# ECSE 506 Final Project
## Paper Summary: Adaptive Controllers Over Finite Parameter Sets

Amro Al Baali
260629840

April 27, 2020

## 1 Introduction

Control theory has been around to answer questions about controlling a system. Some systems, whether physical or digital, need to be controlled in order to perform the task set by the user (e.g., minimize a cost function). Control theory is used to determine control policies; that is, control actions that should be taken by the decision maker (i.e., the controller) at a given state of the system.

Classical control theory requires a mathematical model of the system to be controlled [1]. It is not feasible to derive or obtain a perfect model of the system. However, since "all models are wrong, but some models are useful" [2, p. 424], simplifying the system can be deemed useful. A nonlinear system can be modelled as linear (even though there is no truly linear system), a stochastic system can be simplified as deterministic (even though the parameters are almost never known deterministically), or even a parameter can be assumed to be a constant even though it can be changing over time.

Branches of control theory emerged to tackle these simplifications and improve controller performances. For instance, nonlinear control deals with (highly) nonlinear systems [3], stochastic control deals with the existence of uncertainty either in observations or in the process model [4], robust control was introduced as a tool for achieving satisfactory performances for family of plant models (this is done by specifying a nominal model and a size of uncertainty around the parameters used) [5], while adaptive control deals with systems with parameters which vary, or are initially uncertain [5]. This document is a summary and a critical evaluation of [6] that introduces a specific type of adaptive controllers that operates over finite parameter sets and [7] that extends the range of systems for which such controller is optimal.

Kumar introduces and discusses a novel adaptive controller in [6] and then in [7] the author extends the range of systems for which the identification and control scheme is "optimal". Thus, the contributions in [7] are mainly theoretical and consists of theorems and proofs without discussing in depth the practicality of such adaptive controller. Therefore, in this document, the adaptive controller introduced in [6] is discussed first and an example is presented before discussing the theory introduced in [7].

The outline of this summary is as follows. The adaptive controller introduced in [6] is discussed in Section 2.2 with some theoretical results followed by a discussion of the extended claims in [7] in Section 3. An example is provided in Section 4 to demonstrate the results and a critical evaluation of the contribution is presented in Section 5 followed by a concluding remarks in Section 6.

# 2 Background

Adaptive control covers a set of techniques which provide a systematic approach for automatic adjustment of controllers in real time [5], in order to achieve or maintain a desired level of control system performance when the parameters of the dynamical model are unknown and/or change in time. The problem of adaptive control can be broken down into two parts: system identification and system control. For example, an aircraft changes its mass (as fuel burns) as it flies through the sky, therefore, an adaptive controller must first estimate the mass of the aircraft, and then control the aircraft; the controller accounts or *adapts* to the changing parameter (mass of the aircraft).

The results of [7] considerably extend the range of systems for which the identification and control scheme of [6,8] is known to be "optimal". Since such results are theoretical and rely on the adaptive controller presented in [8], then the focus of this section is to introduce this controller.

The assumptions on the system discussed in [6] are as follows

i) The dynamical system is a Markov chain model with finite state and control spaces, $\mathcal{X}$ and $\mathcal{U}$, respectively;

ii) The parameter set $\mathcal{A}$ is finite;

iii) The performance of the system to be controlled is measured by the average of the costs incurred over an infinite operating time period (defined in (6));

iv)
- Either $p(i, j; u, \alpha) > 0$ for all $(u, \alpha) \in \mathcal{U} \times \mathcal{A}$,
- or $p(i, j; u, \alpha) = 0$ for all $(u, \alpha) \in \mathcal{U} \times \mathcal{A}$.

However, the assumptions are relaxed in [7] to become

i) The state and control spaces, $\mathcal{X}$ and $\mathcal{U}$, respectively, are Polish spaces;

ii) The parameter set $\mathcal{A}$ is finite;

iii) the underlying cost criterion can be either a discounted cost criterion on a long-term average cost criterion;

iv)
- Either $p(i, j; u, \alpha) > 0$ for all $(u, \alpha) \in \mathcal{U} \times \mathcal{A}$,
- or $p(i, j; u, \alpha) = 0$ for all $(u, \alpha) \in \mathcal{U} \times \mathcal{A}$.

## 2.1 Description of the adaptive controller

For each $\alpha \in \mathcal{A}$, the long-term average cost problem (9) is solved to obtain

1. an optimal feedback control law $\phi(\cdot, \alpha) : \mathcal{X} \to \mathcal{U}$ for each model $\alpha \in \mathcal{A}$, and
2. the optimal long-term average cost $J^*(\alpha)$ achievable for each model $\alpha \in \mathcal{A}$.

The user has access to choosing the following parameters

1. any integer $m$ such that $m > m_{ij}$ for every $(i, j) \in \mathcal{X} \times \mathcal{X}$ where $\{m_{ij}\}$ are defined as follows. For every $(i, j) \in \mathcal{X} \times \mathcal{X}$ there exists an integer $m_{ij}$ and a sequence of states $i = k_0, \ldots, k_{m_{ij}} = j$

such that

$$p(k_s, k_{s+1}; u, \alpha) \quad \forall\ (u, \alpha) \in \mathcal{U} \times \mathcal{A} \text{ and } s = 0, 1, \ldots, m_{ij} - 1. \tag{1}$$

2. Any strictly monotone increasing function $f$ such that

$$f(J^*(\alpha)) > 0, \qquad \alpha \in \mathcal{A}. \tag{2}$$

3. Any positive valued function $o$ such that $\lim_{t\to\infty} o(t) = \infty$ while $\lim_{t\to\infty} o(t)/t = 0$.

At equally spaced times, $t = 0, m, 2m, \ldots, km, \ldots$, an estimate $\hat{\alpha}_t$ of the unknown parameter is chosen such that it maximizes the "biased" likelihood function

$$\bar{D}_t(\alpha) := [f(J^*(\alpha))]^{-o(t)} \prod_{s=0}^{t-1} p(x_s, x_{s+1}; u_s, \alpha) \tag{3}$$

over all $\alpha \in \mathcal{A}$. It should be noted that without the first term in (3) the function would simply be the likelihood function thus making the estimator an ML estimator. This estimation criterion can be interpreted as a "mild" bias in favor of models with lower costs.

