Finite-time System Identification and Adaptive Control in Autoregressive Exogenous Systems

Sahin Lale[1] ALALE@CALTECH.EDU
Kamyar Azizzadenesheli[2] KAMYAR@PURDUE.EDU
Babak Hassibi[1] HASSIBI@CALTECH.EDU
Anima Anandkumar[1] ANIMA@CALTECH.EDU
[1] *California Institute of Technology*
[2] *Purdue University*

Abstract

Autoregressive exogenous (ARX) systems are the general class of input-output dynamical system used for modeling stochastic linear dynamical system (LDS) including partially observable LDS such as LQG systems. In this work, we study the problem of system identification and adaptive control of unknown ARX systems. We provide finite-time learning guarantees for the ARX systems under both open-loop and closed-loop data collection. Using these guarantees, we design adaptive control algorithms for unknown ARX systems with arbitrary strongly convex or non-strongly convex quadratic regulating costs. Under strongly convex cost functions, we design an adaptive control algorithm based on online gradient descent to design and update the controllers that are constructed via a convex controller reparametrization. We show that our algorithm has $\tilde{O}(\sqrt{T})$ regret via explore and commit approach and if the model estimates are updated in epochs using closed-loop data collection, it attains the optimal regret of polylog(T) after T time-steps of interaction. For the case of non-strongly convex quadratic cost functions, we propose an adaptive control algorithm that deploys the optimism in the face of uncertainty principle to design the controller. In this setting, we show that the explore and commit approach has a regret upper bound of $\tilde{O}(T^{2/3})$, and the adaptive control with continuous model estimate updates attains $\tilde{O}(\sqrt{T})$ regret after T time-steps.

Keywords: ARX systems, system identification, adaptive control, regret

1. Introduction

Autoregressive Exogenous (ARX) Systems: ARX systems are simple yet central dynamical systems in time-series modelings. They represent stochastic linear dynamical systems (LDS) in the input-output form which have a wide range of applicability to real dynamical systems and amenability for precise analysis. Due to their ability to approximate linear systems in a parametric model structure, ARX systems have been crucial in many areas including chemical engineering, power engineering, medicine, economics, and neuroscience (Norquay et al., 1998; Bacher et al., 2009; Fetics et al., 1999; Huang and Jane, 2009; Burke et al., 2005). The ARX systems have corresponding LTI state-space representations and in their most general form, they can be represented as follows,

$$x_{t+1} = Ax_t + Bu_t + Fy_t, \qquad y_t = Cx_t + e_t. \tag{1}$$

The dynamics are governed by $\Theta = (A, B, C, F)$ where x_t is the internal state, y_t is the output, u_t is the input and e_t is the measurement noise. Notice that by knowing the initial condition x_0

and Θ, one can recover the state sequence. These models provide a *general* representation of LDS with *arbitrary* stochastic disturbances. In particular, via different distributions of e_t, they are able to model partially observed LDS (PO-LDS) with various process and measurement noises. For instance, LQG control systems, which are the canonical settings in control, can be modeled as ARX systems. In an LQG control system, the process and measurement noises have Gaussian distributions which corresponds (in predictive form) to an ARX system, where e_t has a particular Gaussian distribution determined by the state-space parameters and noise distributions (Kailath et al., 2000).

System Identification and Adaptive Control: They are the central problems in control theory and reinforcement learning (Lai et al., 1982). System identification aims to learn the unknown dynamics of the system from the collected data, whereas adaptive control pursues the goal of minimizing the cumulative control cost of dynamical systems with unknown dynamics. Thus, adaptive control inherently includes the system identification process to design a favorable controller. The data collection to achieve these tasks can be performed via independent control inputs yielding open-loop data collection, or via feedback controllers resulting in closed-loop data collection (Ljung, 1999).

Finite-time System Identification and Adaptive Control: In contrast to classical results in both of these problems that analyze the asymptotic performances, recently, there has been a flurry of studies that consider the finite-time performance and learning guarantees in both. In finite-time system identification setting pioneered by Campi and Weyer (2002, 2005), currently, the main focus has been on obtaining optimal learning rate of $1/\sqrt{T}$ after T samples. Using open-loop data collection to avoid correlations in the inputs and outputs, Oymak and Ozay (2018); Sarkar et al. (2019); Tsiamis and Pappas (2019); Simchowitz et al. (2019) suggest methods that achieve this rate for stable LDS. However, due to the difficulty in handling the correlations caused by the feedback controller, the closed-loop system identification guarantees are scarce. Recently, Lale et al. (2020a) propose the first finite-time system identification algorithm that attains the optimal learning rate guarantee for both open and closed-loop data collection.

In finite-time adaptive control, the efforts have been centered around achieving sub-linear regret which measures the difference between the cumulative cost of the adaptive controller and the optimal controller that knows the system dynamics. Most of the prior works follow the explore and commit approach. This approach proposes to first use open-loop data collection to solely explore the system and then estimate the system dynamics and fix a policy to be applied for the remaining time-steps (Lale et al., 2020b; Mania et al., 2019; Simchowitz et al., 2020). The recent introduction of the first finite-time closed-loop system identification algorithm in Lale et al. (2020a) allowed the design of “truly” adaptive control algorithms that naturally use past experiences to improve the model estimates and the controller continuously. Deploying closed-loop data collection, Lale et al. (2020c,a) provide adaptive control algorithms for PO-LDS that achieve optimal regret results.

Contributions: In this work, we study finite-time system identification and adaptive control problems in ARX modeled systems with sub-Gaussian noise. First, we state the finite-time guarantees for learning the ARX systems that hold for both open and closed-loop data collection. Deploying the least-squares problem introduced in Lale et al. (2020a), we show that the estimation error of model parameters decays with $\tilde{O}(1/\sqrt{T})$ rate after collecting T samples with persistent excitation.

