frac{\log(2\sqrt{2}\kappa)}{-\log\rho})$. When $T \geq N$, we have*

$$\text{PolicyRegret} \leq \tilde{O}(N^{1/3}T^{2/3}) + \exp(O(\mathbb{M})),$$

*where $\tilde{O}(\cdot)$ hides a $\log(N)$ factor.*

Corollary 11 shows the order of our regret bound under proper conditions. The first term $\tilde{O}(N^{1/3}T^{2/3})$ is common in the policy regret bound of online bandit learning with memory and has been shown to be the optimal regret order (Dekel et al., 2014). Since online control is closely related to online learning with memory, $\tilde{O}(N^{1/3}T^{2/3})$ is likely to also be the optimal regret order for our online control setting. Obtaining a formal lower bound is our ongoing work.

5. The last batch's index is $J-1$.

Notice that the exponential term $\exp(O(\mathbb{M}))$ does not depend on the horizon $T$, so for large enough $T$, our average regret bound PolicyRegret$/T$ scales as $O(1/T^{1/3})$, which diminishes to 0. This indicates that our algorithm can almost recover the optimal performance of the controllers in $\mathbb{B}$ after learning long enough. It is also worth mentioning that such an exponential term appears in other online control settings without a stabilization assumption. For example, in (Chen and Hazan, 2021), the exponential term depends on system dimensionality in a setting with linear systems and *linear* controllers, while our exponential term depends on the number of unstabilizing controllers since we do not have knowledge or restrictions on the controller structures. It is our future work to also consider controller structures to improve the exponential term for nonlinear systems.

**Proof sketch for Theorem 10.** Our proof consists of two parts: we first bound an "auxiliary regret" of our algorithm, and then bound the difference between the auxiliary regret and the policy regret.

**Lemma 12 (Auxiliary regret bound)** *Define the* auxiliary regret *of Algorithm 1 as*

$$\mathrm{AuxRegret}(\mathcal{A}) = \tau\, \mathbb{E}_{(I_j)_{j\geq 0}} \textstyle\sum_{j=0}^{J-1} g_j(I_j; I_{j-1:0}) - \min_{k\in\mathcal{B}} \tau\, \mathbb{E}_{(I_j)_{j\geq 0}} \sum_{j=0}^{J-1} g_j(k; I_{j-1:0}),$$

*where $I_j$ is seleted by algorithm $\mathcal{A}$. Under the conditions in Theorem 10, we have*

$$\mathrm{AuxRegret} \leq \alpha_4 \eta N T + \alpha_5 \eta N (L_f(1+L_\pi)\kappa)^{4\mathbb{M}} \mathrm{poly}(\|x_0\|, \bar{\pi}_0) + \tau \log N/\eta.$$

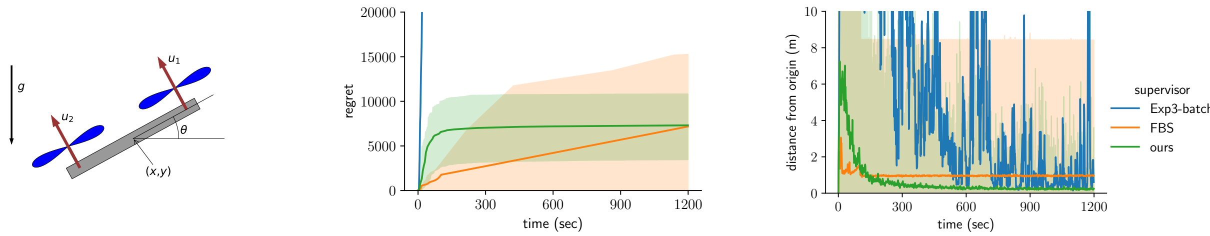
**Lemma 13 (Difference between auxiliary regret and policy regret)** *Under the conditions in Theorem 10, we have* $\mathrm{PolicyRegret} \leq \mathrm{AuxRegret} + \alpha_6 (L_f(1+L_\pi)\kappa)^{2\mathbb{M}} \mathrm{poly}(\|x_0\|, \bar{\pi}_0) + \alpha_7 J.$