The estimate is not updated between times $km$ and $(k + 1)m$. That is,

$$\hat{\alpha}_t = \hat{\alpha}_{km} \qquad km \leq t \leq (k + 1)m. \tag{4}$$

The control applied at time $t$ is

$$u_t = \phi(x_t, \hat{\alpha}_t). \tag{5}$$

It should be noted that the control policy $\phi(\cdot, \alpha)$ is calculated for every $\alpha \in \mathcal{A}$ *a prior* using classical methods such as policy improvement, value iteration, etc. Thus, computing the control action $u_t$ is simply a look-up function. This implies that what makes an adaptive controller "good" is the estimation part. That is, better estimator of $\alpha_t$ directly imply that the adaptive controller is better.

## 2.2 Cost function

The goal is to design an adaptive controller which, at each time $t$, chooses a control input $u_t$, based on the observed past history $(x_0, u_0, x_1, u_1, \ldots, x_t)$ of the system in such a way that the long-term average cost (6) incurred is (almost surely) a minimum. The long-term average cost is defined in Definition 2.1.

**Definition 2.1** (*Long-term average cost*)**.** The long-term average cost incurred $J$ is defined as

$$J(\psi, \alpha) = \lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} c\left(x_s, x_{s+1}, u_s\right) \tag{6}$$

$$= \lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} c\left(x_s, x_{s+1}, \psi(x_s)\right). \tag{7}$$

Computing the long-term average cost using (6) is infeasible. Rather, Theorem 3.4 is used to derive a more feasible way to compute it. Specifically,

$$J(\psi,\alpha) = \sum_{x\in\mathcal{X}} \pi(x,\psi,\alpha) \sum_{y\in\mathcal{X}} p(x,y;\psi(x),\alpha)c(x,y,\psi(x)), \tag{8}$$

where $\pi$ is vector of stationary transition probabilities which is defined in Definition 2.2. The optimal cost for model $\alpha$ is given by

$$\begin{aligned} J^*(\alpha) &= \min_{\psi:\mathcal{X}\to\mathcal{U}} J(\psi,\alpha) &(9)\\ &= J(\phi,\alpha), &(10)\\ \phi(\alpha) &= \arg\min_{\psi:\mathcal{X}\to\mathcal{U}} J(\psi,\alpha). &(11) \end{aligned}$$

**Lemma 2.1.** The control law $\phi(\cdot,\alpha) : \mathcal{X} \to \mathcal{U}$ achieving the minimum in (9) is an optimal feedback control law for the parameter $\alpha$.

Policy improvement algorithm and linear programming are some of the procedures that are used to compute the values of $J^*(\alpha)$ and $\phi(\cdot,\alpha)$.

**Definition 2.2** (*Stationary transition probabilities*)**.** Let $\psi : \mathcal{X} \to \mathcal{U}$ be an arbitrary feedback control law. Let $\mathbf{P}(\psi,\alpha)$ be the matrix of closed-loop transition probabilities defined as $\mathbf{P}_{(i,j)} := p(i,j;\psi(i),\alpha)$. Let $\mathbf{1} := \begin{bmatrix}1 & \dots & 1\end{bmatrix}^\mathsf{T}$. Then there exists a *stationary transition probability* vector $\boldsymbol{\pi}(\psi,\alpha) := \left[\pi(1,\psi,\alpha),\dots,\pi(|\mathcal{X}|,\psi,\alpha)\right]$ such that

$$\begin{aligned} \lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}\mathbf{P}^s(\psi,\alpha) &= \mathbf{1}\boldsymbol{\pi}(\psi,\alpha) &(12)\\ &= \begin{bmatrix} \pi(1,\psi,\alpha) & \cdots & \pi(|\mathcal{X}|,\psi,\alpha) \\ \vdots & \vdots & \vdots \\ \pi(1,\psi,\alpha) & \cdots & \pi(|\mathcal{X}|,\psi,\alpha). \end{bmatrix}. &(13) \end{aligned}$$

$\boldsymbol{\pi}(\psi,\alpha)$ satisfies

1. $\sum_{x\in\mathcal{X}} \pi(x,\psi,\alpha) = 1$, and
2. $\boldsymbol{\pi}(\psi,\alpha) = \boldsymbol{\pi}(\psi,\alpha)\mathbf{P}(\psi,\alpha)$ which can also be written as

$$\mathbf{P}(\psi,\alpha)^\mathsf{T}\boldsymbol{\pi}(\psi,\alpha)^\mathsf{T} = \boldsymbol{\pi}(\psi,\alpha)^\mathsf{T}. \tag{14}$$

In other words, (14) implies that $\boldsymbol{\pi}(\psi,\alpha)$ is the transpose of the eigenvector of $\mathbf{P}(\psi,\alpha)^\mathsf{T}$, or equivalently, it's the left-eigenvector of $\mathbf{P}(\psi,\alpha)$.

Table 1 summarizes the notations used in this document.