Secondly, we study the adaptive control problem in ARX modeled systems with sub-Gaussian noise. Leveraging the finite-time system identification results, we propose adaptive control frameworks for the ARX systems with strongly convex or non-strongly convex quadratic cost functions:

Table 1: Comparison with prior works for PO-LDS. Our results extend similar regret guarantees to general ARX systems with sub-Gaussian noise disturbances, subsuming the prior works. E&C := Explore-and-commit approach CLU := Closed-loop model estimate updates

Work	Regret	Setting	Cost	Noise	Method
Mania et al. (2019)	$\sqrt{T}$	PO-LDS	Str. Convex	Gaussian	E&C
Simchowitz et al. (2020)	$\sqrt{T}$	PO-LDS	Str. Convex	Semi-adversarial	E&C
Lale et al. (2020a)	$\text{polylog}(T)$	PO-LDS	Str. Convex	Gaussian	CLU
Lale et al. (2020b)	$T^{2/3}$	PO-LDS	Convex	Gaussian	E&C
Lale et al. (2020c)	$\sqrt{T}$	PO-LDS	Convex	Gaussian	CLU
Theorem 3	$\sqrt{T}$	ARX	Str. Convex	Sub-Gaussian	E&C
Theorem 4	$\text{polylog}(T)$	ARX	Str. Convex	Sub-Gaussian	CLU
Theorem 5	$T^{2/3}$	ARX	Convex	Sub-Gaussian	E&C
Theorem 6	$\sqrt{T}$	ARX	Convex	Sub-Gaussian	CLU

1. ARX systems with strongly convex cost functions: For this cost function setting, which can possibly be time-varying, we provide an adaptive control algorithm framework that deploys online learning for controller design and exploits the strong convexity. Using online gradient descent with a convex policy reparametrization of linear controllers, we show that adaptive control problem turns into an online convex optimization problem and optimal regret results can be achieved in this setting. To this end, we first show that the explore and commit approach, which fixes the model estimate after open-loop data collection, attains regret of $\tilde{O}(\sqrt{T})$ after T time-steps of interaction via the proposed framework. Here $\tilde{O}(\cdot)$ presents the order up to logarithmic terms. We then show that if the model estimates are updated in epochs using the data collected in closed-loop, this adaptive control framework of ARX systems yields the optimal regret rate of $\text{polylog}(T)$.

2. ARX models with fixed non-strongly convex quadratic cost function: For this setting, we propose an adaptive control framework that deploys the principle of optimism in the face of uncertainty (OFU) (Auer, 2002) to balance exploration vs. exploitation trade-off in the controller design. The OFU principle prescribes to use the optimal policy of the model that has the lowest optimal cost, *i.e.* the optimistic model, within the plausible set of systems according to system identification guarantees. We show that using this framework with the explore and commit approach yields regret of $\tilde{O}(T^{2/3})$. Ultimately, we prove that the adaptive control based on OFU principle attains regret of $\tilde{O}(\sqrt{T})$ if the model estimates are continuously updated using closed-loop data in ARX systems.

These results extend the prior results in PO-LDS to the general class of ARX systems with sub-Gaussian noise which can be adopted in various real-world time-series modelings (Table 1).[1]

2. Preliminaries

The Euclidean norm of a vector x is denoted as $\|x\|_2$. For a given matrix A, $\|A\|_2$ denotes its spectral norm, $\|A\|_F$ is its Frobenius norm, $A^\top$ is its transpose, $A^\dagger$ is its Moore-Penrose inverse, and $\text{Tr}(A)$ is the trace. $\rho(A)$ denotes the spectral radius of A, *i.e.*, the largest absolute value of its

[1] Due to the limited space, the Appendix, which contains the proofs, and the details are omitted. Interested readers are referred to the extended version of this work found online.

eigenvalues. The j-th singular value of a rank-n matrix A is denoted by $\sigma_j(A)$, where $\sigma_{\max}(A) := \sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_n(A) := \sigma_{\min}(A) > 0$. I is the identity matrix with appropriate dimensions. $\mathcal{N}(\mu, \Sigma)$ denotes a multivariate normal distribution with mean vector μ and covariance matrix Σ.

Consider the unknown ARX model of Θ given in (1). At each time-step t, the system is at state x_t and the agent observes y_t. Then, the agent applies a control input u_t, observes the loss function ℓ_t, pays the cost of $c_t = \ell_t(y_t, u_t)$, and the system evolves to a new x_{t+1} at time step $t+1$.

Assumption 2.1 (Sub-Gaussian Noise) *There exists a filtration $(\mathcal{F}_t)$ such that for all $t \geq 0$, and $j \in [0, \ldots, m]$, $e_{t,j}$s are R^2-sub-Gaussian, i.e., for any $\gamma \in \mathbb{R}$, $\mathbb{E}\left[\exp\left(\gamma e_{t,j}\right) | \mathcal{F}_{t-1}\right] \leq \exp\left(\gamma^2 R^2/2\right)$ and $\mathbb{E}\left[e_t e_t^\top | \mathcal{F}_{t-1}\right] = \Sigma_E \succ \sigma_e^2 I$ for some $\sigma_e^2 > 0$.*

Following general construction of ARX models we assume that A is stable such that $\Phi(A) = \sup_{\tau \geq 0} \|A^\tau\| / \rho(A)^\tau$ is finite. This is a mild assumption and captures extensive number of systems including detectable partially observable linear dynamical systems (Kailath et al., 2000).

3. System Identification

Using the dynamics in (1), for any positive integer h, the output of the system can be written as

$$y_t = \sum\nolimits_{k=0}^{h-1} CA^k \left(Bu_{t-k-1} + Fy_{t-k-1}\right) + e_t + CA^h x_{t-h}. \tag{2}$$

The behavior of an ARX system is uniquely governed by its Markov parameters.

