Joint Stabilization and Regret Minimization through Switching in Systems with Actuator Redundancy

Jafar Abbaszadeh Chekan, Kamyar Azizzadenesheli, and Cedric Langbort

Abstract

Adaptively controlling and minimizing regret in unknown dynamical systems while controlling the growth of the system state is crucial in real-world applications. In this work, we study the problem of stabilizing and regret minimization of linear dynamical systems with system-level actuator redundancy. We propose an optimism-based algorithm that utilizes the actuator redundancy and the possibility of switching between actuating modes to guarantee the boundedness of the state. This is in contrast to the prior works that may result in an exponential (in system dimension) explosion of the state of the system. We theoretically study the rate at which our algorithm learns a stabilizing controller and prove that it achieves a regret upper bound of $O(\sqrt{T})$.

I. INTRODUCTION

Switched systems that encompass several sub-systems and a switching rule that designates which subsystem is the governing one has been widely applied in control of mechanical systems, aircraft, automotive industry, air traffic control to name but a few [1]. A widely used paradigm of switching control is systems with actuator redundancy that enables it to be actuated with different subset of actuators. This class of systems contributes to the stability improvement as well as the transient response (2 and [3]). The benefit of switching between actuation modes has been addressed in the literature of cyber-physical security ([4], [5], [6], and [7]) and fault tolerant control as well ([8], [9], and [10]).

Leveraging actuator redundancy and switching between different set of actuating modes can also improve the learning algorithm efficiency which is the goal of this paper. For a system with known dynamics, it is obvious that adding one redundant actuator results in a better performance in terms of optimality as it gives one more degree of freedom in optimal control design. However, this is not necessarily true in unknown setting as this addition increases the number of parameters to be learned. The dimension dependency of the regret minimization algorithms in the literature are mainly exponential [11], [12] and such dependence is inevitable in general [13]. Stabilizing the unknown dynamics of a control system and minimizing the regret are among the main goals of algorithm design. The authors of [12] tackle this problem by proposing the ExpOPT algorithm which deploys additional exploration in the early stages upon which we build our own algorithm. Further concern in control design of system in unknown dynamics model setting is tighter upper bound of state norm which may not be met at initial stages of algorithm deployment. State-blow-up can cause serious problems specifically in linearized dynamical systems as it may cause a large deviation from vicinity where the approximation error is bearable. Noting that this deviation can cause inefficiency in learning and even lead to a destabilizing controller. We aim to circumvent this challenge by leveraging actuator redundancy and choosing an appropriate actuating mode for early exploration phase and specifying its deployment duration. Optimism in the face of uncertainty, OFU, as a class of adaptive control algorithm, which couples estimation and control design procedures, has shown its ability to outperform the naive certainty equivalence algorithm [14]. Campi and Kumar in [14] propose a cost-biased parameter estimator which is an OFU based approach to address the optimal control problem for linear quadratic Gaussian systems with guaranteed asymptotic optimality. However, this algorithm only guarantees the convergence of average cost to that of the optimal control in limit and for finite time does not provide any bound on the measure of performance loss. Abbsi-Yadkori and Szepesvari [15] propose a learning-based algorithm to address the adaptive control design of LQ control system in finite time with worst-case regret bound of $O(\sqrt{T})$ with exponential dependence in the problem ambient dimensions. Using ridge regression estimator, they propose to construct a high-probability confidence set around the unknown parameters of the system and design an algorithm that optimistically plays with respect to this set. Along this line, many works attempt to improve the exponential dependence with further assumption, e.g. highly sparse dynamics [16], or access to stabilizing controller. Furthermore, authors in [17] propose an OFU-based learning algorithm with mild assumptions and $O(\sqrt{T})$ regret. This class of algorithms was extended in [18], [19] to LQG setting where there is only partial and noisy observations on the state of system. In addition, in [12], an algorithm with more exploration is proposed for both controllable systems. In this paper we will apply some results of this paper to controllable system class with actuator redundancy.

The remainder of the paper is organized as follows: Section III reviews the preliminaries and assumptions. Section III presents the problem statement and formulation in which we introduce the initial exploration (IExp) algorithm and stabilizing OFU algorithm (SOFUA) and discuss in detail how switching can contribute to the system stabilization and regret minimization. We carry out the regret bound analysis for (IExp+SOFUA) algorithms in the section IV. Finally,

J. Abbaszadeh Chekan and C. Langbort are with the Aerospace Engineering department and Coordinated Science Lab (CSL), University of Illinois at Urbana Champaign, IL, USA. jafar2 @illinois.edu
K. Azizadehmeshali is with the Department of Computer Science, Purdue University, Indiana, USA kamyar@purdue.edu
Section [VI] summarizes the paper’s key contributions. For the sake of brevity the proofs of the theorems and lemmas have been provided in the arxiv version of draft.

II. PRELIMINARIES

Consider a linear time invariant dynamics and its associated cost function given by:

\[ x_{t+1} = A_s x_t + B_s u_t + \omega_{t+1} \quad (1a) \]
\[ c_t = x_t^\top Q x_t + u_t^\top R u_t \quad (1b) \]

where \( x_t \) and \( u_t \) denote the state and the control input at time step \( t \), and \( A_s \in \mathbb{R}^{n \times n} \) and \( B_s \in \mathbb{R}^{n \times d} \) are the plant and input matrices which are initially unknown to the learner. \( Q \in \mathbb{R}^{n \times n} \) and \( R \in \mathbb{R}^{d \times d} \) represent known and positive definite matrices and \( \omega_{t+1} \) is a noise signal, and initial state \( x_0 \). The average cost as the result of applying control sequence \( \{u_t\}_{t \geq 0} \) is given by:

\[ J(u_0, u_1, ...) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} E_t [x_t^\top Q x_t + u_t^\top R u_t] \quad (2) \]

The problem is to design a controller such that the average cost (2) to be minimized. Let \( J^* \) denote the lowest average cost (optimal cost).

We measure the performance of a learning algorithm using the notation of regret, defined as

\[ R_T = \mathbb{E} \left[ \sum_{t=1}^{T} (x_t^\top Q x_t + u_t^\top R u_t - J^*) \right] \quad (3) \]

where control inputs are prescribed by the learning algorithm. \( R_T \) is the measure of how much the lack of insight into the model affects performance. Following [11] we assume that the system is controllable and observable. By defining \( \Theta^*_s = (A_s, B_s) \), the system transitions dynamics can be rewritten as:

\[ x_{t+1} = \Theta^*_s z_t + \omega_{t+1}, \quad z_t = \begin{pmatrix} x_t \\ u_t \end{pmatrix} \quad (4) \]

The first assumption is on the noise.

**Assumption 2.1:** There exists a filtration \( F_t \) such that

(2.1) \( z_t \) and \( x_t \) are \( F_t \)-measurable.

(2.2) for any \( t \geq 0, \)

\[ \mathbb{E}[x_{t+1}|F_t] = \Theta^*_s z_t \]

(2.3) \( \mathbb{E}[\omega_{t+1}^\top \omega_{t+1}^\top|F_t] = \sigma^2_{\omega} I_n \) for some \( \sigma^2_{\omega} > 0, \)

(2.4) \( \omega_t \) are component-wise sub-Gaussian e.g. there exists \( \sigma_{\omega} > 0 \) such that for any \( \gamma \in \mathbb{R} \) and \( j = 1, 2, ..., n \)

\[ \mathbb{E}[e^{\gamma \omega_{t+1}^j}|F_t] \leq e^{\gamma^2 \sigma^2_{\omega}/2}. \]

where \( \omega_{t+1}^j \) is the \( j \)-th element of \( \omega_{t+1} \).

In this paper, we aim to leverage actuator redundancy in control systems in order to improve the learning. When the parameters of system are known, adding a redundant actuator can improve the optimality of the designed control. However, in unknown settings, a redundant actuator means more parameters to be learned which in turn may cause a loss of efficiency. In this paper, we design an strategy through which the system can perform the learning and control task efficiently. We quantify the proposed strategy by regret bound guarantee.

Accordingly, we focus on special classes of system \( \{\mathcal{A}, \mathcal{B}\} \) in relation with the following notation. Let \( \mathcal{B}_s \) be the set of all columns, \( b_i^s \) (i \( \in \{1, ..., d\} \)), of \( \mathcal{B}_s \), which correspond to each individual actuator of the system. \( 2^{\mathcal{B}_s} \) denotes the power set of \( \mathcal{B}_s \) whose subsets, \( \mathcal{B}_s^j \)'s \( j = 1, ..., 2^d \) is constructed by combination of columns \( b_i^s \)’s.