# 3 Summary of results

The main results of [7] are

| Symbol | Definition |
|---|---|
| $\mathcal{X}$ | State space |
| $\mathcal{U}$ | Control space |
| $\mathcal{A}$ | Parameter space |
| $\alpha$ | Parameter to be estimated |
| $\alpha^\circ$ | True parameter |
| $\hat{\alpha}$ | Estimated parameter |
| $x$ | System states |
| $u$ | Control input |
| $p(i, j; u, \alpha)$ | Probability of transfer from state $i$ to state $j$ under the action of $u$ |
| $c(i, j, u)$ | One-stage cost of going from state $i$ to state $j$ under the action $u$ |
| $\psi(\cdot, \alpha)$ | $\mathcal{X} \to \mathcal{U}$ Control law for parameter $\alpha$ |
| $\phi(\cdot, \alpha)$ | $\mathcal{X} \to \mathcal{U}$ Optimal control law for parameter $\alpha$ |
| $J(\psi, \alpha)$ | Long-term average cost |
| $J^*(\psi, \alpha)$ | Optimal long-term average cost $J(\phi, \alpha)$ |
| $1(\cdot)$ | Identifier function ($1(\text{true}) = 1$, $1(\text{false}) = 0$) |

Table 1: Notations used in this paper.

i) Ever Cesaro-limit point of the parameter estimates is "equivalent" to the true but unknown parameter;

ii) the adaptive control law Cesaro-converges to the set of optimal control laws for a variety of cost criteria;

iii) if the cost criterion is of the long-term average type, then the cost actually incurred by using the adaptive controller is optimal and cannot be improved even if one knew the value of the unknown parameter at the start.

The core of [6] is in introducing a novel estimator by pre-multiplying the likelihood function by the term $[f(J^*(\alpha))]^{-o(t)}$. The other novel contribution is the introduction of the $m$ term which is the time interval at which the same estimate $\hat{\alpha}$ is used to determine the control policy. In [7], the author generalizes the systems for which such controller is optimal. Therefore, the paper mostly consists of proofs of convergence properties. The rest of this section summarizes the theorems without including the proofs as the proofs are long.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space of trajectories of the system where $\Omega$ is the sample space, $\mathcal{F} := \sigma(x_0, u_0, \ldots, x_t)$ is the $\sigma$-algebra generated by the history up to time $t$, and $\mathbb{P}$ the probability measure induced on $\Omega$ by the adaptive control law applied to the true parameter $\alpha^\circ$.

**Theorem 3.1.** There exists a set $N \subseteq \Omega$, $\mathbb{P}(N) = 0$ such that if $\omega \in N^{\mathrm{C}}$ and if $\alpha^*$ satisfies $\limsup_{t\to\infty} 1/t \sum_{s=0}^{t-1} 1(\hat{\alpha}_s(\omega) = \alpha^*) > 0$, then

$$p(x, y, \phi(x, \alpha^*), \alpha^*) = p(x, y, \phi(x, \alpha^*), \alpha^\circ) \qquad (x, y) \in \mathcal{X} \times \mathcal{X}. \tag{15}$$

$\triangle$

**Theorem 3.2.** There exists a set $N \subset \Omega$, $\mathbb{P}(N) = 0$ such that if $\omega \in N^{\mathrm{C}}$ and if $\alpha^*$ is any limit point of $\{\hat{\alpha}_t(\omega)\}_{t=1}^{\infty}$, then

$$J(\alpha^*) \leq J(\alpha^\circ). \tag{16}$$

$\triangle$

The idea of the adaptive controller is that the designer chooses $J(\cdot)$ in such a way that if any $\alpha^*$ satisfies the two properties

i) $J(\alpha^*) \leq J(\alpha^\circ)$

ii) $p(x, y, \phi(x, \alpha^*), \alpha^*) = p(x, y, \phi(x, \alpha^*), \alpha^\circ)$,

then $\phi(\cdot, \alpha^*) : \mathcal{X} \to \mathcal{U}$ is a “good” feedback control law for the true parameter $\alpha^\circ$.

**Theorem 3.3.** Let $\Phi = \{\phi_\alpha : \alpha \in \mathcal{A}, J(x, \phi_\alpha, \alpha^\circ) = J(x, \phi_{\alpha^\circ}, \alpha^\circ)\ \forall (x, y) \in \mathcal{X} \times \mathcal{X}\}$ be the set of control laws which are best for parameter $\alpha^\circ$, then

$$\lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} 1(\phi_{\hat{\alpha}_s} \in \Phi) = 1 \qquad \text{a.s.}, \tag{17}$$

where $1(\cdot)$ is the identifier function. $\triangle$

**Theorem 3.4.** Consider a system with a Polish state space $\mathcal{X}$ for which

$$\operatorname{prob}(x_{t+1} \in \mathcal{B} |\ x_t) = \int_{\mathcal{B}} p(x_t, y)\mu(\mathrm{d}y), \tag{18}$$

where $\mu$ is a nonnegative Borel measure on $\mathcal{X}$, $p : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is continuous and dominated by $q(y) \geq p(x, y)$ for all $x$ with $\int_{\mathcal{X}} q(y)\mu(\mathrm{d}y) < \infty$. It is assumed that there is an integer $m$ such that for every $x \in \mathcal{X}$ and $0 \subseteq \mathcal{X}$ open, there is an integer $n$ (depending possibly on $x$ and $0$) $n < m$, with

$$\operatorname{prob}(x_{t+n} \in 0 |\ x_t = x) > 0. \tag{19}$$

If $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is bounded, Borel measurable, with $c(x, y)$ continuous in $x$ for every fixed $y$, then

i) there exists a unique $\pi(\cdot)$ for which

$$\int_{\mathcal{X}} \pi(x)\mu(\mathrm{d}x) = 1 \tag{20}$$

and

$$\int_{\mathcal{X}} \pi(x)p(x, y)\mu(\mathrm{d}x) = \pi(y) \qquad y \in \mathcal{X}. \tag{21}$$