Definition 1 (Markov Parameters) *The set of matrices that maps the previous inputs to the output is called input-to-output Markov parameters and the ones that map the previous outputs to the output are denoted as output-to-output Markov parameters of the system Θ. In particular, the matrices that map inputs and outputs to the output in (2) are the first h parameters of the Markov operator, $\mathbf{G} = \{G^i_{u\to y}, G^i_{y\to y}\}_{i\geq 1}$ where $\forall i \geq 1$, $G^i_{u\to y} = CA^{i-1}B$ and $G^i_{y\to y} = CA^{i-1}F$ which are unique.*

Let $\mathbf{G_{u\to y}}(h) = [G^1_{u\to y} G^2_{u\to y} \ldots G^h_{u\to y}] \in \mathbb{R}^{m\times hp}$ and $\mathbf{G_{y\to y}}(h) = [G^1_{y\to y} G^2_{y\to y} \ldots G^h_{y\to y}] \in \mathbb{R}^{m\times hm}$ denote the h-length Markov parameters matrices. Consider the following h-length operator $\mathcal{G}$ and the subsequences of h input-output pairs from the data collected, either open or closed-loop or both,

$$\mathcal{G} = [\mathbf{G_{u\to y}}(h) \;\; \mathbf{G_{y\to y}}(h)] \in \mathbb{R}^{m\times h(m+p)}, \quad \phi_i = [u_{i-1}^\top \ldots u_{i-h}^\top \;\; y_{i-1}^\top \ldots y_{i-h}^\top]^\top \in \mathbb{R}^{h(m+p)} \tag{3}$$

for $h \leq i \leq t$. Using $\mathcal{G}$, at each time step t, the output of the system can be written as

$$y_t = \mathcal{G}\phi_t + e_t + CA^h x_{t-h}. \tag{4}$$

Since A is stable, for $h = c_h \log(T)$, for some problem dependent constant c_h and total execution duration of T, the last term in (4) provides a negligible bias term of $1/T^2$. Therefore, we solve the following regularized least squares problem to estimate the Markov parameters of the system:

$$\widehat{\mathcal{G}}_t = \arg\min_{\mathcal{G}} \lambda \|X\|_F^2 + \sum\nolimits_{i=h}^{t} \|y_i - \mathcal{G}\phi_i\|_2^2. \tag{5}$$

The problem in (5) is first introduced in Lale et al. (2020a) to recover LQG systems in predictor form, which is a special case of ARX systems with sub-Gaussian noise. The following learning guarantee for (5) follows from Theorem 3 of Lale et al. (2020a), which is presented for i.i.d. Gaussian innovation terms yet holds for sub-Gaussian measurement disturbances of ARX systems.

Theorem 2 (Learning Markov Parameters of ARX Systems) *Let $\widehat{\mathcal{G}}_t$ be the solution to (5) at time t. For the given choice of h, define $V_t = \lambda I + \sum_{i=h}^{t} \phi_i \phi_i^\top$. Let $\|\mathcal{G}\|_F \leq S$. For $\delta \in (0,1)$, with probability at least $1-\delta$, for all $t \leq T$, $\mathcal{G}$ lies in the set $\mathcal{C}_{\mathcal{G}}(t)$, where*

$$\mathcal{C}_{\mathcal{G}}(t) = \{\mathcal{G}' : \operatorname{Tr}((\widehat{\mathcal{G}}_t - \mathcal{G}')V_t(\widehat{\mathcal{G}}_t - \mathcal{G}')^\top) \leq \beta_t\},$$

for $\beta_t = \left(\sqrt{mR\log(\delta^{-1}\det(V_t)^{1/2}\det(\lambda I)^{-1/2})} + S\sqrt{\lambda} + t\sqrt{h}/T^2\right)^2$. Furthermore, for persistently exciting inputs, i.e., $\sigma_{\min}(V_t) \geq \sigma_\star^2 t$ for some $\sigma_\star > 0$, and bounded ϕ_i, with high probability, the least square estimate $\widehat{\mathcal{G}}_t$ obeys $\|\widehat{\mathcal{G}}_t - \mathcal{G}\|_F = \tilde{\mathcal{O}}(1/\sqrt{t})$

This result shows that under persistent of excitation, the least squares problem (5) provides consistent estimates and the estimation error decays with the optimal rate. Note that both input-to-output and output-to-output Markov parameters of ARX system are submatrices of $\mathcal{G}$. Therefore, the given bound trivially holds for $\|\mathbf{G}_{\mathbf{u}\to\mathbf{y}}(h) - \widehat{\mathbf{G}}_{\mathbf{u}\to\mathbf{y}}(h)\|$ and $\|\mathbf{G}_{\mathbf{y}\to\mathbf{y}}(h) - \widehat{\mathbf{G}}_{\mathbf{y}\to\mathbf{y}}(h)\|$.

4. Adaptive Control of ARX Systems with Strongly Convex Cost

In this section, we will first introduce linear dynamic controllers (LDC) and provide a convex policy reparametrization, disturbance feedback controllers (DFC) (Simchowitz et al., 2020; Lale et al., 2020a), to approximate LDC controllers. We then provide the details of the setting of ARX systems regarding the loss and regret definition. Finally, we consider two variants of an algorithm that uses DFC policies in adaptive control of ARX system and provide the regret performances.

Linear Dynamic Controllers (LDC): An LDC (π) is a linear controller with internal state dynamics $s_{t+1}^\pi = A_\pi s_t^\pi + B_\pi y_t$ and $u_t^\pi = C_\pi s_t^\pi + D_\pi y_t$ where $s_t^\pi \in \mathbb{R}^s$ is the state of the controller, y_t is the input to the controller, *i.e.* observation from the system, and u_t^π is the output of the controller. $(A_\pi, B_\pi, C_\pi, D_\pi)$ control the internal dynamics of the LDC. LDC include a large number of controllers including H_2 and H_∞ controllers of fully and partially observable LDS (Hassibi et al., 1999). The optimal control law for ARX models with quadratic cost is also an LDC (Section 5).