We define the dynamics of candidate of actuating modes \( i \) as follows:

\[ x_{t+1} = A_s x_t + B_s^i u_t^i + \omega_{t+1} \quad i \in \{1, ..., |\mathcal{B}^*|\} \quad (5) \]

in which \( |\mathcal{B}^*| \) denotes the cardinality of set \( \mathcal{B}^* \) defined by

\[ \mathcal{B}^* = \{(A_s, B_s^i) | B_s^i \in 2^{\mathcal{B}_s}, (A_s, B_s^i) \text{ is controllable}\} \]

and \( B_s^i \in \mathbb{R}^{n \times d_i} \), where \( d_i \) denotes number of actuators of the mode \( i \).
Letting the mode $i$ be defined by $\Theta_{i}^{T} = (A, B_{i}^{j})$, with a slight abuse of notation, we denote the parameter matrix of mode $1$, encompassing all actuators, by $\Theta$, (i.e. $\Theta_{1}^{T} = \Theta$).

The cost function has a quadratic form of (11). The matrices $A_{i}$ and $B_{i}$ are unknown to the learner that aims to learn and control the system simultaneously while switching between a number of actuating modes. These actuation modes are constructed leveraging the redundancy of actuators.

For each actuating mode $i$ defined by the pair $\Theta_{i}^{T} = (A, B_{i}^{j})$, where $B_{i}^{j} \in B^{j}$ solving the DARE,

$$P(\Theta_{i}^{T}) = Q + A^{T} P(\Theta_{i}^{T}) A - A^{T} P(\Theta_{i}^{T}) B_{i}^{j} (B_{i}^{jT} P(\Theta_{i}^{T}) B_{i}^{j} + R_{i})^{-1} B_{i}^{jT} P(\Theta_{i}^{T}) A.$$  

(6)

where $R_{i} \in \mathbb{R}^{d_{i} \times d_{i}}$ is input weight matrix of actuating mode $i$.

Under controllability condition, the linear optimal control law $u^{i}(t) = K(P(\Theta_{i}^{T})) x(t)$, where,

$$K(\Theta_{i}^{T}) = -(B_{i}^{jT} P(\Theta_{i}^{T}) B_{i}^{j} + R_{i})^{-1} B_{i}^{jT} P(\Theta_{i}^{T}) A.$$  

is stabilizing and the average cost of control law with $\Theta = \Theta_{i}$ is the optimal average cost $J(\Theta_{i}^{T}) = \text{trace}(P(\Theta_{i}^{T}))$.

We reformulated the problem $\text{5}$ in the following way:

$$x_{t+1} = A_{i} x_{t} + B_{i}^{jT} \bar{u}_{t} + \omega_{t+1}$$  

(7)

in which $\bar{u} = \bar{K}_{i} x$. We need the following construction before defining $\bar{K}_{i}$. Recalling that the matrix $B_{i}^{j}$ is constructed by the subset columns $b^{j}$’s ($j \in \bar{A} \subset \{1, ..., m\}$), we define $\bar{K}_{i}$ as a $d_{i} \times n$ matrix whose rows $j \in \bar{A}$ are rows of $K_{i}$ by order and zero elsewhere.

Having the set of candidate switching modes $B_{\text{can}} \subseteq B$, the following assumption is the other core assumption of our analysis which holds for all actuating modes.

**Assumption 2.2:** Let the sets $S_{0}^{i}$ and $S_{1}^{i}$ be defined as follows:

$$S_{0}^{i} = \{ \Theta_{i}^{T} \in R^{n \times (n+d_{i})} \mid \text{trace}(\Theta_{i}^{T} \Theta_{i}^{T}) \leq s^{i} \}$$

$$S_{1}^{i} = \{ (A, B_{i}^{j}) \in R^{n \times (n+d_{i})} \mid (A, B^{j}) \in B_{\text{can}} \} \text{ where } Q = M^{T} M, \text{ and } \|A + B_{i}^{j} K_{i}\| \leq \Upsilon_{i} < 1$$

Then, given the information $s^{i}$ and $\Upsilon_{i}$, $\Theta_{i}^{T} \forall i \in B_{\text{can}}$ belongs to the intersection of these sets, i.e. $\Theta_{i}^{T} \in S^{i}$ where $S^{i} \subseteq S_{0}^{i} \cap S_{1}^{i}$. By slightly abusing the notation for the full actuation mode mode $i = 1$, we simply let $\Upsilon_{1} = \Upsilon$, $s^{1} = s$, and $S^{1} = S$. It is obvious that $s^{i} \leq s \forall i \in B_{\text{can}}$.

In addition, boundedness of $S^{i}$’s results in boundedness of $P(\Theta_{i}^{T})$ with a finite constant $D^{i}$ [20]:

$$D^{i} = \sup \{ \|P(\Theta_{i}^{T})\| \mid \Theta_{i} \in S^{i} \},$$  

(8)

then we define $D = \max_{i \in B^{i}} D^{i}$. And $\exists \kappa > 1$ such that

$$\kappa^{i} = \sup \{ \|K(\Theta_{i}^{T})\| \mid \Theta_{i} \in S^{i} \}.$$  

(9)

In the proposed strategy we need some side information from system parameters for all the actuating modes which are summarized in the following assumption.

**Assumption 2.3:** (Side Information) There are known positive constants $\Upsilon_{i}$, $\eta_{i}$, $\vartheta_{i}$, and $\gamma^{i}$ $\forall i \in \{1, ..., |B_{\text{can}}|\}$ such that

$$\sup_{\Theta_{i} \in S^{i}} \|A_{i}^{j} + B_{i}^{jT} K(\Theta_{i}^{T})\| \leq \eta_{i},$$  

(10)

and

$$\|B_{i}^{j}\| \leq \vartheta_{i},$$  

(11)

and

$$J_{\star}(\Theta_{i}^{T}) - J_{\star}(\Theta_{i}) \leq \gamma^{i}.$$  

(12)

At first glance the assumption (12) may seem restrictive, however, it is not true as we do not use this side information in the control design procedure and it is merely used for regret bound analysis.

**III. PROBLEM STATEMENT AND FORMULATION**

In this section, we propose an ExpOPT ([12]) based algorithm that leverages actuator redundancy in the "more exploration" step to avoid blow up in the state norm while minimizing a regret bound. We break down the strategy into two phases of initial exploration (Exp) and optimism (Opt) and build upon the recent results regarding the existence of a stabilizing neighborhood around the the system parameters. Reaching such a set is one of the main goals of initial phase of the algorithm.
we can show that there exists a stabilizing neighborhood around the system parameters that any controller designed using Algorithm 1

\begin{algorithm}
\begin{enumerate}
\item \textbf{Inputs:} $T^\text{c}_e > T^\text{c}_e^*, S > 0, \delta > 0, Q, R, L, \sigma, \lambda > 0$
\item Set actuation mode $\mathcal{M} = i^*$
\item set $V^i_t = \lambda I, \hat{\Theta}^i = 0$
\item $\hat{\Theta}^i_0 = \arg\min_{\Theta^i \in C^i_{\hat{\Theta}_0^i} \cap \mathbb{S}^i} J(\Theta^i)$
\item for $t = 0, 1, \ldots, T^\text{c}_e$
\item \hspace{0.5cm} if $\det(V^i_t) > \delta \det(V^i_T)$ or $t = 0$
\item \hspace{1cm} Calculate $\hat{\Theta}^i_t$ by (14) and set $\tau = t$
\item \hspace{1cm} Find $\hat{\Theta}^i_t$ s.t. $J(\hat{\Theta}^i_t) \leq \inf_{\Theta^i \in C^i_{\hat{\Theta}_0^i} \cap \mathbb{S}^i} J(\Theta^i) + \frac{1}{\sqrt{t}}$
\item \hspace{0.5cm} else
\item \hspace{1cm} $\hat{\Theta}^i_t = \hat{\Theta}^i_{t-1}$
\item \hspace{1cm} end if
\item \hspace{0.5cm} end for
\item \hspace{0.5cm} end for
\item \hspace{0.5cm} Return $V_{T^\text{c}_e^*+1}$ and corresponding $C^i_{T^\text{c}_e^*}$
\end{enumerate}
\end{algorithm}

A. The initial exploration phase

Before describing our strategy and the corresponding algorithms, we deem it necessary to go through a few important points. Given the fact that for a controllable system, discrete algebraic Riccati equation has unique positive definite solution, we can show that there exists a stabilizing neighborhood around the system parameters that any controller designed using parameters in that neighborhood stabilizes the system.

In order to reach this neighborhood, we need to explore the system for some time. However, during the exploration phase the state of system may blow up. In [12] it is shown that the state norm is upper-bounded by $c(n + d)(n + d)$ with a problem dependent parameter $c$, which emphasizes the role of the number of actuators $d$ on state norm during initial exploration phase. For systems with redundant actuation such as those described in Section II, it is thus natural to think of performing this exploration phase using a single actuator. The question, then, is which one to choose to guarantee the smallest blow-up of state (instability). The duration of the exploration phase, $T_e$, that guarantees reaching the stabilizing neighborhood is the other important concern which should be addressed. Before answering these questions, we provide a brief explanation of the Algorithm 1 which assumes that exploration actuation mode $i^*$ and $T^\text{c}_e$ have been determined.