The proof of Theorem 10 follows by combining the bounds in Lemma 12 and 13. The detailed proofs of the lemmas are deferred to (Li et al., 2023). We only discuss some high-level ideas below. First, auxiliary regret allows the regret benchmark to depend on the same history as that of our algorithm. It is simply called "regret" in the classical online learning setting when the cost does not depend on the history decisions. Therefore, we can borrow ideas from the regret bound proof for standard Exp3 to prove Lemma 12. However, standard Exp3 assumes uniformly bounded costs, while our problem suffers unbounded costs. To address this issue, we leverage the state bounds in Theorem 9. One technical contribution is that we bound the auxiliary regret by the bound on the total cost, $\sum_j g_j(I_j, I_{j-1:0})$, instead of the uniform bound on $g_j(I_j, I_{j-1:0})$ as in the literature (Lin et al., 2022; Arora et al., 2012). This is because the uniform bound on the cost scales as $\exp(O(\mathbb{M}))$, so directly applying this uniform bound will lead to a regret bound of order $\exp(O(\mathbb{M}))T^{2/3}$, which is much worse than our current bound $\tilde{O}(T^{2/3}) + \exp(O(\mathbb{M}))$. In fact, the uniform bound is not ideal in our case because we only suffer large states during the transient phase and enjoy small states after unstabilizing controllers are de-activated, which is also reflected in our numerical results.

Second, Lemma 13 is the only lemma that utilizes Definition 2, which establishes how fast the current state 'forgets' the history. When the current state does not depend on the history, the auxiliary regret and the policy regret are identical. Under Definition 2, the current state forgets the history exponentially fast, so by having a long enough batch size, we can bound the difference between the auxiliary regret and the policy regret. Details are in the supplementary (Li et al., 2023).

## 5. Numerical experiments

This section provides simulation results on a planar "quadrotor" illustrated in Figure 2(a) (Tedrake, 2022). We consider state $(x, y, \theta, \dot{x}, \dot{y}, \dot{\theta})$, where $(x, y)$ denotes the position and $\theta$ denotes the angle, and control inputs $(u_1, u_2)$ from the two propellers. The dynamics are $m\ddot{x} = -(u_1 +$




(a) Planar quadrotor (b) Policy regret (c) Distance trajectories

Figure 2: Comparison of Algorithm 1 with Exp3-batch in (Lin et al., 2022) and Falsification-Based Switching (FBS) in (Al-Shyoukh and Shamma, 2009) on a simulated planar quadrotor. The solid lines represent the mean value over 100 trials. The shaded regions in (b) and (c) represent the 75% percentile and the min/max over every trial, respectively.

$u_2)\sin\theta,\ m\ddot{y} = -(u_1+u_2)\cos\theta - mg,\ I\ddot{\theta} = r(u_1-u_2)$, where $m$ is the mass, $I$ is the moment of inertia, and $r$ is the arm length. Our task is to fly the quadrotor towards a target. We consider 81 proportional–derivative candidate controllers as in (Lee et al., 2010), whose parameters include gains $(k_p, k_d, k_p^\theta, k_d^\theta)$ on the position and attitude, and estimations of $m, I, r$. We consider inaccurate estimation of $m$ to test the robustness of our algorithm. More details on the setting are deferred to (Li et al., 2023) due to space limits.

Figure 2(b-c) compare our Exp3-ISS with Exp3-batch in (Lin et al., 2022) and Falsification-based Switching (FBS), which focuses on the stability and does not optimize the cost (Al-Shyoukh and Shamma, 2009). When comparing our algorithm with Exp3-batch, we can observe that Exp3-batch performs much worse than our algorithm in terms of both policy regret and the trajectories, with large spikes and fluctuations in the trajectory plot. When comparing our algorithm with FBS, we observe that, although FBS performs better than Exp3-ISS at the beginning, FBS generates a linearly increasing regret in expectation, which while Exp3-ISS enjoys regret sublinear in $T$. This is because FBS "settles" on the first stabilizing controller it identifies and does not explore to find better controllers. Therefore, unless nearly all of the controllers are unstabilizing, FBS avoids high cost of exploration at the beginning. However, since FBS essentially selects one stabilizing controller at random, linear regret is unavoidable unless FBS selects the optimal stabilizing controller by random chance. Figure 2(c) shows similar trends: though our algorithm generates larger distances at the beginning, our distances quickly diminishes to be smaller than FBS after enough exploration.

## 6. Conclusion and future directions

This paper proposes an online switching control algorithm by integrating the adversarial bandit algorithm Exp3 with a stability certification. Our algorithm stabilizes the system and provides sublinear policy regret despite the existence of unstabilizing candidate controllers. There are many interesting future directions, e.g., (i) discussing output feedback, where the stability certification in (Al-Shyoukh and Shamma, 2009) might be useful, (ii) considering an infinite or continuous policy pool by leveraging problem structure and continuity, (iii) fundamental regret and stability lower bounds for online switching control, (iv) time-varying dynamics where switching policies is necessary for stabilizing the system, (v) relaxing the global exponential stability assumptions to local and/or asymptotic stability, and (vi) combining switching-based control with estimation-based control as in multi-model adaptive control, etc.

## References

Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119. PMLR, 2019.

A Pedro Aguiar and Joao P Hespanha. Trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty. *IEEE transactions on automatic control*, 52(8):1362–1379, 2007.