Output uncertainties $\bar{b}_t(\mathcal{G})$: The output can be decomposed to its components via $\mathbf{G}$ as follows,

$$y_t = \sum\nolimits_{k=0}^{t-1} G_{u\to y}^{k+1} u_{t-k-1} + G_{y\to y}^{k+1} y_{t-k-1} + CA^t x_0 + e_t.$$

The output uncertainties of ARX system at time t is denoted as follows:

$$\bar{b}_t(\mathcal{G}) = y_t - \left(\sum\nolimits_{k=0}^{t-1} G_{u\to y}^{k+1} u_{t-k-1} + G_{y\to y}^{k+1} y_{t-k-1}\right) = CA^t x_0 + e_t. \tag{6}$$

This definition is similar to Nature's output adopted in Simchowitz et al. (2020); Lale et al. (2020a). It represents the only unknown components on the output. Notice that, one can identify the uncertainty in the output at any time step uniquely using the history of inputs, outputs and the Markov parameters. This gives the ability of counterfactual reasoning, *i.e.*, consider what the output would have been, if the agent had taken different sequence of inputs and observed different outputs.

4.1. Adaptive Control Setting

Disturbance Response Controllers (DFC): For adaptive control of ARX systems with strongly convex cost functions, we adopt a convex policy parametrization called DFC. A DFC of length h' is defined as a set of parameters, $\mathbf{M}(h') := \{M^{[i]}\}_{i=0}^{h'-1}$ acting on the last h' output uncertainties, *i.e.*,

$$u_t^{\mathbf{M}} = \sum\nolimits_{i=0}^{h'-1} M^{[i]} \bar{b}_{t-i}(\mathcal{G}). \tag{7}$$

Algorithm 1 Adaptive Control of ARX Systems with Strongly Convex Cost

```
Input: ID, $T$, $h$, $h'$ $T_w$, $\tau$, $S > 0$, $\delta > 0$, $\eta_t$
if ID = Explore & Commit then Set $T_{\text{warm}} = T_w$, else Set $T_{\text{warm}} = \tau$
—— WARM-UP ——
for $t = 0, 1, \dots, T_{\text{warm}}$ do
    Deploy $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$ and store $\mathcal{D}_{T_{\text{warm}}} = \{y_t, u_t\}_{t=1}^{T_{\text{warm}}}$ and set $\mathbf{M}_t$ as any member of $\mathcal{M}_r$
—— ADAPTIVE CONTROL ——
for $i = 0, 1, \dots$ do
    Calculate $\widehat{\mathcal{G}}_i$ via (5) using $\mathcal{D}_i = \{y_t, u_t\}_{t=1}^{2^i T_{\text{warm}}}$
    if ID = Explore & Commit then Set $\widehat{\mathcal{G}}_i = \widehat{\mathcal{G}}_0$ → IN E&C, ONLY $\widehat{\mathcal{G}}_0$ USED FOR CONTROL
    Compute $\bar{b}_j(\widehat{\mathcal{G}}_i) := y_j - (\sum_{k=0}^{h-1} \widehat{G}_{u\to y}^{k+1} u_{j-k-1} + \widehat{G}_{y\to y}^{k+1} y_{j-k-1}), \forall j \le t$
    for $t = 2^i T_{\text{warm}}, \dots, 2^{i+1} T_{\text{warm}} - 1$ do
        Observe $y_t$, and compute $\bar{b}_t(\widehat{\mathcal{G}}_i) := y_t - (\sum_{k=0}^{h-1} \widehat{G}_{u\to y}^{k+1} u_{t-k-1} + \widehat{G}_{y\to y}^{k+1} y_{t-k-1})$
        Commit to $u_t^{\mathbf{M}_t} = \sum_{j=0}^{H'-1} M_t^{[j]} \bar{b}_{t-j}(\widehat{\mathcal{G}}_i)$, observe $\ell_t$, and pay a cost of $\ell_t(y_t, u_t^{\mathbf{M}_t})$
        Update $\mathbf{M}_{t+1} = proj_{\mathcal{M}_r}\left(\mathbf{M}_t - \eta_t \nabla f_t\left(\mathbf{M}_t, \widehat{\mathcal{G}}_i\right)\right)$, $\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{y_t, u_t\}$
```

This convex policy parameterization follows the classical Youla parameterization (Youla et al., 1976) and used for adaptive control of PO-LDS in Simchowitz et al. (2020); Lale et al. (2020a). DFC policies are truncated approximations of LDC policies and for any LDC policy there exists a DFC policy which provides equivalent performance (see Appendix A).

Define the closed, convex and compact sets of DFCs, $\mathcal{M}$ and $\mathcal{M}_r$ such that the controllers $\mathbf{M}(h'_0) = \{M^{[i]}\}_{i=0}^{h'_0-1} \in \mathcal{M}$ are bounded and $\mathcal{M}_r$ is an r-expansion of $\mathcal{M}$, *i.e.*, $\sum_{i\geq 0}^{h'_0-1} \|M^{[i]}\| \leq \kappa_\psi$ and $\mathcal{M}_r = \{\mathbf{M}(h') = \mathbf{M}(h'_0) + \Delta : \mathbf{M}(h'_0) \in \mathcal{M}, \sum_{i\geq 0}^{h'-1} \|\Delta^{[i]}\| \leq r\kappa_\psi\}$, where $h'_0 = \lfloor \frac{h'}{2} \rfloor - h$. Therefore, all controllers $\mathbf{M}(h') \in \mathcal{M}_r$ are also bounded $\sum_{i\geq 0}^{h'-1} \|M^{[i]}\| \leq \kappa_\psi(1+r)$. Throughout the interaction with the system, the agent has access to $\mathcal{M}_r$.

Loss function: The loss function $\ell_t(\cdot,\cdot)$ is strongly convex, smooth, sub-quadratic and Lipschitz with a parameter L, such that for all t, $0 \prec \underline{\alpha}_{loss} I \preceq \nabla^2 \ell_t(\cdot,\cdot) \preceq \overline{\alpha}_{loss} I$ for a finite constant $\overline{\alpha}_{loss}$ and for any Γ with $\|u\|, \|u'\|, \|y\|, \|y'\| \leq \Gamma$, we have,

$$|\ell_t(y,u) - \ell_t(y',u')| \leq L\Gamma(\|y-y'\| + \|u-u'\|) \quad \text{and} \quad |\ell_t(y,u)| \leq L\Gamma^2. \tag{8}$$

Regret definition: Let $\mathbf{M}_\star$ be the optimal, in hindsight, DFC policy in the given set $\mathcal{M}$, *i.e.*, $\mathbf{M}_\star = \arg\min_{\mathbf{M}\in\mathcal{M}} \sum_{t=1}^{T} \ell_t(y_t^{\mathbf{M}}, u_t^{\mathbf{M}})$. For ARX systems with strongly convex loss function, the adaptive control algorithm's performance is evaluated by its regret with respect to $\mathbf{M}_\star$ after T steps of interaction and it is denoted as $\textsc{Regret}(T) = \sum_{t=1}^{T} c_t - \ell_t(y^{\mathbf{M}_\star}, u^{\mathbf{M}_\star})$.