The Initial exploration uses the self-normalized process to construct a high probability confidence set around the true parameters of the system. By receiving new measurement $z_t$ the least square estimation error, $e(\Theta^i | z_t)$, is obtained as,

$$
e(\Theta^i | z_t) = \lambda \text{Tr}(\Theta^i \Theta^i \text{T}) + \sum_{s=0}^{t-1} \text{Tr}((x_{s+1} - \Theta^i \text{T} z^i - s_0)(x_{s+1} - \Theta^i \text{T} z^i - s_0 \text{T})) \tag{13}$$

with regularization parameter $\lambda > 0$. This yields the $l^2$-regularized least square estimate:

$$
\hat{\Theta}^i_t = \arg\min_{\Theta^i} e(\Theta^i | z_t) = (Z^i \text{T} Z^i + \lambda I)^{-1} Z^i \text{T} X. \tag{14}
$$

where $Z^i$ and $X$ are matrices whose rows are $z^i, \ldots, z_{t-1} \text{T}$ and $x^i, \ldots, x_t \text{T}$. Defining covariance matrix $V^i_t$ as follows:

$$
V^i_t = \lambda I + \sum_{s=0}^{t-1} z^i \text{T} z^i = \lambda I + Z^i \text{T} Z^i
$$
[11] shows that with probability at least \(1 - \delta\), where \(0 < \delta < 1\), the true parameters of system \(\Theta_t\) belongs to the confidence set defined by:

\[
C_t^* (\delta) = \{ \Theta^T \in \mathbb{R}^{n \times (n+d)} | \quad \text{Tr}((\hat{\Theta}_t^* - \Theta^*)^T V_t^*(\hat{\Theta}_t^* - \Theta^*)) \leq \beta_t^* (\delta) \}
\]

where

\[
\beta_t^* (\delta) = (\sigma \omega) \sqrt{2 n \log \left( \frac{\det(V_t^*)^{-1/2} \det(\lambda I)^{-1/2}}{\delta} \right) + \lambda^{1/2} S^2}
\]

(15)

After finding high-probability confidence sets for the unknown parameter, the core step of algorithm proposed in [11] is implementing Optimism in the Face of Uncertainty (OFU) principle. At any time \(t\), we need to lower bound the smallest eigenvalue of covariance matrix \(V_t\) to determine whether updates to the control policy are needed where \(\tau\) is the last time of policy update.

As can be seen in the regret bound analysis of [11], recurrent updates in the policy may worsen the performance, so a criterion is needed to prevent frequent policy switches. A such, at each time step \(t\) the algorithm checks the condition \(\det(V_t^*) > 2 \det(V_t^*)\) to determine whether updates to the control policy are needed where \(\tau\) is the last time of policy update.

Note that in (15) we have not specified any structure for the control input used to construct the extended state \(z_t\). The formulation above holds for any actuation mode, being mindful that the dimension of the covariance matrix changes. Output of Algorithm 1 is a confidence set around the parameters of full-actuation mode which is defined in a similar way as follows:

\[
C_t(\delta) = \{ \Theta^T \in \mathbb{R}^{n \times (n+d)} | \quad \text{Tr}((\hat{\Theta}_t - \Theta)^T V_t(\hat{\Theta}_t - \Theta)) \leq \beta_t (\delta) \}
\]

(18)

where

\[
\beta_t (\delta) = (\sigma \omega) \sqrt{2 n \log \left( \frac{\det(V_t)}{\delta} \right) + \lambda^{1/2} S^2}.
\]

(19)

The control input for the exploration phase is written as \(\tilde{u}_t = K(\hat{\Theta}_t^*) x_t + \nu_t\) where \(K(\hat{\Theta}_t^*) x_t\) corresponds to the optimism part of strategy (\(K\) is constructed using \(K\)) and \(\nu_t \sim \mathcal{N}(\mu, \sigma_p^2 I) \in \mathbb{R}^d\) with \(\sigma_p^2 = 2n^2 \sigma_{\nu}^2\) is random exploration term to be characterized. This additional term aims at finding stabilizing neighborhood faster.

Given the constructed central confidence set, we aim to specify the time duration \(T^*_c\) that guarantees the parameter estimate is within stabilizing neighborhood. For this, we need to lower bound the smallest eigenvalue of co-variance matrix \(V_t\). The following lemma adapted from [12], named persistence of excitation during the extra exploration, provides this lower bound.

**Lemma 3.1:** For the initial exploration period of \(T_\omega \geq \frac{4\rho^2}{\sigma_p^2} \log (12/\delta)\) we have

\[
\lambda_{\min}(V_{\omega}) \geq \sigma_r^2 T_{\omega}
\]

(20)

with probability at least \(1 - \delta\) where \(\sigma_r^2 = \frac{\sigma_p^2}{10} + \sigma_p^2\), \(c_p := \min\{c_p, c''_p\}\), and

\[
c_p = \frac{\bar{\sigma}_r^2 - \sigma_r^2 - 4\sigma_r^2 (1 + \frac{\bar{\sigma}_r^2}{2\sigma_r^2}) \exp(-\frac{\bar{\sigma}_r^2}{\sigma_r^2})}{\bar{\sigma}_r^2}
\]

\[
c''_p = \frac{\bar{\sigma}_r^2 - 4\sigma_r^2 (1 + \frac{\bar{\sigma}_r^2}{2\sigma_r^2}) \exp(-\frac{\bar{\sigma}_r^2}{\sigma_r^2}) - 0.5 \sigma_r^2 \exp(-\frac{3\bar{\sigma}_r^2}{2\sigma_r^2})}{\bar{\sigma}_r^2}
\]

for any \(\sigma_r^2 \leq \bar{\sigma}_r^2\) and \(\bar{\sigma}_r^2\) such that \(c_p, c''_p > 0\).

The following lemma gives an upper-bound for the parameter estimation error at the end of time \(T\) which will be used to compute the minimum extra exploitation time \(T_\omega\).

**Lemma 3.2:** Suppose assumption 1 and 2 holds. For \(T \geq \text{poly} (\sigma_r^2, \bar{\sigma}_r^2, n \log (1/\delta))\) having additional exploration leads to

\[
||\hat{\Theta}_T - \Theta_*||_2 \leq \frac{1}{\sigma_r \sqrt{T}} \left( \sigma \omega \sqrt{2 n \log \left( \frac{\det(V_T)^{1/2}}{\delta \det(\lambda I)^{1/2}} + \sqrt{\lambda S} \right) } \right)
\]

(21)
Proof: The proof is straightforward. First, a confidence set around the true but unknown parameters of the system is obtained which is given by (18) and (19). Then applying (20) given by Lemma 3.1 completes the proof.

There is one more step to obtain the extra exploration duration, $T_\omega$, which is obtaining an upper-bound for the right hand side of (21). Performing such a step allows us to state the following central result.

**Lemma 3.3:** At the end of initial exploration, for any mode $\forall i \in B_*$ the following inequality holds

$$||\hat{\Theta}_{T_\omega} - \Theta_*||_2 \leq \frac{\kappa_i^w}{\sqrt{T_\omega}}$$  \hspace{1cm} (22)

where

$$\kappa_i^w := \frac{\sigma_\omega}{\sigma_*}$$

$$\sqrt{n(n+d) \log (1 + \frac{\mathcal{P}}{\lambda(n+d)} + 2n \log \frac{1}{\delta} + \sqrt{\lambda S}} \hspace{5cm} \text{(23)}$$

and,

$$\mathcal{P} := X_{T_\omega}^2 (1 + \kappa^2) T_\omega + 4\sigma_\omega^2 d \log(4nT_\omega(T_\omega + 1)/\delta)$$

in which $d_i$ denotes the number of actuators of the mode $i$ and $T_\omega$ stands for initial exploration duration of the mode $i$. Furthermore, if we define

$$T_c := \frac{4(1 + \kappa)^2 \kappa^2}{(1 - \bar{Y})^2}$$  \hspace{1cm} (24)

then for $T_\omega > T_c$, $||\hat{\Theta}_{T_\omega} - \Theta_*||_2 \leq \frac{1 - Y}{2(1 + \kappa)\bar{Y}}$ holds with probability at least $1 - 2\delta$.

The proof is provided in Appendix [VII].

**B. Determining the optimal mode for IExp**

We still need to specify the best actuating mode $i^*$ for initial exploration along with its corresponding upperbound $X_t^{i*}$. To this aim, we need following lemma adapted from [11].

**Lemma 3.4:** ([11]) For any $0 \leq t \leq T$ and $\forall i \in \{1, ..., |B^*|\}$ we have that

$$\max_{s \leq t, s \notin \tau_i} ||((\Theta_* - \hat{\Theta}_s)^T z_s^i)|| \leq GZ^{n+d_i}/(\bar{Y}) \beta_i^T(\delta/4)^{2(n+d_i + 1)}$$  \hspace{1cm} (25)

where $\tau_i$ is a set of finite number of times, with maximum cardinality $n + d_i$, occurring between any time interval $[0, t]$ such that $||((\Theta_* - \hat{\Theta}_s)^T z_s^i)||$ is not well controlled in those instances (see Lemma 17 and 18 in [11] for more details).