Also,

$$\int_{0} \pi(x)\mu(\mathrm{d}x) > 0 \qquad 0 \subseteq \mathcal{X}; \tag{22}$$

ii)

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}c(x_s,x_{s+1})=\int_{\mathcal{X}}\pi_x(x)\int_{\mathcal{X}}p(x,y)c(x,y)\mu(\mathrm{d}y)\mu(\mathrm{d}x)\qquad \text{a.s.}\tag{23}$$

iii) There exists a bounded continuous $w(\cdot)$ such that

$$w(x)=\int_{\mathcal{X}}\pi(y)d(y)\mu(\mathrm{d}y)\tag{24}$$

$$=d(x)+\int_{\mathcal{X}}p(x,y)w(y)\mu(\mathrm{d}y)\qquad \forall x\in\mathcal{X},\tag{25}$$

where

$$d(x):=\int_{\mathcal{X}}p(x,y)c(x,y)\mu(\mathrm{d}y).\tag{26}$$

△

Theorem 3.4 is used in deriving the feasible long-term average cost function (8).

**Theorem 3.5.** i) Let $\Phi:=\{\phi_\alpha:\ \alpha\in\mathcal{A},J(\phi_\alpha,\alpha^\circ)=J(\phi_{\alpha^\circ},\alpha^\circ)\}$ be the set of all control laws which are best for $\alpha^\circ$. Then

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}1(\phi_{\hat{\alpha}_s}\in\Phi)=1\qquad \text{a.s.}\tag{27}$$

ii)

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=0}^{t-1}c(x_s,x_{s+1},\phi(x_s,\hat{\alpha}_s))=J(\phi_{\alpha^\circ},\alpha^\circ)\qquad \text{a.s.}\tag{28}$$

△

Theorem 3.5 is the core of the paper. It says that the control policy $\phi$ obtained by using the estimated parameter $\hat{\alpha}$, $\phi_{\hat{\alpha}}$ is in the set of all control laws which are best for $\alpha^\circ$ almost surely (with probability 1). That is, with probability 1, the control policy using the estimated parameter $\hat{\alpha}$ matches the control policy using the true parameter $\alpha^\circ$. Furthermore, the theorem also says that, in the limit $t\to\infty$, the long-term average cost of the system while using the estimated parameter $\hat{\alpha}$ converges to the long-term average of the system using the true parameter $\alpha^\circ$.

# 4 Example

In this section, two numerical examples are presented that demonstrates the ability of such adaptive controller to give optimal results almost surely. The Example 4.1 is from [6] demonstrating a Markov decision process (MDP), while Example 4.2 is estimating a parameter in a linear system.

**Example 4.1.** Consider the system with spaces $\mathcal{X}=\{1,2\}$, $\mathcal{U}=\{1,2\}$ with parameter set $\mathcal{A}=$

$\{1,2,3\}$ with $\alpha^\circ$ being the true parameter. Further, let

$$p(1,1;1,1) = p(1,1;1,3) = 0.5 \tag{29}$$
$$p(1,1;1,2) = 0.9 \tag{30}$$
$$p(1,1;2,1) = p(1,1;2,2) = 0.8 \tag{31}$$
$$p(1,1;2,3) = 0.2 \tag{32}$$
$$p(2,1;u,\alpha) = 1 \qquad \forall (u,\alpha) \in \mathcal{U} \times \mathcal{A} \tag{33}$$
$$c(i,j,u) = 3 + (2-i)(7.8 - 0.3u - 0.6). \tag{34}$$

Using MATLAB[1], the optimal cost function is computed as outlined in Section 2.2 to give

$$J^*(1) = 2 \tag{35}$$
$$J^*(2) = 3 \tag{36}$$
$$J^*(3) = 1. \tag{37}$$

The optimal control policy was then determined to be

$$\phi(i,1) = 1 \qquad \forall i \in \mathcal{X} \tag{38}$$
$$\phi(i,2) = \phi(i,3) = 2 \qquad \forall i \in \mathcal{X}. \tag{39}$$

The auxiliary functions were arbitrarily chosen as

$$f(y) = y \tag{40}$$
$$o(t) = \sqrt{t} \tag{41}$$

which satisfy the conditions outlined in Section 2.1 and $m = 1$.

First, it will be demonstrated that the ML estimator fails badly as follows. Let $x_0 = 1$ and $u_0 = 1$. Assume the next state is then $x_1 = 1$ (this happens with probability 0.5). Then since $p(1,1;1,2) = 0.9$ while $p(1,1;1,1) = p(1,1;1,3) = 0.5$, the ML estimate will be $\hat{\alpha}_1^{\mathrm{ML}} = 2$. From (39) it follows that the control input is then $u_1 = 2$. Since $p(i,j;2,1) = p(i,j;2,2)$ then it follows that

$$\prod_{s=0}^{1} p\left(x_s, x_{s+1}; u_s, 1\right) < \prod_{s=0}^{1} p\left(x_s, x_{s+1}; u_s, 2\right) \tag{42}$$

and hence, the ML estimate at $t = 2$ will not be 1 (i.e., $\hat{\alpha}_2^{\mathrm{ML}} \in \{2,3\}$). From (39), this would imply that the next control input is $u_2 = 2$. Thus, it can be seen by induction that the control input will remain at the control input $u_t = 2$ indefinitely; this can be seen in Figure 1. This is a non-optimal control input as demonstrated by the average cost function in Figure 2. On the other hand, the optimal adaptive controller performs better and its performance is close to that of a controller using the true parameter $\alpha = \alpha^\circ$ as can be seen by Figure 2.