A.P. Aguiar and J.P. Hespanha. Logic-based switching control for trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty. In *Proceedings of the 2004 American Control Conference*, volume 4, pages 3004–3010 vol.4, 2004. doi: 10.23919/ACC.2004.1384369.

Ibrahim Al-Shyoukh and Jeff S Shamma. Switching supervisory control using calibrated forecasts. *IEEE transactions on automatic control*, 54(4):705–716, 2009.

David Angeli. A lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control*, 47(3):410–421, 2002.

Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*, pages 1747–1754, 2012.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

Michelangelo Bin, Emanuele Crisostomi, Pietro Ferraro, Roderick Murray-Smith, Thomas Parisini, Robert Shorten, and Sebastian Stein. Hysteresis-based supervisory control with application to non-pharmaceutical containment of covid-19. *Annual reviews in control*, 52:508–522, 2021.

Nicholas M Boffi, Stephen Tu, and Jean-Jacques E Slotine. Regret bounds for adaptive nonlinear control. In *Learning for Dynamics and Control*, pages 471–483. PMLR, 2021.

Lukas Brunke, Melissa Greeff, Adam W Hall, Zhaocong Yuan, Siqi Zhou, Jacopo Panerati, and Angela P Schoellig. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, 5:411–444, 2022.

Xinyi Chen and Elad Hazan. Black-box control for linear dynamical systems. In *Conference on Learning Theory*, pages 1114–1143. PMLR, 2021.

Ofer Dekel, Jian Ding, Tomer Koren, and Yuval Peres. Bandits with switching costs: T 2/3 regret. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 459–467, 2014.

Tomislav Dragičević, Josep M Guerrero, Juan C Vasquez, and Davor Škrlec. Supervisory control of an adaptive-droop regulated dc microgrid with battery management capability. *IEEE Transactions on power Electronics*, 29(2):695–706, 2013.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1467–1476. PMLR, 2018.

Pablo Garcia, Juan P Torreglosa, Luis M Fernandez, and Francisco Jurado. Optimal energy management system for stand-alone wind turbine/photovoltaic/hydrogen/battery hybrid system with supervisory control based on fuzzy logic. *International journal of hydrogen energy*, 38(33): 14146–14158, 2013.

Joao P Hespanha. Tutorial on supervisory control. In *Lecture Notes for the workshop Control using Logic and Switching for the 40th Conf. on Decision and Contr., Orlando, Florida*, 2001.

Joao P Hespanha, Daniel Liberzon, and A Stephen Morse. Overcoming the limitations of adaptive control by means of logic-based switching. *Systems & control letters*, 49(1):49–65, 2003.

Sham Kakade, Akshay Krishnamurthy, Kendall Lowrey, Motoya Ohnishi, and Wen Sun. Information theoretic regret bounds for online nonlinear control. *Advances in Neural Information Processing Systems*, 33:15312–15325, 2020.

Shishir Kolathaya, Jacob Reher, Ayonga Hereid, and Aaron D. Ames. Input to state stabilizing control lyapunov functions for robust bipedal robotic locomotion. In *2018 Annual American Control Conference (ACC)*, pages 2224–2230, 2018. doi: 10.23919/ACC.2018.8430946.

Matthew Kuipers and Petros Ioannou. Multiple model adaptive control with mixing. *IEEE transactions on automatic control*, 55(8):1822–1836, 2010.

Taeyoung Lee, Melvin Leok, and N. Harris McClamroch. Geometric tracking control of a quadrotor UAV on SE(3). In *CDC*, pages 5420–5425. IEEE, 2010.

Yingying Li, Xin Chen, and Na Li. Online optimal control with linear dynamics and predictions: Algorithms and regret analysis. *Advances in Neural Information Processing Systems*, 32, 2019.

Yingying Li, Subhro Das, and Na Li. Online optimal control with affine constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8527–8537, 2021a.

Yingying Li, Subhro Das, Jeff Shamma, and Na Li. Safe adaptive learning-based control for constrained linear quadratic regulators with regret guarantees. *arXiv preprint arXiv:2111.00411*, 2021b.

Yingying Li, Yujie Tang, Runyu Zhang, and Na Li. Distributed reinforcement learning for decentralized linear quadratic control: A derivative-free policy optimization approach. *IEEE Transactions on Automatic Control*, 2021c.

Yingying Li, James A Preiss, Na Li, Yiheng Lin, Adam Wierman, and Jeff Shamma. Online switching control with stability and regret guarantees. *arXiv preprint arXiv:2301.08445*, 2023.

Yiheng Lin, James Preiss, Emile Anand, Yingying Li, Yisong Yue, and Adam Wierman. Online adaptive controller selection in time-varying systems: No-regret via contractive perturbations. *arXiv preprint arXiv:2210.12320*, 2022.