For the system actuating in the mode $i$ the state recursion can be written as follows:

$$x_{t+1} = \Gamma_t x_t + r_t$$  \hspace{1cm} (26)

where,

$$\Gamma_{t+1}^i = \begin{cases} \hat{A}_i + \hat{B}_i K_i(\hat{\Theta}_t) & t \notin \tau_i \\ A_i + B_i K_i(\hat{\Theta}_t) & t \in \tau_i \\ \end{cases}$$  \hspace{1cm} (27)

and

$$x_{t+1}^i = \begin{cases} M_i^i z_t + B_i \nu_t + \omega_{t+1} & t \notin \tau_i \\ \omega_{t+1} + B_i \nu_t & t \in \tau_i \\ \end{cases}$$  \hspace{1cm} (28)

where $M_i^i = (\Theta_* - \hat{\Theta}_t) z_t$. By propagating the state back to time step zero, the state update equation can be written as:

$$x_t = \prod_{s=0}^{t-1} \Gamma_s x_0 + \sum_{k=1}^{t-1} \left( \prod_{s=k}^{t-1} \Gamma_s^i \right) r_k$$  \hspace{1cm} (29)

Assuming $x_0 = 0$ and applying the side information given in Assumption 2.3, i.e. (10) and assuming $||B_i||\sigma_\nu$ and $\sigma_\omega$ are independent, then one can upper-bound the state norm as follows:

$$||x_t^i|| \leq \frac{1}{1 - \bar{Y}} (\frac{\eta}{\bar{Y}})^{n+d_i} \left[ \frac{G_i z_t^{n+d_i}}{\lambda(\eta/\bar{Y})^{n+d_i} + 1} \beta_i^T(\delta/4)^{2(n+d_i + 1)} \right]$$

$$\leq \frac{1}{1 - \bar{Y}} \left( \frac{\eta}{\bar{Y}} \right)^{n+d_i} \left[ 2n \log \frac{1}{\delta} \right] =: \alpha_i^T, \ \forall i \in \{1, ..., |B^*|\}.$$

$$|||B_i||\sigma_\nu + \sigma_\omega|| \sqrt{2n \log \frac{1}{\delta}} =: \alpha_i^T, \ \forall i \in \{1, ..., |B^*|\}.$$  \hspace{1cm} (30)
normalization where \(d_i\) stands for the number of actuators of an actuating mode \(i\) and similarly any subscripts and superscripts \(i\) denotes the actuating mode \(i\).

The policy explicated in Algorithm 1 keeps the states of the underlying system bounded with the probability at least \(1 - \delta\) during initial exploration which is defined as the "good event" \(F^i_t\)

\[
F^i_t = \{ \omega \in \Omega \mid \forall s \leq T^i_c, \|x^i_s\| \leq \alpha^i_t \}. \tag{31}
\]

A second "good event" is associated with the central confidence set defined as:

\[
E_i = \{ \omega \in \Omega \mid \forall s \leq t, \Theta_s \in C_s(\delta/4) \} \tag{32}
\]

Given the side information \(T_c\) and \(\eta_s\) for all actuating modes \(i \in \{1, \ldots, |B^*|\}\), using the bound (30), we aim to find an actuating mode \(i^*\) that provides the lowest possible upper bound of state at first phase. This guarantees the state norm does not blow-up while minimizing the regret. The following lemma specifies the best mode \(i^*\) to reach this goal as well as the corresponding upper bound \(X^i_{t^*,c}\) already used in (33).

**Theorem 3.1:** For any \(0 \leq t \leq T^i_c\), when the system actuates in the mode \(i \in B^*\), it holds that

1) \(I_{F^i_t} \max_{1 \leq s \leq t} \|x^i_s\| \leq X^i_t\), where

\[
X^i_t = Y^i_{t,\nu} + \bar{X}^i = Y^i_{t,\nu} + \bar{X}^i \tag{33}
\]

where

\[
Y^i_{t,\nu} := \max \left( e, \lambda(n + d^i)(e - 1), \left( -\bar{F}^i + \sqrt{\bar{F}^2 + \bar{L}^i} \right) / (2\bar{L}^i) \right), \tag{34}
\]

with

\[
\bar{F}^i = D_1^i \left( (n + d^i) n\sigma^\nu + \sqrt{\lambda}S \right) \log t + D_2^i \sqrt{\log t / \delta} + \frac{D_1^i (n + d^i) n\sigma^\nu \log (1 + C^2 t)}{(n + d^i) \lambda},
\]

in which

\[
D_1^i := \frac{1}{1 - Y^i_t} (\eta^i_t)^{n + d^i} \bar{G}^i,
\]

\[
D_2^i := \frac{\sigma^\nu}{1 - Y^i_t} (\eta^i_t)^{n + d^i} (\| B^* \| \sigma^\nu + \sigma^\nu)
\]

And \(\bar{L}^i = \bar{K}^i(n + d^i + 1)\) where

\[
\bar{K}^i = 2D_1^i (n + d^i) n\sigma^\nu.
\]

2) The best actuating mode \(i^*\) for initial exploration is,

\[
i^* = \arg\min_{i \in B^*} Y^i_{t,\nu} + \bar{X}^i \tag{35}
\]

3) The upper-bound of state norm of system actuating in the mode \(i^*\) during initial exploration phase can be written as follows:

\[
\|X^i_{t,c}\| \leq c^{i^*} (n + d^i)^{n + d^{i^*}} \tag{36}
\]

for some system parameters dependent constant \(c^{i^*}\).

The proof can be found in Appendix VII.

**Remark 3.1:** While optimization problem (35) cannot be solved analytically because \(T^i_c\) itself depends on \(X^i_{t^*,c}\), it can be determined using the Algorithm 2.

After running the IExp algorithm for \(t \leq T^i_{t^*}\) noting that the confidence set is tight enough and we are in the stabilizing region, Algorithm 3 which leverages all the actuators comes into play. This algorithm has the central confidence set given by the Algorithm 1 as an input. If the event \(E_1\) holds, Algorithm 3 keeps the states of the underlying system bounded with probability \(1 - \delta\) in optimization phase (second phase) which is defined as the "good event" \(F^i_t\) for \(t > T^i_{t^*}\):

\[
F^i_{t^*} = \{ \omega \in \Omega \mid \forall 0 \leq t, \|x^i_s\| \leq X^i_{t^*,c} \}. \tag{37}
\]

where

\[
X^i_{t^*} = \frac{32n\sigma^2 \sqrt{n + \kappa^2}}{(1 - \delta)^2} \log \frac{n(T - T^i_{t^*})}{\delta}. \tag{38}
\]

The upper-bound (38) is simply obtained from propagating the linear dynamics

\[
x_{t+1} = (A^*_c + B^* \bar{K}(\bar{G})) x_t + \omega_t = M_c x_t + \omega_t \tag{39}
\]
Algorithm 2 Find best actuating mode $i^*$ and its corresponding $T_{c^*}^i$

1: \textbf{Inputs}: $\lambda, \kappa, S > 0, \delta > 0, Q, \sigma_\omega, \vartheta_i, \eta_i, \tilde{T}_i \forall i \in B_s$
2: \textbf{for} $\forall i \in B_s$ \textbf{do}
3: \hspace{1em} $T_{itr}^i = 1$
4: \hspace{1em} \textbf{for} $t = 1, 2, ..., T_{itr}^i$ \textbf{do}
5: \hspace{2em} compute $\kappa_c^i$ by (23)
6: \hspace{2em} compute $T_c^i$ by (24)
7: \hspace{2em} \textbf{if} $t < T_c^i$ \textbf{then}
8: \hspace{3em} $T_{itr}^i = T_{itr}^i + 1$
9: \hspace{2em} \textbf{else}
10: \hspace{3em} $T_{itr}^i = t$
11: \hspace{1em} \textbf{end if}
12: \hspace{1em} \textbf{end for}
13: \textbf{end for}
14: Compute $X_{T_0}^i = Y_{i,T_0}^{n+d_i+1} \forall i \in B_s$
15: Solve $i^* = \arg\min_{i \in B_s} X_{T_0}^{i^*}$
16: \textbf{Outputs}: $i^*$ and $T_{c^*}^i \approx T_c^i$

where

$$M_t = (A_s - \tilde{A}_{t-1} + \tilde{A}_{t-1} + B_s \tilde{K}(\tilde{\Theta}_{t-1}) + \tilde{B}_{t-1} \ast K(\tilde{\Theta}_{t-1}) - \tilde{B}_{t-1} \ast K(\tilde{\Theta}_{t-1}))$$

is the closed-loop dynamics and $x_{T_0}^{i^*}$ is initial state. With controllability assumption for the $t > T_c^{i^*}$, if the event $E_t$ holds then $\|M_t\| < \frac{1 + \Upsilon}{2} < 1$ for all $t \geq T_c^{i^*}$. This results in

$$\|x_t\| \leq \left(\frac{1 + \Upsilon}{2}\right)^{t-T_c^{i^*}} \|x_{T_c^{i^*}}\| + \frac{2}{1 - \Upsilon} \max_{T_c^{i^*} \leq s \leq t} \|z_s\|. \tag{40}$$

By applying union bound argument on the second term from right hand side of (40) and using the bound (36), it is straight forward to show that

$$\|x_t\| \leq \left(\frac{1 + \Upsilon}{2}\right)^{t-T_c^{i^*}} c^{i^*} (n + d_i)^{n+d_i} + \chi_s$$

where

$$\chi_s := \frac{2\sigma_\omega}{1 - \Upsilon} \sqrt{2n \log \frac{n(T - T_c^{i^*})}{\delta}}. \tag{41}$$

For $t > T_c^{i^*} + \frac{(n+d_i) \log(n+d_i) + \log c^{i^*} - \log \chi_s}{\log \frac{1}{T_c^{i^*} - T_c^{j^*}}} := T_{cc}$ we have $\|x_t\| \leq 2\chi_s$ that gives (38) (See [12] for details).