The optimal estimator estimates the parameter $\alpha$ to be the true value $\alpha^\circ = 1$ with probability 1 (almost surely). This is demonstrated in Figure 3. Thus, the optimal adaptive controller maintains a control input of $u_t = 1$ when $\hat{\alpha} = 1$ (i.e., $u_t = 1$ *almost surely*). The optimal controller picks

[1] The code used is attached with this document.

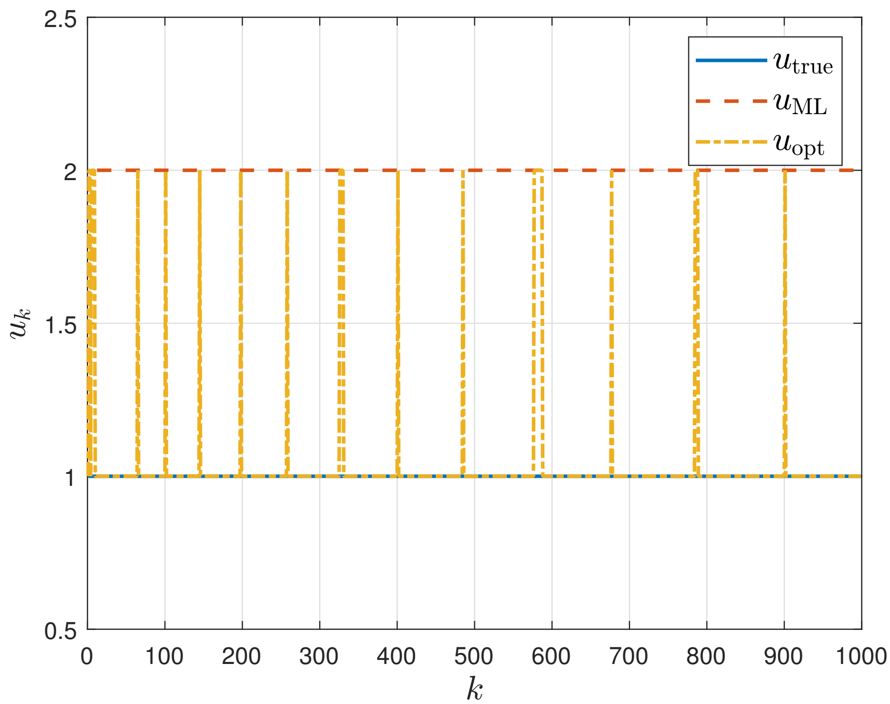


Figure 1: Example 4.1: Control inputs using the optimal control policy $u_k = \phi(x_k, \hat{\alpha})$ for $\hat{\alpha}$ computed using either ML or the optimal estimators. $u_{\text{true}} = 1$, $u_{\text{ML}} = 2$, while $u_{\text{opt}}$ maintains a value of 2 most of the time.

$u = 2$ when the estimate $\hat{\alpha} \neq \alpha^{\circ}$ which happens with probability 0. This can be seen by the peaks in Figure 1; this phenomenon does not contradict the theory as the claim is that the controller will choose the optimal control input *almost surely* (with probability 1). The long-time average cost of the system using the optimal adaptive controller is close to the long-time average cost of the system using the optimal controller while knowing the true parameter value $\alpha = \alpha^{\circ}$ as can be seen in Figure 2.

**Example 4.2.** Consider a mass-spring system given by the equation

$$\dot{x} + \alpha x = u, \tag{43}$$

where $x$ is the displacement of the mass with respect to the origin, $\alpha$ is the unknown spring constant (parameter), and $u$ is the external force applied on the system. Discretizing the system using forward Euler gives

$$\begin{aligned} x_{k+1} &= f(x_k, u_k; \alpha) && (44) \\ &= \text{round}\left((1 - T\alpha)x_k + Tu_k\right), && (45) \end{aligned}$$

where $T \in \mathbb{R}_+$ is the discretization step and round is a rounding function that ensures $x_{k+1} \in \mathcal{X}$.

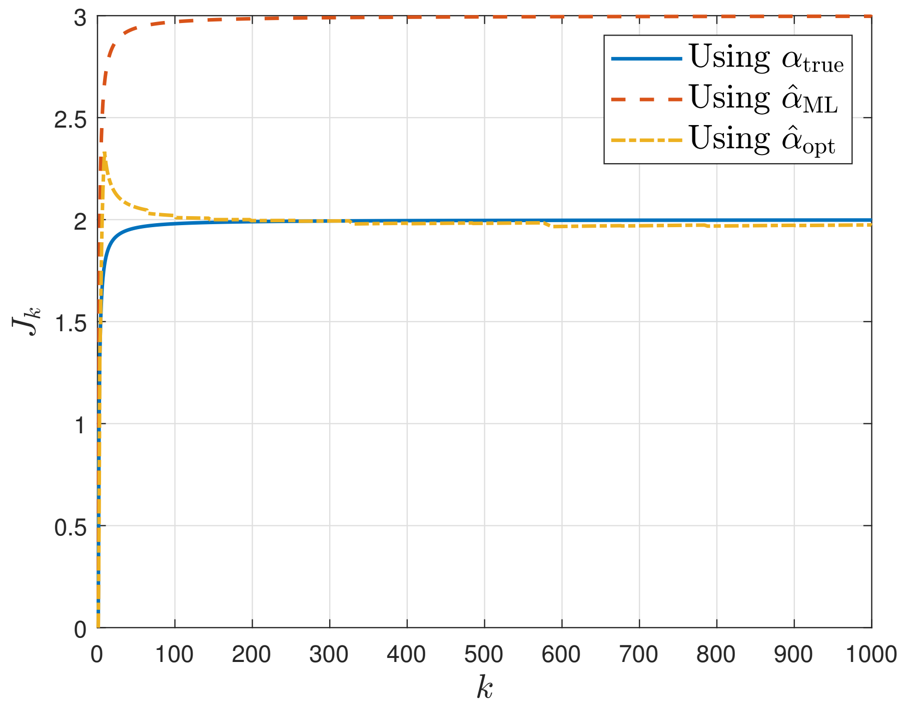


Figure 2: Example 4.1: Long-term average cost using the optimal control policy $u_k = \phi(x_k, \hat{\alpha})$ for $\hat{\alpha}$ computed using either ML or the optimal estimators.