The proposed algorithm for the ARX systems with strongly convex cost is given in Algorithm 1. It has two possible approaches depending on the persistence of excitation of given DFC set $\mathcal{M}_r$: explore and commit approach or adaptive control with closed-loop estimate updates.

4.2. Adaptive Control via Explore and Commit Approach

In the explore and commit approach, Algorithm 1 has two phases: an exploration (warm-up) phase with the duration of $T_w = \mathcal{O}(\sqrt{T})$ and an exploitation phase for the remaining $T - T_w$ time-steps.

Warm-up: During the warm-up period, Algorithm 1 applies $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$ in order to recover the Markov parameters of the system. The duration of warm-up T_w is chosen to guarantee reliable estimate of Markov parameters of ARX system and the stability of DFC controllers in exploitation phase. The exact duration of warm-up is given in Appendix C.
Exploitation: At the end of warm-up, Algorithm 1 estimates the Markov parameters of ARX system, $\mathcal{G}$, using the data gathered in warm-up. It deploys the regularized least-squares estimation of (5) to obtain $\widehat{\mathcal{G}}$. At each time-step t, Algorithm 1 uses this estimate and the past inputs to approximate the output uncertainties, $\bar{b}_t(\widehat{\mathcal{G}}) = y_t - \sum_{k=0}^{h-1} \widehat{G}^{k+1}_{u\to y} u_{t-k-1} + \widehat{G}^{k+1}_{y\to y} y_{t-k-1}$. These approximate output uncertainties are then used to execute a DFC policy $\mathbf{M}_t \in \mathcal{M}_r$ as given in (7). Upon applying the control input, the algorithm observes the output of the system along with the loss function $\ell_t(\cdot,\cdot)$ and pays the cost of $c_t = \ell_t(y_t, u_t^{\mathbf{M}_t})$. At each time-step, Algorithm 1 employs the counterfactual reasoning introduced in Simchowitz et al. (2020) to compute a counterfactual loss. Briefly, it considers what the loss would be if the current DFC policy has been applied from the beginning. This provides a noisy metric to evaluate the performance of the current DFC policy. The details of the counterfactual reasoning are in Appendix E. Finally, Algorithm 1 deploys projected online gradient descent on the counterfactual loss to update and keep the DFC policy within the given set $\mathcal{M}_r$ for the next time-step. This process is repeated for the remaining $T - T_w$ time-steps.

Note that deploying DFC policies turns adaptive control problem into an online convex optimization problem which is computationally and statistically efficient. Moreover, using online gradient descent for controller updates exploits the strong convexity grants the following regret rate.

Theorem 3 *Given $\mathcal{M}_r$, a closed, compact and convex set of* DFC *policies, Algorithm 1 with explore and commit approach attains* REGRET$(T) = \tilde{\mathcal{O}}(\sqrt{T})$ *with high probability.*

The proof is in Appendix E. In the proof, we first show that the choice of T_w guarantees that the open-loop data is persistently exciting and the Markov parameter estimates are refined. Then, we show that the estimates of the output uncertainties, the DFC policy inputs and the outputs of the ARX system are bounded. Following the regret decomposition of Theorem 5 of Simchowitz et al. (2020), we show that with the choice of T_w, the regret of running gradient descent on strongly convex losses scales quadratically with the Markov parameters estimation error. This roughly gives REGRET$(T) = \tilde{\mathcal{O}}\left(T_w + (T - T_w)/(\sqrt{T_w})^2\right)$ which is minimized by $T_w = \mathcal{O}(\sqrt{T})$, giving the advertised bound.

4.3. Adaptive Control with Closed-Loop Model Estimate Updates

Prior to describing Algorithm 1 with closed-loop model estimate updates, we need a further condition on the sets $\mathcal{M}$ and $\mathcal{M}_r$, such that the DFC policies in these sets persistently excite the underlying ARX system. The exact definition of the persistence of excitation is given in Appendix B. Note that this condition is mild and briefly implies having a full row rank condition on a significantly wide matrix that maps past e_t to inputs and outputs. One can also show that if a controller satisfies this, then there exists a neighborhood around it that consists of persistently exciting controllers. In the adaptive control with closed-loop model estimates approach, Algorithm 1 also has two phases: a fixed length warm-up phase and an adaptive control phase in epochs.
Warm-up: Algorithm 1 applies $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$ for a fixed duration of τ that solely depends on the underlying system. This phase guarantees the access to a refined first estimate of the system, the persistence of excitation and the stability of the controllers during adaptive control.
Adaptive control in epochs: After warm-up, Algorithm 1 starts controlling the system and operates in epochs with doubling length, *i.e.*, the i'th epoch is of duration $2^{i-1}\tau$ for $i \geq 1$. Unlike the explore

and commit approach, at the beginning of each epoch, it uses all the data gathered so far to estimate the Markov parameters via (5). It then uses this estimate throughout the epoch to approximate the output uncertainties and implement the DFC policies. At each time step, the DFC policies are updated via projected online gradient descent on the computed counterfactual loss. The main difference from the explore and commit approach is that Algorithm 1 updates the model estimates during adaptive control which further refines the estimates and improves the controllers.