Liang Liu and Xuebo Yang. Robust adaptive state constraint control for uncertain switched high-order nonlinear systems. *IEEE Transactions on Industrial Electronics*, 64(10):8108–8117, 2017.

Dhruv Malik, Ashwin Pananjady, Kush Bhatia, Koulik Khamaru, Peter Bartlett, and Martin Wainwright. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. In *The 22nd international conference on artificial intelligence and statistics*, pages 2916–2925. PMLR, 2019.

Gianni Marchetti, Massimiliano Barolo, Lois Jovanovic, Howard Zisser, and Dale E Seborg. An improved pid switching control strategy for type 1 diabetes. *ieee transactions on biomedical engineering*, 55(3):857–865, 2008.

Lexuan Meng, Eleonora Riva Sanseverino, Adriana Luna, Tomislav Dragicevic, Juan C Vasquez, and Josep M Guerrero. Microgrid supervisory controllers and energy management systems: A literature review. *Renewable and Sustainable Energy Reviews*, 60:1263–1273, 2016.

Edgar Minasyan, Paula Gradu, Max Simchowitz, and Elad Hazan. Online control of unknown time-varying dynamical systems. *Advances in Neural Information Processing Systems*, 34:15934–15945, 2021.

Amirhossein Nikoofard, Tor Arne Johansen, Hessam Mahdianfar, and Alexey Pavlov. Design and comparison of constrained mpc with pid controller for heave disturbance attenuation in offshore managed pressure drilling systems. *Marine Technology Society Journal*, 48(2), 2014.

Sagar V Patil, Yu-Chen Sung, and Michael G Safonov. Unfalsified adaptive control for nonlinear time-varying plants. *IEEE Transactions on Automatic Control*, 67(8):3892–3904, 2021.

James A Preiss and Gaurav S Sukhatme. Suboptimal coverings for continuous spaces of control tasks. In *Learning for Dynamics and Control*, pages 547–558. PMLR, 2021.

Paulo Rosa, Jeff S Shamma, Carlos Silvestre, and Michael Athans. Stability overlay for adaptive control laws. *Automatica*, 47(5):1007–1014, 2011.

Björn S Rüffer, Nathan Van De Wouw, and Markus Mueller. Convergent systems vs. incremental stability. *Systems & Control Letters*, 62(3):277–285, 2013.

Kiran S Sajjanshetty and Michael G Safonov. Transient performance bounds for adaptive control. In *2018 Annual American Control Conference (ACC)*, pages 4075–4080. IEEE, 2018.

Shankar Sastry. *Nonlinear systems: analysis, stability, and control*, volume 10. Springer Science & Business Media, 2013.

Mohamad T Shahab and Daniel E Miller. Asymptotic tracking and linear-like behavior using multi-model adaptive control. *IEEE Transactions on Automatic Control*, 67(1):203–219, 2021.

Guanya Shi, Kamyar Azizzadenesheli, Michael O'Connell, Soon-Jo Chung, and Yisong Yue. Meta-adaptive nonlinear control: Theory and algorithms. *Advances in Neural Information Processing Systems*, 34:10013–10025, 2021.

Eduardo D Sontag. Input to state stability: Basic concepts and results. In *Nonlinear and optimal control theory*, pages 163–220. Springer, 2008.

Margareta Stefanovic and Michael G Safonov. Safe adaptive switching control: Stability and convergence. *IEEE Transactions on Automatic Control*, 53(9):2012–2021, 2008.

Russ Tedrake. *Underactuated Robotics*. 2022. URL https://underactuated.csail.mit.edu.

Tung-Ching Tsao and Michael G Safonov. Unfalsified direct adaptive control of a two-link robot arm. *International Journal of Adaptive Control and Signal Processing*, 15(3):319–334, 2001.

Hiroyasu Tsukamoto, Soon-Jo Chung, and Jean-Jaques E Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021.

Yang Wang and Stephen Boyd. Fast model predictive control using online optimization. *IEEE Transactions on control systems technology*, 18(2):267–278, 2009.

Jing Yu, Dimitar Ho, and Adam Wierman. Online stabilization of unknown networked systems with communication constraints. *arXiv preprint arXiv:2203.02630*, 2022.

Kemin Zhou and John Comstock Doyle. *Essentials of robust control*, volume 104. Prentice hall Upper Saddle River, NJ, 1998.

Maede Zolanvari, Marcio A Teixeira, Lav Gupta, Khaled M Khan, and Raj Jain. Machine learning-based network vulnerability analysis of industrial internet of things. *IEEE Internet of Things Journal*, 6(4):6822–6834, 2019.