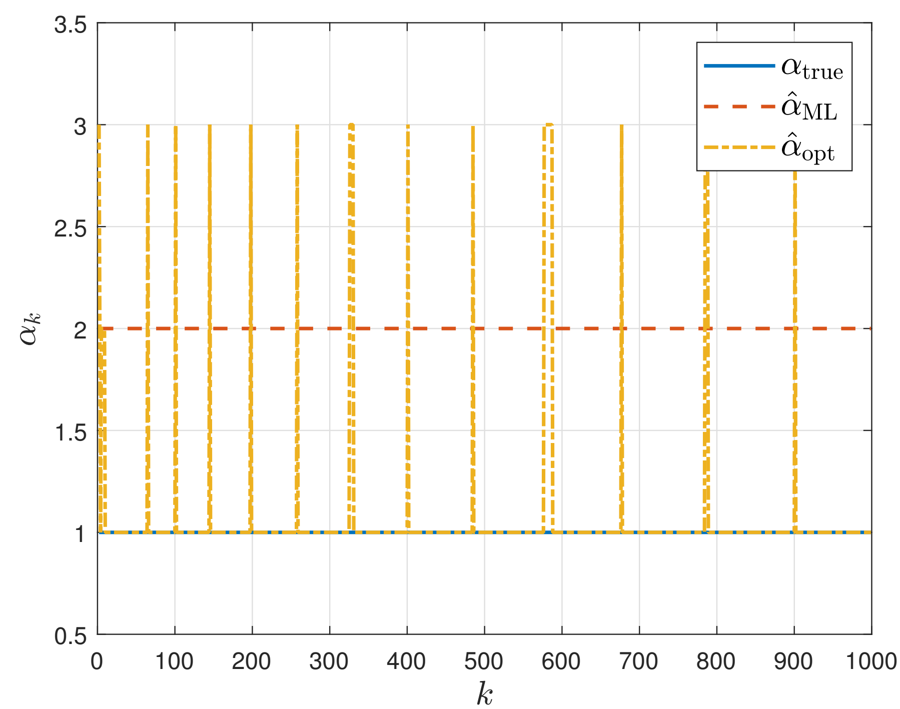


Figure 3: Example 4.1: Progression of the parameter estimates $\hat{\alpha}$ using either ML or the optimal estimators. $\alpha_{\text{true}} = \alpha^\circ = 1$.

For this numerical example, the spaces are given by

$$\mathcal{X} = \{-1, 0, 1\} \tag{46}$$
$$\mathcal{U} = \{-1, 0, 1\} \tag{47}$$
$$\mathcal{A} = \{0, 0.5, 1\}, \tag{48}$$

while the true parameter is $\alpha^\circ = 0.5$. Furthermore, the discretization step is $T = 1$. The goal of the decision maker is to maintain the mass position at $x_{\text{target}} = 1$. This goal is translated to the following cost function

$$c_t(x_s, x_{s+1}, \alpha) = |x_s - x_{\text{target}}|, \tag{49}$$

where 1-norm $|\cdot|$ was chosen arbitrarily. It should be noted that choosing other distance measures such as the squared distance would not make a difference because of the special structure of the state space (46).

The probability transition is given by

$$p(x_k, x_{k+1}; u_k, \alpha) = \begin{cases} 0.8, & \text{if } x_{k+1} = f(x_k, u_k, \alpha^\circ), \\ 0.2, & \text{otherwise.} \end{cases} \tag{50}$$

The auxiliary functions are given by

$$f(y) = y, \tag{51}$$
$$o(t) = \sqrt{t}. \tag{52}$$

Value iteration algorithm was used to compute the optimal control policies. Other than that, the algorithm is the same as the one used in Example 4.1.

As mentioned earlier, the goal is to maintain the mass position at $x_{\text{target}} = 1$. If the control input is zero, then the spring will pull the mass back to zero (round ensures that the system does not get stuck in a limit cycle of $x_k = 1$). Therefore, the optimal control input is $u_k = 1$ at all times. Figure 4 shows that both the ML-based and the optimal adaptive controllers perform well. Albeit, the ML-based adaptive controller performs better at the beginning. However, the ML estimator fails drastically in estimating $\alpha$ as can be seen in Figure 6; it estimates it to be 0. The long-term cost function plot in Figure 5 shows how the optimal adaptive controller performs better than the ML-based adaptive controller.

## 5 Evaluation of the contribution

This document was supposed to serve as the summary for [7]. However, the paper was an extension to the adaptive controller introduced in [6]. Therefore, I had to review the paper that introduces the adaptive controller [8] before reviewing the extended version. I will be evaluating both papers [6] and [7] as two parts of a single contribution.