Theorem 4 *Given $\mathcal{M}_r$ with* DFC*s that persistently excite the underlying ARX system, Algorithm 1 with closed-loop model estimate updates attains* REGRET$(T) =$ *polylog*(T)*, with high probability.*

The proof is in Appendix E and it follows similarly with Theorem 3. One major difference that allows to achieve the optimal regret rate is the use of data collected during adaptive control to improve the Markov parameter estimates. This approach roughly gives the following decomposition REGRET$(T) = \mathcal{O}(\tau + \text{polylog}(T)\sum_{i=1}^{\log(T)} 2^{i-1}\tau/(\sqrt{2^{i-1}\tau})^2)$. Notice that unlike explore and commit approach, the estimation error decays at each epoch gives the advertised logarithmic regret.

5. Adaptive Control of ARX Systems with Non-Strongly Convex Quadratic Cost

In this section, we present the setting of ARX systems with non-strongly convex quadratic cost and the regret definition that competes against the optimal controller for this setting. Finally, we propose an optimism based adaptive control algorithm with two variants and provide the regret guarantees.

5.1. Adaptive Control Setting

The unknown ARX system belongs to a set $\mathcal{S}$ which consists of systems that are (A, B) and (A, F) controllable and (A, C) observable. The ARX system has quadratic cost on u_t and y_t, *i.e.*, $c_t = y_t^\top Q y_t + u_t^\top R u_t$ where $Q \succeq 0$ and $R \succ 0$, hence non-strongly convex. For this ARX system, the minimum average expected cost problem is given as follows

$$J_\star(\Theta) = \lim_{T\to\infty} \min_{u=[u_1,\ldots,u_T]} \frac{1}{T}\mathbb{E}\left[\sum\nolimits_{t=1}^{T} y_t^\top Q y_t + u_t^\top R u_t\right].$$

Using the average cost optimality equation, one can derive the optimal control law for this problem (Appendix G). The optimal control law of ARX systems, π^*, is a linear feedback policy,

$$u_t^* = K_x^* x_t + K_y^* y_t = -(R + B^\top \mathbf{P} B)^{-1} B^\top \mathbf{P}\left(A x_t + F y_t\right) \tag{9}$$

where $\mathbf{P}$ is the unique positive semidefinite solution to the discrete-time algebraic Riccati equation:

$$\mathbf{P} = C^\top Q C + (A+FC)^\top \mathbf{P}(A+FC) - (A+FC)^\top \mathbf{P} B(R+B^\top \mathbf{P} B)^{-1} B^\top \mathbf{P}(A+FC). \tag{10}$$

Note that π^* is an LDC policy with the optimal minimum average expected cost of $J_\star(\Theta) = \mathrm{Tr}(\Sigma_E(Q + F^\top(\mathbf{P} - \mathbf{P}B(R+B^\top \mathbf{P}B)^{-1}B^\top \mathbf{P})F))$. We assume that the systems in the set $\mathcal{S}$ are *contractible* such that the optimal controller produces contractive closed-loop system dynamics for the state and the output, *i.e.* $\|A + BK_x^*\| \leq \rho < 1$ and $\|F + BK_y^*\| \leq \upsilon < 1$. Finally, the regret measure in this setting is REGRET$(T) = \sum_{t=0}^{T}(c_t - J_*(\Theta))$.

Optimism in the face of uncertainty (OFU) principle: OFU principle has been widely adopted in sequential decision making tasks in order to balance exploration and exploitation. It suggests to

Algorithm 2 Adaptive Control of ARX Systems with Non-Strongly Convex Quadratic Cost

```
Input: ID, $T$, $T_w$, $\tau$, $h$, $S>0$, $\delta>0$, $n$, $m$, $p$, $Q$, $R$, $\rho$, $\upsilon$
if ID = Explore & Commit then Set $T_{\text{warm}} = T_w$, else Set $T_{\text{warm}} = \tau$
—— Warm-Up ——
for $t = 0, 1, \ldots, T_{\text{warm}}$ do
    Deploy $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$ and store $\mathcal{D}_0 = \{y_t, u_t\}_{t=1}^{T_{\text{warm}}}$
—— Adaptive Control ——
for $i = 0, 1, \ldots$ do
    Calculate $\widehat{\mathcal{G}}_i$ via (5) using $\mathcal{D}_i = \{y_t, u_t\}_{t=1}^{2^i T_{\text{warm}}}$
    Deploy SysId-ARX $(h, \widehat{\mathcal{G}}_i, n)$ for $\hat{A}_i, \hat{B}_i, \hat{C}_i, \hat{F}_i$
    Construct $\mathcal{C}_i \coloneqq \{\mathcal{C}_A(i), \mathcal{C}_B(i), \mathcal{C}_C(i), \mathcal{C}_F(i)\}$ s.t. w.h.p. $(A, B, C, F) \in \mathcal{C}_i$
    Find a $\tilde{\Theta}_i = (\tilde{A}_i, \tilde{B}_i, \tilde{C}_i, \tilde{F}_i) \in \mathcal{C}_i \cap \mathcal{S}$ s.t.   $J(\tilde{\Theta}_i) \leq \inf_{\Theta' \in \mathcal{C}_i \cap \mathcal{S}} J(\Theta') + T^{-1}$
    if ID = Explore & Commit then Set $\tilde{\Theta}_i = \tilde{\Theta}_0$ → In E&C, only $\tilde{\Theta}_0$ used for control
    for $t = 2^i T_{\text{warm}}, \ldots, 2^{i+1} T_{\text{warm}} - 1$ do
        Execute the optimal controller for $\tilde{\Theta}_i$
```

estimate the model up to confidence interval and proposes to act according to the optimal controller of the model that has the lowest optimal cost within the confidence interval, *i.e.*, the optimistic model. For adaptive control in this setting, we deploy the controllers designed via OFU principle.

The proposed algorithm for the ARX systems with non-strongly convex quadratic cost is given in Algorithm 2. It has two variants depending on the persistence of excitation of the optimal controller π^*: explore and commit approach or adaptive control with closed-loop estimate updates.

5.2. Adaptive Control via Explore and Commit Approach

Similar to prior setting, in the explore and commit approach, Algorithm 2 has two phases: an exploration (warm-up) phase with the duration of $T_w = \mathcal{O}(T^{2/3})$ and an exploitation phase.
Warm-up: Algorithm 2 uses $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$ for exploration. The exact T_w is given in Appendix D and it guarantees reliable estimation of system parameters and the stability of OFU based controller.