Algorithm 3 Stabilizing OFU Algorithm (SOFUA)

1: \textbf{Inputs}: $T, S > 0, \delta > 0, Q, R, L, V_{T_c}, C_{T_c}, \Theta_{T_c}$
2: $\hat{\Theta}_T^* = \arg\min_{\Theta \in C_{T_c}^*} J(\Theta)$
3: \textbf{for} $t = T_c^{i^*}, T_c^{i^*} + 1, T_c^{i^*} + 2, ...$ \textbf{do}
4: \hspace{1em} if $\det(V_t) > 2 \det(V_{T_c})$ or $t = T_c^{i^*}$ then
5: \hspace{2em} Calculate $\Theta_t$ by (13) and set $\tau = t$
6: \hspace{2em} Find $\Theta_t$ such that $J(\Theta_t) \leq \inf_{\Theta \in C_t(\delta) \cap S} J(\Theta) + \frac{1}{\sqrt{t}}$
7: \hspace{1em} \textbf{else}
8: \hspace{2em} $\hat{\Theta}_t = \hat{\Theta}_{t-1}$
9: \hspace{1em} \textbf{end if}
10: For the parameter $\hat{\Theta}_t$ solve Ricatti and calculate $u_t = K(\hat{\Theta}_t)x_t$
11: Apply the control and observe new state $x_{t+1}$
12: Save $(z_t, x_{t+1})$ into dataset
13: $V_{t+1} = V_t + z_t z_t^T$
14: \textbf{end for}
IV. REGRET BOUND ANALYSIS

The decomposition of IExp+SOFUA algorithm regret is similar to the one provided by [12] but in upper-bounding it we have an additional term. Adapted from [12] for the regret decomposition we have:

\[
R(T) = \sum_{t=0}^{T} \left( x_t^T Q x_t + u_t^T R u_t + 2v_t^T R u_t + v_t^T R v_t \right) \\
- T(J_*(\Theta_*, \omega_t))
\]  

(42)

Following same steps as of [12] one can write:

\[
\sum_{t=0}^{T} \left( x_t^T Q x_t + u_t^T R u_t \right) = \sum_{t=0}^{T} \left( \sigma_2^2 \text{Tr}(P(\hat{\Theta}_{t-1}^*)) B_x B_x^T \right) \\
+ \sum_{t=0}^{T} \left( \sigma_2^2 \text{Tr}(P(\hat{\Theta}_{t-1}^*)) \right) + \sum_{t=T_0+1}^{T} \left( \sigma_2^2 \text{Tr}(P(\hat{\Theta}_{t-1}^*)) \right) \\
+ R_1 - R_2 - R_3
\]  

(43)

where

\[
R_1 = \sum_{t=0}^{T} \left( x_t^T P(\hat{\Theta}_{t-1}) x_t - \mathbb{E}[x_{t+1}^T P(\hat{\Theta}_t) x_{t+1} | \mathcal{F}_{t-1}] \right).
\]  

(44)

\[
R_2 = \sum_{t=0}^{T} \mathbb{E}[x_{t+1}^T (P(\hat{\Theta}_{t-1}) - P(\hat{\Theta}_t)) x_{t+1} | \mathcal{F}_{t-1}]
\]  

(45)

and

\[
R_3 = \sum_{t=0}^{T} \left( (\tilde{A}_{t-1} x_t + \tilde{B}_{t-1} u_t)^T P(\hat{\Theta}_{t-1}) (\tilde{A}_{t-1} x_t + \tilde{B}_{t-1} u_t) \\
- (A_* x_t + B_* u_t)^T P(\hat{\Theta}_{t-1}) (A_* x_t + B_* u_t) \right)
\]  

(46)

From given side information (see [12] in Assumption 2.3), one can write

\[
\bar{\sigma}_2^2 \text{Tr}(P(\hat{\Theta}_{t-1}^*)) = J_*(\hat{\Theta}_{t-1}^*, \omega_t) \leq J_*(\Theta_*^*, \omega_t) \\
\leq J_*(\Theta_*, \omega_t) + \gamma^*
\]  

(47)

Combining (47) with (43) and (42) under the events \( F_{T_c^*}^* \cap E_{T_c^*} \) for \( t \leq T_c^* \) and \( F_{\omega}^* \cap E_T \) for \( t \geq T_c^* \) the regret can be upper-bounded as follows:

\[
R(T) \leq \sigma_2^2 T_{\omega} D ||B_*||^2_{F} + \gamma^* T_{\omega} + R_0 + R_1 - R_2 - R_3
\]

where

\[
R_0 = \sum_{t=0}^{T_0} (2v_t^T R u_t + v_t^T R v_t)
\]  

(48)

Now, we proceed with bounding the regret bound terms individually.

**Lemma 4.1:** ([11]) On the event \( F_{T_c^*}^* \cap E_{T_c^*} \) for \( t \leq T_c^* \) and \( F_{\omega}^* \cap E_T \) for \( t \geq T_c^* \), with probability at least \( 1 - \delta/2 \) for \( t > T_c \) the term \( R_1 \) is upper-bounded as follows:

\[
R_1 \leq k_{c,1} (n + d_*)^{n+d_*}(\sigma_\omega + ||B_*|| \sigma_\nu) \\
\times n \sqrt{T_c^*} \log((n + d^*)T_c^*/\delta) \\
+ k_{c,2} \sigma_\omega n \sqrt{m} \sqrt{t - T_c^*} \log(n(t - T_c^*)/\delta) \\
+ k_{c,3} n \sigma_\omega \sqrt{t - T_c^*} \log(nt/\delta) \\
+ k_{c,4} n(\sigma_\omega + ||B_*|| \sigma_\nu)^2 \sqrt{T_c^*} \log(nt/\delta)
\]  

(49)

for some problem dependent coefficients \( k_{c,1}, k_{c,2}, k_{c,3}, k_{c,4} \).

**Proof:** Proof follows as same steps of [12], with only difference that exploration phase is done with the actuating mode \( i^* \) with number of actuators \( d_*^* \).

The following lemma upper-bounds the term \( R_2 \).
The inputs to the OFU algorithm are $T_i$ where
$$\text{The proof can be found in Appendix VII.}$$

Step with actuating mode $i$ regret of $O_1$ step on the regret bound. The following lemma adapted from [1 2] provides this bound.

The only remaining part is upper-bounding the term $R_3$ defined by
$$\text{Lemma 4.2: (Bounding $R_2$) On the event $E \cap F$, it holds true that the term $R_2$ defined by (45) is upper-bounded as}$$
$$|R_2| \leq 2Dc^2(n + d_{t, *})^2(n + d)$$
$$\times \left\{1 + \log_2 \left(1 + \frac{T^*}{\lambda} (n + d_{t, *})^2(1 + \kappa^2) \right) \right\}^{n + d_{t, *}}$$
$$+ 2D \frac{32n\sigma^2}{(1 - \gamma)^2} \log \frac{\delta}{\delta} \log_2 \left(\frac{\lambda}{\sigma(T^* + 1)}\right)^{n + d}$$
$$\text{in which}$$
$$\bar{\lambda} := \lambda + T^* (n + d_{t, *})^2(1 + \kappa^2)$$
$$+ 2dT^* \sigma^2 \log \left(\frac{4nT^*}{\lambda} (T^* + 1)\right)$$
$$+ (T - T^*) \frac{32n\sigma^2}{(1 - \gamma)^2} \log \frac{n(T - T^*)}{\delta}.$$}

The proof can be found in Appendix VII.

Lemma 4.3: (Bounding $R_3$) On the event $E^*_T \cap E_{T^*}$ for $t \leq T^*$ and $E^{opt}_T \cap E_T$ for $t \geq T^*$, the term $R_3$ defined by (46) has the following upper bound:
$$|R_3| = O\left((n + d_{t, *})^2(n + d)T^* + (n + d)n^2 \sqrt{T - T^*}\right).$$

The proof can be found in Appendix VII.