The introduction of the adaptive controller boils down to the novel estimator that maximizes the modified likelihood function (3). There was no novelty in the controller itself. Rather, most of the focus was on proving that the controller is optimal in the sense that it would perform equally well as if the parameter was known *a priori*. The rest of the paper presents theorems about the

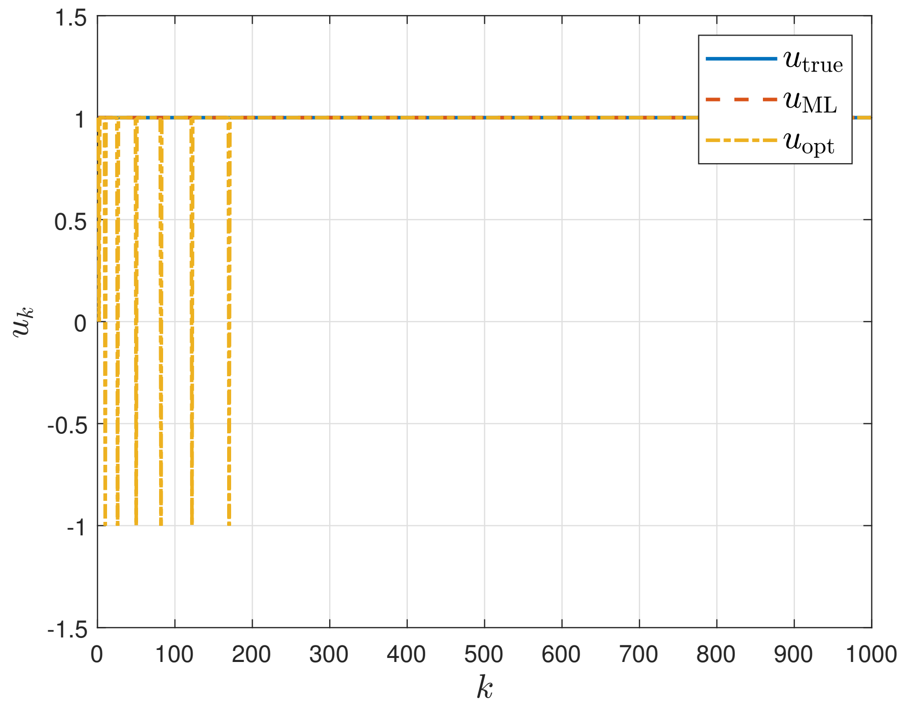


Figure 4: Example 4.2: Control inputs using the optimal control policy $u_k = \phi(x_k, \hat{\alpha})$ for $\hat{\alpha}$ computed using either ML or the optimal estimators.

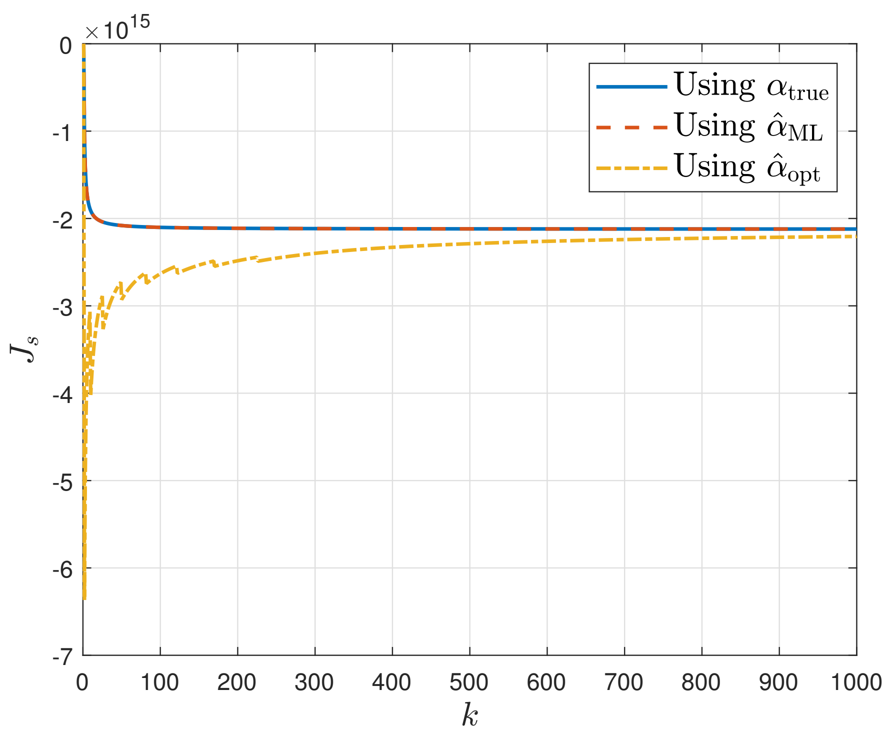


Figure 5: Example 4.2: Long-term average cost using the optimal control policy $u_k = \phi(x_k, \hat{\alpha})$ for $\hat{\alpha}$ computed using either ML or the optimal estimators.

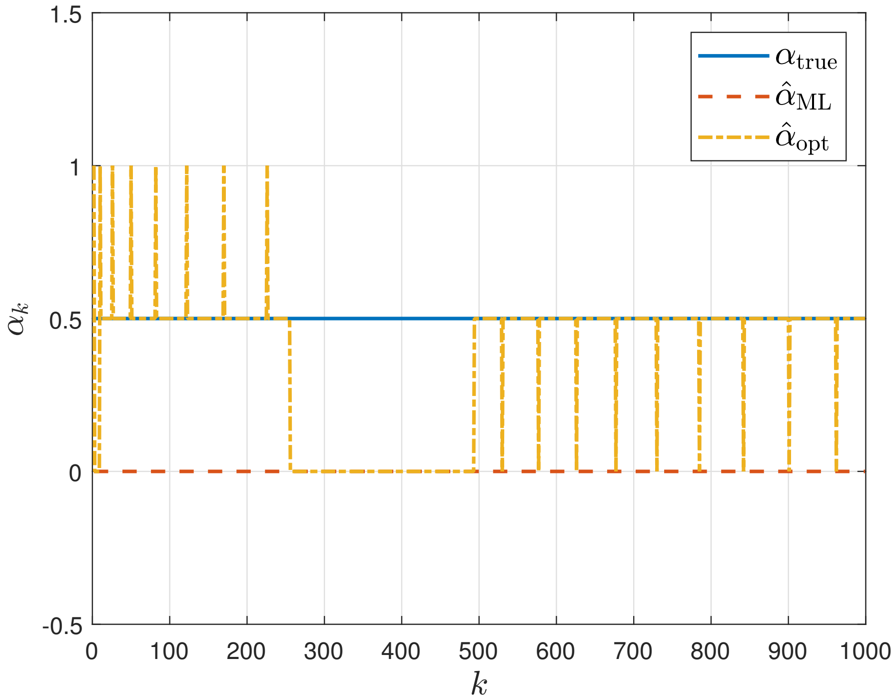


Figure 6: Example 4.2: Progression of the parameter estimates $\hat{\alpha}$ using either ML or the optimal estimators. $\alpha_{\text{true}} = \alpha^{\circ} = 1$.

convergence properties of the controller.