Exploitation: At the end of warm-up, Algorithm 2 estimates the Markov parameters of ARX system via (5) and constructs confidence sets $(\mathcal{C}_A, \mathcal{C}_B, \mathcal{C}_C, \mathcal{C}_F)$ for the system parameters up to similarity transform using SysId-ARX, a variant of Ho-Kalman realization algorithm (Ho and Kálmán, 1966). The procedure follows similarly with SYS-ID of Lale et al. (2020c) and the details are given in Appendix F. Algorithm 2 then deploys the OFU principle and chooses the optimistic system parameters, $\tilde{\Theta}$, that lie in the intersection of the confidence sets and $\mathcal{S}$. Finally, Algorithm 2 constructs the optimal control law for $\tilde{\Theta}$ via (9) and (10) and executes it for the remaining $T - T_w$ time-steps.

Theorem 5 *Given an unknown ARX system with non-strongly convex quadratic cost, Algorithm 2 with explore and commit approach attains* Regret$(T) = \tilde{\mathcal{O}}(T^{2/3})$*, with high probability.*

The proof is in Appendix F. In the proof, we first show that the choice of T_w guarantees persistence of excitation in open-loop data and the stability of inputs and outputs. Then, we derive the Bellman optimality equation for ARX systems which we use for decomposing regret via OFU principle. This roughly gives Regret$(T) = \tilde{\mathcal{O}}\left(T_w + (T - T_w)/\sqrt{T_w}\right)$ which is minimized by $T_w = \mathcal{O}(T^{2/3})$.

5.3. Adaptive Control with Closed-Loop Model Estimate Updates

Before describing Algorithm 2 with closed-loop model estimate updates, we need a further condition such that the optimal controller for the underlying ARX system persistently excited the system. This is again a mild condition and briefly implies that a significantly wide matrix which maps the past e_t to inputs and outputs and formed via optimal controller is full row rank. The precise condition is given in Appendix B. Note that if the system parameter estimates are accurate enough, the controller designed with system parameter estimates persistently excite the ARX system. Similar to strongly convex cost setting, in the adaptive control with closed-loop estimates approach, Algorithm 2 has two phases: a fixed length warm-up phase and an adaptive control in epochs.
Warm-up: Algorithm 2 uses $u_t \sim \mathcal{N}(0, \sigma_u^2 I)$ for a fixed warm-up duration τ which grants refined estimates of the system parameters, persistence of excitation and stability for adaptive control phase.
Adaptive control in epochs: After warm-up, Algorithm 2 starts adaptive control in doubling length epochs, *i.e.*, i'th epoch has the duration of $2^{i-1}\tau$. At the beginning of i'th epoch, it estimates the system parameters via (5), constructs the confidence sets and deploys OFU principle to recover an optimistic model, $\tilde{\Theta}_i$. Finally, it executes the optimal control law for $\tilde{\Theta}_i$ until the end of epoch i. Thus, the main difference from explore and commit approach is the use of closed-loop data to further refine the model estimates. This improves the regret performance and the proof is in Appendix F.

Theorem 6 *Given an unknown ARX system with non-strongly convex quadratic cost whose optimal controller persistently excites the system, Algorithm 2 with closed-loop model estimate updates attains* $\text{Regret}(T) = \tilde{\mathcal{O}}(\sqrt{T})$*, with high probability.*

6. Related Works

System Identification: The classical open or closed-loop system identification methods mostly consider the asymptotic performance of the proposed algorithms or demonstrate positive and negative empirical studies (Verhaegen, 1994; Forssell and Ljung, 1999; Van Overschee and De Moor, 1997; Ljung, 1999). These works mostly consider LQR or LQG systems in their state-space form. However, Chiuso and Picci (2005); Jansson (2003) provide asymptotic studies of closed-loop system identification of LQG systems in predictive form which corresponds to the exact ARX systems formulation of LQG. Moreover, the ARX systems, in particular, have been studied extensively in system identification perspective due to their input-output form (Diversi et al., 2010; Bercu and Vazquez, 2010; Sanandaji et al., 2011; Stojanovic et al., 2016). In these works, the authors discuss the role of persistence excitation in consistent asymptotic recovery of ARX system parameters. On the other hand, the finite-time learning guarantees, which is the focus of this work, are not known.

Adaptive Control: The classical works in adaptive control also study the asymptotic performance of the designed controllers (Lai et al., 1982; Lai and Wei, 1987; Fiechter, 1997). In the ARX systems setting, Prandini and Campi (2000a,b); Campi and Kumar (1998) study the asymptotic convergence to optimal controller of ARX systems using an early interpretation of OFU principle. The current paper is the finite-time counterpart of these studies and completes an important part of the picture in adaptive control of ARX systems by providing optimal regret guarantees. It also extends the prior efforts in adaptive control of LQR and LQG systems in regret minimization perspective to the general ARX systems setting (Abbasi-Yadkori and Szepesvári, 2011; Dean et al., 2018; Abeille and Lazaric, 2018; Agarwal et al., 2019a,b; Cohen et al., 2019; Faradonbeh et al., 2018, 2020a,b; Lale et al., 2020a,b,c; Mania et al., 2019; Simchowitz and Foster, 2020; Simchowitz et al., 2020).

References

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26, 2011.

Marc Abeille and Alessandro Lazaric. Improved regret bounds for thompson sampling in linear quadratic control problems. In *International Conference on Machine Learning*, pages 1–9, 2018.

Naman Agarwal, Brian Bullins, Elad Hazan, Sham M Kakade, and Karan Singh. Online control with adversarial disturbances. *arXiv preprint arXiv:1902.08721*, 2019a.

Naman Agarwal, Elad Hazan, and Karan Singh. Logarithmic regret for online control. In *Advances in Neural Information Processing Systems*, pages 10175–10184, 2019b.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar energy*, 83(10):1772–1783, 2009.