The only remaining part is upper-bounding the term $R_0$ given by (45) which is the direct effect of "more exploration" step on the regret bound. The following lemma adapted from [12] provides this bound.

Lemma 4.4: (Bounding $R_0$) On the event $E \cap F^*$, the term $R_0$ defined by (48) has the following upper bound:
$$|R_0| \leq d\sigma \sqrt{B_0 + d\|R\|\sigma^2}$$
$$\times \left(T_\omega + \sqrt{T_\omega \log \frac{4dT_\omega}{\delta} \sqrt{\log \frac{4}{\delta}}\right)$$
$$\text{where}$$
$$B_0 = 8(1 + T_\omega \kappa^2 \|R\|^2(n + d_{t, *})^2(1 + \kappa^2))$$
$$\times \log \left(\frac{4d}{\delta} (1 + T_\omega \kappa^2 \|R\|^2(n + d_{t, *})^2(1 + \kappa^2))^{1/2}\right).$$

The following theorem by combining the terms for individual terms, gives the regret upper-bound.

**Theorem 4.1:** (Regret Bound of IExp+SOFUA for controllable systems) Under Assumptions 2.1 and 2.2 with probability at least $1 - \delta$ the algorithm SOFUA together with additional exploration algorithm IExp which runs for $T_e$ time steps achieves regret of $O\left((n + d^*)^2(n + d)T_e\right)$ for $t \leq T_e$ and $O(\text{poly}(n + d)\sqrt{T - T_e})$ for $t > T_e$.

**Proof:** Recalling the regret
$$R(T) \leq \sigma^2T_\omega D\|B_e\|_2^2 + \gamma^T T_\omega + R_0 + R_1 - R_2 - R_3$$
combining the Lemmas 4.1 4.2 4.3 and 4.4 proves the theorem.

V. SIMULATION RESULTS

In this section, we investigate the effect of actuator redundancy and possibility of switching between different set of actuators on the boundedness of state trajectory. For this aim, we consider a control system (adopted from [21]) with drift and control matrices to be set as follows:

$$A = \begin{pmatrix} 0.54 & -0.11 \\ -0.026 & 0.63 \end{pmatrix}, \quad B = \begin{pmatrix} -85 & 4.4 \\ -2.5 & 2.8 \end{pmatrix} \times 10^{-4},$$

For the control cost we choose the cost matrices as follows:

$$Q = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1/50^2 & 0 \\ 0 & 0.1^6 \end{pmatrix}.$$

The inputs to the OFU algorithm are $T = 10000$, $\delta = 1/T$, $\lambda = 1$, $L = 0.1$, $s = 1$ and we repeat simulation 10 times.

we consider two cases of different in the actuating mode used for initial exploration phase. For first case, we perform this step with actuating mode $i^*$ (to be specified by the Algorithm 2) and for the other, all of the actuators are exploited to this
aim. The time duration of initial phase for both cases is obtained by the Algorithm 1. At time $T_c$, the Algorithm 3 comes to play which is applying optimism without more exploration noise on full actuation mode.

It has graphically been shown in [21] that the objective function $J(\Theta) = tr(P(\Theta))$ is generally non-convex and when it comes to one dimensional system ($n, m = 1$) it is only convex in drift matrix, $A$. Because of this fact, we decided to solve optimization problem (17) using a projected gradient descent method, with basic step

$$\Theta_{t+1} \leftarrow \text{PROJ}_{C_t(\delta)} \left( \Theta_t - \gamma \nabla_{\Theta} (tr(P(\Theta))) \right)$$ (53)

where $\nabla_{\Theta} f$ is the gradient of $f$ with respect to $\Theta$. $C_t(\delta)$ is the confidence set, $\text{PROJ}_g$ is Euclidean projection on $g$ and finally $\gamma$ is the step size. Computation of gradient $\nabla_{\Theta}$ as well as formulation of projection has been explicit in [21], similar to which we choose the learning rate as follows:

$$\gamma = \sqrt{\frac{0.001}{\text{tr}(V_t)}}$$

We apply the gradient method for 200 iterations to solve each OFU optimization problem and apply the projection technique until the projected point lies inside the confidence ellipsoid. The simulation results will be provided soon.

VI. CONCLUSION

In this work, we proposed an OFU principle-based controller for systems with actuator redundancy, which combines a step of “more-exploration” (to produce a stabilizing neighborhood of the true parameters while guaranteeing a bounded state during exploration) with one of “optimism”, which efficiently controls the system. Due to the redundancy, it is possible to further optimize the speed of convergence of the exploration phase to the stabilizing neighborhood by choosing over actuation modes, then to switch to full actuation to guarantee an $O(\sqrt{T})$ regret in closed-loop, with polynomial dependency on the system dimension.

A natural extension of this work is to classes of systems in which some modes are only stabilizable. Speaking more broadly, the theme of this paper also opens the door to more applications of switching as a way to facilitate learning-based control of unknown systems, some of which are the subject of current work.

References

[1] D. Liberzon and A. S. Morse, “Basic problems in stability and design of switched systems,” IEEE control systems magazine, vol. 19, no. 5, pp. 59–70, 1999.
[2] S. R. Kulkarni and P. J. Ramadge, “Model and controller selection policies based on output prediction errors,” IEEE Transactions on Automatic Control, vol. 41, no. 11, pp. 1594–1604, 1996.
[3] K. S. Narendra and J. Balakrishnan, “Adaptive control using multiple models,” IEEE transactions on automatic control, vol. 42, no. 2, pp. 171–187, 1997.
[4] A. Kanellopoulos and K. G. Vamvoudakis, “A moving target defense control framework for cyber-physical systems,” IEEE Transactions on Automatic Control, 2019.
[5] J. Xu, P. Guo, M. Zhao, R. F. Erbacher, M. Zhu, and P. Liu, “Comparing different moving target defense techniques,” in Proceedings of the First ACM Workshop on Moving Target Defense, 2014, pp. 97–107.
[6] S. Weerkakody and B. Sinopoli, “Detecting integrity attacks on control systems using a moving target approach,” in 2015 54th IEEE Conference on Decision and Control (CDC). IEEE, 2015, pp. 5820–5826.
[7] M. Takahashi and T. Takagi, “Adaptive fault-tolerant control based on hybrid redundancy,” Asia-Pacific Journal of Chemical Engineering, vol. 7, no. 5, pp. 642–650, 2012.
[8] Q. Yang, S. S. Ge, and Y. Sun, “Adaptive actuator fault tolerant control for uncertain nonlinear systems with multiple actuators,” Automatica, vol. 60, pp. 92–99, 2015.
[9] 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.
[10] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, “Explore more and improve regret in linear quadratic regulators,” in Proceedings of the 24th Annual Conference on Learning Theory, 2011, pp. 1–26.
[11] X. Chen and E. Hazan, “Black-box control for linear dynamical systems,” arXiv preprint arXiv:2007.06650, 2020.
[12] M. C. Campi and P. Kumar, “Adaptive linear quadratic gaussian control: the cost-biased approach revisited,” SIAM Journal on Control and Optimization, vol. 56, no. 2, pp. 388–415, 2017.
[13] 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.
[14] M. Ibrahimi, A. Javanmard, and B. V. Roy, “Efficient reinforcement learning for high dimensional linear quadratic systems,” in Advances in Neural Information Processing Systems, 2012, pp. 2656–2664.
[15] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, “Optimism-based adaptive regulation of linear-quadratic systems,” arXiv preprint arXiv:1711.07230, 2017.
[16] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, “Regret bound of adaptive control in linear quadratic gaussian (lqg) systems,” arXiv, 2020.
[17] ———, “Adaptive control and regret minimization in linear quadratic gaussian (lqg) setting,” arXiv preprint arXiv:2003.05999, 2020.
[18] B. D. Anderson and J. B. Moore, Linear Optimal Control [by] Brian DO Anderson [and] John B. Moore. Prentice-hall, 1971.
[19] Y. Abbasi-Yadkori, “Online learning for linearly parametrized control problems,” 2013.
Proof: (Proof of Lemma 3.3) we can write
\[
det(V_t) \leq \prod_{i=1}^{n+d} \left( \lambda + \sum_{k=1}^{t-1} z_{ki}^2 \right)
\]
\[
\leq \left( \sum_{i=1}^{n+d} \left( \lambda + \sum_{k=1}^{t-1} z_{ki}^2 \right) \right)^{n+d}
\]
\[
= \left( \frac{(n+d)\lambda + \sum_{k=1}^{t-1} 2\|z_k\|^2 + 2\|\zeta_k\|^2}{n+d} \right)^{n+d}
\]
In second inequality, we applied AM-GM inequality and in the third inequality we apply the property \((a+b)^2 \leq 2a^2 + 2b^2\). Furthermore, \(\|z_k\|^2 \leq (1 + 2\kappa^2)X^2\). Given \(\|\zeta_k\| \leq 2\sigma_\nu \sqrt{d\log(4nt(t+1)/\delta)}\) which holds with probability at least \(1 - \delta/2\) one can write:
\[
det(V_t) \leq \frac{\lambda}{(n+d)\lambda + \sum_{k=1}^{t-1} 2\|z_k\|^2 + 2\|\zeta_k\|^2}
\]
This completes the proof of (22). Proof of the second statement of lemma is given in [12].