The assumptions play an important role in evaluating a contribution or a paper. The assumption that the parameter set $\mathcal{A}$ is finite is reasonable especially for controllers implemented on digital computers. Even in the case that the unknown parameter is continuous (i.e., $\mathcal{A}$ is infinite but countable), the parameter space can be discretized and quantized and the assumption would hold again. However, the assumption on the cost function (being a discounted type or of a long-term average type) is somewhat restrictive.

The adaptive controller performs well for systems that satisfy the outlined assumptions. However, if the any of the state, control, or parameter spaces are large (still finite), then the memory requirements grow and the adaptive controller might be deemed infeasible. This might be the case for physical systems described by discretized dynamical models. The parameters are likely to be continuous, thus the user has to trade between the accuracy of the estimated parameter and the memory requirements of the system.

The auxiliary functions $f$ and $o$ that pop up in the optimal estimator (3) must meet certain conditions. I thought these conditions were outlined to help in the proofs. However, it was not clear how to choose these functions. Furthermore, it is not clear how the choice of these functions would affect the performance of the controller. In the example provided in Section 4, the functions used were arbitrarily chosen keeping in mind the conditions they must meet.

Furthermore, given that the author claims this is a good contribution for digital computers (since they are finite machines), the author does not discuss computational performance of the adaptive controller; it is not clear whether or not such adaptive controller would perform well online.

Finally, even though the author provides possible applications of such controller, no example was provided to demonstrate the results. But I presume this is the case with theoretical stochastic control papers; most of the focus is on convergence properties of the controller as these are the basis of any good controller; without proving that the controller converges, then the controller might just be "lucky". However, without demonstrating the applications of such controllers on real systems, then they may not be so useful. After all, adaptive controllers are meant to solve problems in the real world where some parameters of the system are not well known.

## 6 Conclusion

The adaptive controller has appealing characteristics when it comes to discrete systems with small finite spaces $\mathcal{X}$, $\mathcal{U}$, and $\mathcal{A}$. The paper focuses on the convergence properties of the controller. The novelty of the adaptive controller is embedded in the estimator of the unknown parameter $\alpha$. Specifically, the novelty is in pre-multiplying the likelihood function with a new parameter. This parameter would "mildly bias" the estimate towards parameters with low cost functions. As such, the controller was compared with ML estimator which is shown to fail in some cases. However, the shortcomings of the paper is summarized in the applicability of the controller. First, the paper does not discuss how to pick the auxiliary functions $f$ and $o$ that are used in the optimal estimator (3). Second, the paper does not discuss how these auxiliary functions affect the performance of the controller. Finally, the paper does not provide an example that would show the controller's performance in a realistic scenario (e.g., computational performance, memory requirements, etc.).

However, with all that criticism, one should keep in mind that the paper was published in 1983. Thus, the mathematical proofs of the convergence properties were used as a foundation to improve on such adaptive controllers; this is demonstrated by the fact that the paper was was cited in [4] on the discussion of non-Bayesian adaptive controllers. Further, it may be the case that in 1983, demonstrating the controller with an example may not be as easy as is the case nowadays using computer software that is widely available.

## References


[1] R. L. Williams II and D. A. Lawrence, *Linear State-Space Control Systems*. John Wiley & Sons, Feb. 2007. Google-Books-ID: UPWAmAXQu1AC.

[2] G. E. P. Box and N. R. Draper, "Empirical Model-Building and Response Surfaces.," 1987.

[3] B. Brogliato, R. Lozano, B. Maschke, and O. Egeland, *Dissipative Systems Analysis and Control: Theory and Applications*. Communications and Control Engineering, London: Springer-Verlag, 2 ed., 2007.

[4] P. R. Kumar and P. Varaiya, *Stochastic systems: estimation, identification and adaptive control*. USA: Prentice-Hall, Inc., 1986.

[5] I. D. Landau, R. Lozano, M. M'Saad, and A. Karimi, "Adaptive Prediction," in *Adaptive Control: Algorithms, Analysis and Applications* (I. D. Landau, R. Lozano, M. M'Saad, and A. Karimi, eds.), Communications and Control Engineering, pp. 193–203, London: Springer, 2011.

[6] P. Kumar and A. Becker, "A new family of optimal adaptive controllers for Markov chains," *IEEE Transactions on Automatic Control*, vol. 27, pp. 137–146, Feb. 1982. Conference Name: IEEE Transactions on Automatic Control.

[7] P. Kumar, "Simultaneous identification and adaptive control of unknown systems over finite parameter sets," *IEEE Transactions on Automatic Control*, vol. 28, pp. 68–76, Jan. 1983. Conference Name: IEEE Transactions on Automatic Control.

[8] P. Kumar and W. Lin, "Optimal adaptive controllers for unknown Markov chains," *IEEE Transactions on Automatic Control*, vol. 27, pp. 765–774, Aug. 1982. Conference Name: IEEE Transactions on Automatic Control.