Bernard Bercu and Victor Vazquez. On the usefulness of persistent excitation in arx adaptive tracking. *International Journal of Control*, 83(6):1145–1154, 2010.

Dave P Burke, Simon P Kelly, Philip De Chazal, Richard B Reilly, and Ciarán Finucane. A parametric feature extraction and classification strategy for brain-computer interfacing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1):12–17, 2005.

Marco C Campi and PR Kumar. Adaptive linear quadratic gaussian control: the cost-biased approach revisited. *SIAM Journal on Control and Optimization*, 36(6):1890–1907, 1998.

Marco C Campi and Erik Weyer. Finite sample properties of system identification methods. *IEEE Transactions on Automatic Control*, 47(8):1329–1334, 2002.

Marco C Campi and Erik Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764, 2005.

Alessandro Chiuso and Giorgio Picci. Consistency analysis of some closed-loop subspace identification methods. *Automatica*, 41(3):377–391, 2005.

Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. *arXiv preprint arXiv:1902.06223*, 2019.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 4188–4197, 2018.

Roberto Diversi, Roberto Guidorzi, and Umberto Soverini. Identification of arx and ararx models in the presence of input and output noises. *European Journal of Control*, 16(3):242–255, 2010.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Input perturbations for adaptive regulation and learning. *arXiv preprint arXiv:1811.04258*, 2018.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. On adaptive linear–quadratic regulators. *Automatica*, 117:108982, 2020a.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Optimism-based adaptive regulation of linear-quadratic systems. *IEEE Transactions on Automatic Control*, 2020b.

Barry Fetics, Erez Nevo, Chen-Huan Chen, and David A Kass. Parametric model derivation of transfer function for noninvasive estimation of aortic pressure by radial tonometry. *IEEE Transactions on Biomedical Engineering*, 46(6):698–706, 1999.

Claude-Nicolas Fiechter. Pac adaptive control of linear systems. In *Annual Workshop on Computational Learning Theory: Proceedings of the tenth annual conference on Computational learning theory*, volume 6, pages 72–80. Citeseer, 1997.

Urban Forssell and Lennart Ljung. Closed-loop identification revisited. *Automatica*, 35(7):1215–1241, 1999.

Babak Hassibi, Ali H Sayed, and Thomas Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H2 and H-infinity Theories*, volume 16. SIAM, 1999.

BL Ho and Rudolf E Kálmán. Effective construction of linear state-variable models from input/output functions. *at-Automatisierungstechnik*, 14(1-12):545–548, 1966.

Kuang Yu Huang and Chuen-Jiuan Jane. A hybrid model for stock market forecasting and portfolio selection based on arx, grey system and rs theories. *Expert systems with applications*, 36(3): 5387–5392, 2009.

Magnus Jansson. Subspace identification and arx modeling. *IFAC Proceedings Volumes*, 36(16): 1585–1590, 2003.

Thomas Kailath, Ali H Sayed, and Babak Hassibi. Linear estimation, 2000.

Tze Leung Lai and Ching-Zong Wei. Asymptotically efficient self-tuning regulators. *SIAM Journal on Control and Optimization*, 25(2):466–481, 1987.

Tze Leung Lai, Ching Zong Wei, et al. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 10(1): 154–166, 1982.

Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Logarithmic regret bound in partially observable linear dynamical systems. *arXiv preprint arXiv:2003.11227*, 2020a.

Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Regret minimization in partially observable linear quadratic control. *arXiv preprint arXiv:2002.00082*, 2020b.

Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Regret bound of adaptive control in linear quadratic gaussian (lqg) systems. *arXiv preprint arXiv:2003.05999*, 2020c.

Lennart Ljung. System identification. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–19, 1999.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalent control of lqr is efficient. *arXiv preprint arXiv:1902.07826*, 2019.

Sandra J Norquay, Ahmet Palazoglu, and JoséA Romagnoli. Model predictive control based on wiener models. *Chemical Engineering Science*, 53(1):75–84, 1998.

Samet Oymak and Necmiye Ozay. Non-asymptotic identification of lti systems from a single trajectory. *arXiv preprint arXiv:1806.05722*, 2018.

Maria Prandini and Marco C Campi. Adaptive lqg control of input-output systems—a cost-biased approach. *SIAM Journal on Control and Optimization*, 39(5):1499–1519, 2000a.

MARIA Prandini and MC Campi. A self-optimizing adaptive lqg control scheme for input-output systems. In *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187)*, volume 2, pages 1110–1115. IEEE, 2000b.

Borhan M Sanandaji, Tyrone L Vincent, Michael B Wakin, Roland Tóth, and Kameshwar Poolla. Compressive system identification of lti and ltv arx models. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 791–798. IEEE, 2011.

Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite-time system identification for partially observed lti systems of unknown order. *arXiv preprint arXiv:1902.01848*, 2019.

Max Simchowitz and Dylan J Foster. Naive exploration is optimal for online lqr. *arXiv preprint arXiv:2001.09576*, 2020.

Max Simchowitz, Ross Boczar, and Benjamin Recht. Learning linear dynamical systems with semiparametric least squares. *arXiv preprint arXiv:1902.00768*, 2019.

Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. *arXiv preprint arXiv:2001.09254*, 2020.

Vladimir Stojanovic, Novak Nedic, Dragan Prsic, and Ljubisa Dubonjic. Optimal experiment design for identification of arx models with constrained output in non-gaussian noise. *Applied Mathematical Modelling*, 40(13-14):6676–6689, 2016.

Anastasios Tsiamis and George J Pappas. Finite sample analysis of stochastic system identification. *arXiv preprint arXiv:1903.09122*, 2019.

Peter Van Overschee and Bart De Moor. Closed loop subspace system identification. In *Proceedings of the 36th IEEE Conference on Decision and Control*, volume 2, pages 1848–1853. IEEE, 1997.

Michel Verhaegen. Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.

Dante Youla, Hamid Jabr, and Jr Bongiorno. Modern wiener-hopf design of optimal controllers–part ii: The multivariable case. *IEEE Transactions on Automatic Control*, 21(3):319–338, 1976.