VII. APPENDIX

Proof: (Proof of Theorem 3.1) Given (30), we first upper-bound the term \(G_i Z_n^{n+d+1} \beta_i(\delta/4)^{n+d+1} \). First, we can write
\[
Z_T = \max_{0 \leq t \leq T} \|z_t\| \leq \sqrt{1 + \kappa^2}X_T
\]
where \(X_T := \max_{0 \leq t \leq T} \|z_t\|\). On the other hand, given the definition of \(\beta_i(\delta/4)\) one can write:
\[
\beta_i(\delta/4) \leq \beta_i(\delta/4)^{n+d+1} \leq 4\sqrt{\beta_i(\delta)}
\]
Combining the results give:
\[
G_i Z_T^{n+d+1} \beta_i(\delta/4) \leq 4\sqrt{1 + \kappa^2}G_i \sqrt{\beta_i(\delta)X_T^{n+d+1}}
\]
We also have
\[
G_i = 2 \left( \frac{2S(n+d_i)}{U_{0.5}^{n+d_i+1}} \right)^{1/(n+d_i+1)}
\]
where
\[
U^i = \frac{U_0}{H} \quad \text{and} \quad U_0 \frac{1}{16n+d_i-2} \leq \sqrt{S^2(n+d_i-2)}
\]
One can simply rewrite \(G_i\) as follows:
\[
G_i = C H \left( \frac{1}{S^{n+d_i+1}} \right)
\]
where
\[
C := 2 \left( 2S(n+d_i)(16^{n+d_i-2} \sqrt{S^2(n+d_i-2)}) \right)^{1/(n+d_i+1)}
\]
and
\[
H_i > \left( 16 \sqrt{\frac{4S^2M_i^2}{(n+d_i)U_0^2}} \right)
\]
with \(M_i\) to be defined as follows
\[
M_i = \sup_{Y>0} \frac{\sigma_\omega \sqrt{n(n+d_i) \log (\frac{1+TY/\delta}{\delta})}}{Y} + \lambda^{1/2}S
\]
Given the fact that $Y = \sup_{0 \leq t \leq T} \|z_t\|$. Then by having nonzero initial state $x_0$ then by defining $Y^* = \sqrt{1 + \nu^2} \|x_0\|$ we have the following upper-bound for $M$

$$M_i \leq \frac{\sigma_\omega \sqrt{n(n + d_i) \log \left( \frac{1 + \lambda Y^*/\lambda}{\eta_i} \right) + \lambda^{1/2} \delta}}{Y^*} := M_{i,\text{max}}.$$ 

Then we can upper bound $G_i$ as follows

$$G_i \leq C \left( \frac{16}{\log \delta} + \frac{\lambda^{1/2} \delta}{(n + d_i)U_{\text{max}}} \right) := G_i$$

Hence, the state norm is bounded as follows:

$$\|x_t\| \leq D_1 \sqrt{\beta_t(\delta) \log(t)X} + D_2 \frac{\sqrt{t}}{\delta}$$

where

$$D_1 := \frac{1}{1 - \frac{\lambda}{n}} X_{\text{max}} \bar{G}_i, \quad D_2 := \frac{n\sqrt{2}}{1 - \frac{\lambda}{n}} (\eta_i)_{n + d_i} \left( \|B^*_i\|\sigma_\nu + \sigma_\omega \right)$$

Adopted from [11], it yields

$$X_i \leq \left( D_1 \sqrt{\beta_t(\delta) \log(t)X} + D_2 \frac{\sqrt{t}}{\delta} \right)^{n + d_i}$$

To upper-bound $\beta_t(\delta)$, we need the following upper bound for $\det(V_i)$

$$\det(V_i) \leq \left( \frac{(n + d_i)\lambda + (1 + \nu^2)X^2_t}{n + d_i} \right)^{n + d_i}$$

By applying elementary calculations, it yields

$$X_i \leq \left( \bar{F}_i + \bar{K}_i \log(X_i) \right)^{n + d_i + 1}$$

where

$$\bar{F}_i = D_1 \left( (n + d_i)\nu_\omega + \sqrt{\lambda S} \right) \log t + D_2 \frac{\sqrt{t}}{\delta} \log t + \frac{1}{n + d_i} \log \left( \frac{1 + \nu^2 t}{(n + d_i)\lambda} \right) \quad \text{and}$$

$$\bar{K}_i = 2D_1 (n + d_i)\nu_\omega.$$ 

Let $a_t = X_t^{(n + d_i + 1)}$, then it yields

$$a_t \leq \bar{F}_i + \bar{K}_i (n + d_i + 1) \log a_t$$

Let $c = \max\{1, \max_{1 \leq s \leq t} \|a_s\|\}$, assume that $t \geq \lambda(n + d_i)$ then it can be shown that

$$c \leq \bar{F}_i + \bar{L}_i \log^2(c)$$

where $\bar{L}_i = \bar{K}_i (n + d_i + 1)$. Using the property $\log x \leq x^a/s$ which holds $\forall s \in R^+$ and letting $s := 1$, one can write

$$c \leq \frac{-\bar{F}_i + \sqrt{\bar{F}_i^2 + 4\bar{L}_i}}{2\bar{L}_i}.$$ 

By elementary calculations first statement of the lemma is shown.

The proofs of statements 2 and 3 are immediate and we skip them for the sake of brevity.

Proof: (Proof of Lemma [12]) Note that except the times instances that there is switch in policy, most terms in RHS of (45) vanish. Denote covariance matrices of central ellipse and actuating mode $i^*$ by $V$ and $V_{i^*}$ respectively, and suppose at time steps $t_{n_1}, ..., t_{n_N}$ the algorithm changes the policy. Therefore, it yields det($V_{i_{n_1}}$) $\geq$ det($V_{i_{n_{n-1}}}^{i_{n-1}}$) for $t \leq T^*_c$ and det($V_{i_{n_j}}$) $\geq$ 2 det($V_{i_{n_{j-1}}}^{i_{n_{j-1}}}$) for $t \geq T^*_c$. This results in

$$\det(V^{i_{n_1}}) \geq 2^{N_1} \lambda^{n + d_i},$$

$$\det(V_T) \geq 2^{N_2} \det(V^{i_{n_{j+1}}})$$

(58) (59)
where \( N_1 \) and \( N_2 \) are number of switches in policy while actuating in the mode \( i^* \) and fully actuating mode respectively. On one hand \( \det(V_{T_c^{i^*}}) \leq \lambda_{\max}^{n+d}(V_{T_c^{i^*}}) \) where

\[
\lambda_{\max}(V_{T_c^{i^*}}) \leq \lambda + \sum_{t=0}^{T_{i^*}^c - 1} \|z_t\|^2 \leq \lambda + (1 + \kappa^2)c^* (n + d_{i^*})^2(1 + \kappa^2) \]

(60)

using \( N_1 \) is upper-bounded by

\[
N_1 \leq \log_2 \left( 1 + \frac{T_{i^*}^c (n + d_{i^*})^2(1 + \kappa^2)}{\lambda} \right)^{n+d_{i^*}}
\]

(61)

On the other hand we have \( \det(V_T) \leq \lambda_{\max}^{n+d}(V_T) \) where

\[
\lambda_{\max}(V_T) \leq \lambda + \sum_{t=0}^{T-1} \|z_t\|^2 \leq \lambda + \sum_{t=0}^{T_{i^*}^c - 1} (\|z_t^*\|^2 + \|\zeta_t\|^2) + \sum_{t=T_{i^*}^c + 1}^{T-1} \|z_t\|^2.
\]

(62)

Furthermore,

\[
\Lambda := \max_{0 \leq t \leq T_c} \|\zeta_t\| \leq 2\sigma_\nu \sqrt{d \log(4nT_{i^*}^c (T_{i^*}^c + 1) / \delta)}
\]

(63)

with probability at least \( 1 - \delta / 2 \) together with upper-bounds of state norm in both initial exploration and optimism parts yields

\[
\lambda_{\max}(V_T) \leq \tilde{\lambda} := \lambda + T_{i^*}^c (n + d_{i^*})^2(1 + \kappa^2)
\]

\[
+ 2dT_{i^*}^c \sigma_\nu^2 \log(4nT_{i^*}^c (T_{i^*}^c + 1) / \delta)
\]

\[
+ (T - T_{i^*}^c) \frac{32n\sigma_\nu^2 (1 + \kappa^2)}{(1 - \hat{\nu})^2} \log n(T - T_{i^*}^c) / \delta
\]

(64)

Considering (20) we have

\[
\lambda_{\min}(V_{T_{i^*}^c + 1}) \geq \sigma_\nu^2(T_{i^*}^c + 1)
\]

(65)

now applying \( \det(V_T) \leq \lambda_{\max}^{n+d} \) results in

\[
N_2 \leq \log_2 \left( \frac{\tilde{\lambda}}{\sigma_\nu^2 (T_{i^*}^c + 1)} \right)^{n+d}
\]

(66)

Considering switch from IExp to SOFUA that can cause switch in the policy, the total number of switch in policy is \( N_1 + N_2 + 1 \). Now, applying (33) and upper bounds of state norm for \( t \leq T_{i^*}^c \) and \( t > T_{i^*}^c \) complete proof.

The following lemma adapts the proof of [11] to our setting which will be useful in bounding \( R_3 \).

**Lemma 7.1:** ([11]) On the event \( F_{T_c^{i^*}} \cap E_{T_c^{i^*}} \) for \( t \leq T_c^{i^*} \) and \( F_{T}^{\text{opt}} \cap E_{T}^{\text{opt}} \) for \( t \geq T_c^{i^*} \), the following holds,

\[
\sum_{t=0}^{T} \| (\Theta_s - \hat{\Theta}_t)^T z_t \|^2 \leq \frac{16(1 + \kappa^2)}{\lambda}
\]

\[
\times \left( 2X_c^2 \beta_{T_c^c}^2 \log \frac{\det(V_{T_c^c})}{\det(I)} + X_c^2 \beta_{T_c^c}^2 \log \frac{\det(V_T)}{\det(V_{T_c^c})} \right) + 2S^2 \Lambda_0^2 t_{i^*}
\]

(67)

where \( X_c^2 = \frac{32n\sigma_\nu^2 (1 + \kappa^2)}{(1 - \hat{\nu})^2} \log n(T - T_{i^*}^c) / \delta \) and \( X_c^2 = c^* (n + d_{i^*})^2(1 + \kappa^2) \)

**Proof:** Let \( s_t = (\Theta_s - \hat{\Theta}_t)^T z_t \), then one can write

\[
\| s_t \| \leq \| (\Theta_s - \hat{\Theta}_t)^T z_t \| + \| (\hat{\Theta}_t - \hat{\Theta}_t)^T z_t \|
\]
For $t > T_c^*$ we have:

\[
\| (\Theta - \hat{\Theta}_t)^\top z_t \| \leq \| V_t^{1/2} (\Theta - \hat{\Theta}_t) \| z_t \| \| V_t^{\top}^{-1} \|
\]

\[
\leq \| V_t^{1/2} (\Theta - \hat{\Theta}_t) \| \| z_t \| \sqrt{\det(V_t)} \det(V_t)
\]

\[
\leq \sqrt{2} \| V_t^{1/2} (\Theta - \hat{\Theta}_t) \| z_t \| V_t^{\top} \|
\]

\[
\leq \sqrt{2} \beta_r(\delta) \| z_t \| V_t^{\top} \|
\]

where $\tau \leq t$ is the last time that policy change happened. We applied Cauchy-Schwartz inequality in (68), (69) follows from the Lemma 11 in [11], applying the update rule renders (70) and finally (71) is obtained using the property $\lambda_{\max} \leq \text{Tr}(M)$ for $M \succeq 0$. Then we can write,

\[
\sum_{t=T_c}^{T} \| (\Theta - \hat{\Theta}_t)^\top z_t \|^2 \leq \frac{8X_2^2 (1 + \kappa^2) \beta_{\epsilon}^2 (\delta)}{\lambda}
\]

\[
\times \sum_{t=T_c^* + 1}^{T} \min\{\| z_t \|^2_{V_t^{\top}^{-1}}, 1\}
\]

\[
\leq 16X_2^2 (1 + \kappa^2) \beta_{\epsilon}^2 (\delta) \log \frac{\det(V_T)}{\det(V_{T_c^*}^{-1})}
\]

(72)

where in the second inequality we applied the Lemma 10 of [11]. For $t \leq T_c^*$ we have

\[
\| (\Theta - \hat{\Theta}_t)^\top z_t \| \leq \| (\Theta^* - \hat{\Theta}_t^*)^\top z_t^* \| + SA
\]

\[
\leq \| V_t^{\top}^{1/2} (\Theta^* - \hat{\Theta}_t^*) \| z_t^* \| V_t^{\top}^{\top} \|
\]

\[
\leq \| V_t^{\top}^{1/2} (\Theta^* - \hat{\Theta}_t^*) \| z_t^* \| V_t^{\top} \|
\]

\[
\times \sqrt{\det(V_t^{\top})} + SA
\]

\[
\leq \sqrt{2} \| V_t^{\top}^{1/2} (\Theta^* - \hat{\Theta}_t^*) \| z_t^* \| V_t^{\top} \|
\]

\[
\leq \sqrt{2} \beta_{\epsilon}^* (\delta) \| z_t^* \| V_t^{\top} \|
\]

where in (73) we decomposed the term $(\Theta - \hat{\Theta}_t)^\top z_t$, and applied triangle inequality, and applied (63). The rest of inequalities follows the same steps as of (68, 71).

Applying this result gives

\[
\sum_{t=0}^{T_c^*} \| (\Theta - \hat{\Theta}_t)^\top z_t \|^2 \leq \frac{16(1 + \kappa^2)X_2^2 \beta_{\epsilon}^* 2^2 (\delta)}{\lambda}
\]

\[
\times \sum_{t=0}^{T_c^*} \min\{\| z_t^* \|^2_{V_t^{\top}^{-1}}, 1\} + 2S^2A^2T_c^*
\]

\[
\leq \frac{32X_2^2 (1 + \kappa^2) \beta_{\epsilon}^* (\delta)}{\lambda} \log \frac{\det(V_t^{\top})}{\det(M)}
\]

\[
+ 2S^2A^2T_c^*
\]

where similar to (72) in the second inequality we applied the Lemma 10 of [11]. Combining (72) and (74) completes proof.

\[
\Box
\]

**Proof:** (Proof of Lemma 4.3) In upper-bounding the term $R_3$ we skip a few straight forward steps which can be found
\[
|R_k| \leq \left( \sum_{t=0}^{T_k^*} \left\| P(\tilde{\Theta}_t)^{1/2}(\tilde{\Theta}_t - \Theta_*)^\top z_t \right\|^2 \right)^{1/2} \\
\times \left( \sum_{t=0}^{T_k^*} \left( \left\| P(\tilde{\Theta}_t)^{1/2}\tilde{\Theta}_t^\top z_t \right\| + \left\| P(\tilde{\Theta}_t)^{1/2}\Theta_*^\top z_t \right\| \right)^2 \right)^{1/2} \\
+ \left( \sum_{t=T_k^*}^{T} \left\| P(\tilde{\Theta}_t)^{1/2}(\tilde{\Theta}_t - \Theta_*)^\top z_t \right\|^2 \right)^{1/2} \\
\times \left( \sum_{t=T_k^*}^{T} \left( \left\| P(\tilde{\Theta}_t)^{1/2}\tilde{\Theta}_t^\top z_t \right\| + \left\| P(\tilde{\Theta}_t)^{1/2}\Theta_*^\top z_t \right\| \right)^2 \right)^{1/2} \\
\leq \sqrt{32D(1 + \kappa^2)\beta T^* \sigma^2} \lambda \log \left( \frac{\det(V_{T_k^*}^*)}{\det(\Lambda T^*)} \right) + 2DS^2\Lambda^2 T_k^* \\
\times \sqrt{4T_k^* D(1 + \kappa^2)X_e^2} \\
+ \sqrt{16D(1 + \kappa^2)\beta T^* (\delta)X_e^2} \lambda \log \left( \frac{\det(V_T)}{\det(V_{T_k^*}^*)} \right) \\
\times \sqrt{4(T - T_k^*) DS^2(1 + \kappa^2)X_e^2}
\]

where in the inequalities (75) and (76) we applied (74) and (72) (from the Lemma 7.1) respectively. The remaining step is following upper-bounds

\[
\log \left( \frac{\det(V_{T_k^*}^*)}{\det(\Lambda T^*)} \right) \leq (n + d) \log \left( 1 + \frac{T_k^* (1 + \kappa^2)X_e^2}{\lambda(n + d)} \right) \\
\log \left( \frac{\det(V_T)}{\det(V_{T_k^*}^*)} \right) \leq (n + d) \\
\times \log \left( \frac{\lambda(n + d) + T_k^* (1 + \kappa^2)X_e^2 + (T - T_k^*) (1 + \kappa^2)X_e^2}{\sigma^2 T_k^* (n + d)} \right)
\]

where is the second inequality we applied (65). Considering the definitions of \(X_e^2\), \(X_c^2\) and \(\Lambda\) we can easily notice the statement of lemma holds true.