



Towards understanding the capability of adaptation for time-varying systems[☆]

F. Xue^a, L. Guo^{a,*}, M. Y. Huang^b

^a*Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, People's Republic of China*

^b*Department of Electrical Engineering, McGill University, 3480 University Street, Montreal, QC, Canada H3A 2A7*

Received 26 December 1999; revised 28 August 2000; received in final form 14 March 2001

The information uncertainty coupled with the structural complexity is more pertinent than the rate of parameter changes in characterizing the capability of adaptation for time-varying systems.

Abstract

The primary function of adaptation is to deal with systems with uncertain and time-varying dynamics. However, a natural question is: how fast must the rate of change of the uncertain dynamics be, before it can no longer be captured by adaptation? In order to initiate an understanding of this problem, we consider in this paper a first-order linear system with time-varying unknown parameters modeled as a finite state hidden Markov chain. It is shown that the key factor inherent in characterizing the capability of adaptation is the information uncertainty coupled with the structural complexity of the systems under control, rather than the existence of a certain critical rate of parameter changes that has been conjectured by many in the area. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Adaptation; Uncertainty; Markov chain; Complexity; Stability

1. Introduction

Adaptive control is designed primarily for dealing with systems with uncertain and changing structure and/or environments, which has been a main focus of investigations in automatic control for the past several decades and has been an area of many successful practical applications.

Intuitively, adaptation should at least be able to capture slowly time-varying structure of a system, as has been shown rigorously for a large class of linear finite dimensional adaptive control systems in both the deterministic framework (e.g. Ioannou & Tsakalis, 1985; Middleton & Goodwin, 1988) and the stochastic framework (e.g. Guo, 1990; Meyn & Brown, 1993). There are also considerable research efforts devoted to the identi-

fication of time-varying systems, see, Benveniste (1987), Millnert (1987), Guo and Ljung (1995), and Zames, Lin, and Wang (1994), among many others. To get a comprehensive understanding of the capability of adaptation, one is naturally concerned about the following questions: (i) How fast can the rate of parameter changes be captured by adaptation? (ii) Is there a critical value, that we can find, of the rate of parameter changes in determining the adaptive stabilizability? (iii) What are the limitations of adaptation? (iv) What are the key factors on which the capability of adaptation depends? There is still a lack of theoretical understanding of these questions.

As a starting point towards understanding the above questions, we consider in this paper a first-order linear control system with the unknown time-varying parameter process modeled as a finite state Markov chain, with states denoted by $\{a_i, 1 \leq i \leq N\}$. The Markovian transition probability p_{ij} will then provide a natural measure of the rate of parameter changes, while the degree of dispersion of the states $\{a_i, 1 \leq i \leq N\}$ may be regarded as a measure of the model complexity.

To understand the capability of adaptation, one needs a precise definition of it first. By adaptation we mean adaptive feedback which captures the uncertain

[☆]This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor K. W. Lim under the direction of Editor Frank Lewis. This work was supported by the National Natural Science Foundation of China and the National Key Project of China.

*Corresponding author. Fax: (86-10) 62587343.

E-mail address: lguo@control.iss.ac.cn (L. Guo).

information of the system by properly utilizing the measured on-line system data.¹ However, as mentioned by Aström and Wittenmark (1995), since it is practically difficult, in general, to distinguish adaptive feedback from ordinary nonlinear feedback by looking at either the software or the hardware of a controller, we will speak of adaptive feedback as any causal functions of the observed output process. This will prevent us from restricting the capability of adaptation at the outset by (artificial) definition. Thus, the capability of adaptation that we are going to investigate is also the capability of generally defined feedback.

With the above formulated framework, we are able to explore rigorously and quantitatively what the capability (and limitations) of adaptation is, and how, if at all, it depends on the rate of changes of the unknown system parameters. It turns out that the key factor inherent in characterizing the capability of adaptation is the information uncertainty of the underlying Markov chain, described by the transition probability $\{p_{ij}\}$, coupled with the model complexity exhibited by the dispersion of the states $\{a_i\}$. However, the rate of parameter changes, is found not to be a key factor in characterizing the capability of adaptation, since there is not a monotonic dependence relationship between the two. This fact is made more evident by studying a simplified example of systems where the Markov chain is of only two states. In this case, an explicit connection between the information uncertainty and the Shannon information entropy is established, showing that the capability of adaptation can reach its maximum in both cases where the rate of parameter changes is fast or slow.

Besides, this paper will also study the worst case where the transition probability $\{p_{ij}\}$ is unknown and treated as arbitrary. A necessary and sufficient condition on the structural complexity is found for robust stabilizability in this case, which demonstrates how conservative the worst case framework can be.

The aforementioned results will be presented in the following section, with proofs of these results given in Section 3. Section 4 concludes the paper.

2. The main results

Consider the following first-order time-varying linear stochastic system:

$$y_{t+1} = a(\theta_t)y_t + u_t + w_{t+1}, \quad t \geq 1, y_1 \in R^1, \quad (1)$$

where y_t , u_t , w_t and θ_t are the system output, input, noise and unknown time-varying parameter processes, respec-

tively. Assume that

- (A1) $\{\theta_t\}$ is a homogenous *hidden* Markov chain with finite states taking values in $\{1, 2, \dots, N\}$, and with probability transition matrix $P = (p_{ij})_{N \times N}$, where $p_{ij} \triangleq P(\theta_{t+1} = j | \theta_t = i)$.
- (A2) $\{w_t\}$ is a martingale difference sequence independent of $\{\theta_t\}$, with $Ew_t^2 = \sigma_w^2 > 0, \forall t \geq 1$.
- (A3) $a_i \neq a_j, \forall i \neq j$, where $a_i \triangleq a(i), 1 \leq i \leq N$.

Obviously, it would be a trivial control problem if $\{\theta_t\}$ were observable or available. For example, the simple feedback $u_t = -a(\theta_t)y_t$ will be stable and optimal for regulation, regardless of the degree of dispersion of $\{a_i\}$ as long as $a(\theta_t)$ is available. In our present case, however, the control problem is far from trivial, since the uncertainty of $\{\theta_t\}$ compounded with the complexity (or degree of dispersion in this case) of $\{a_i\}$ will make it impossible to control in many cases no matter how we design the adaptive feedback. To make this fact more precise and to pursue further, we first give a definition of adaptive feedback together with stabilizability.

Definition 2.1. A control sequence $\{u_t, t \geq 1\}$ is called *adaptive feedback*, if at each time $t \geq 1$, u_t belongs to $\sigma(y_1, \dots, y_t)$, the σ -algebra generated by $\{y_1, \dots, y_t\}$, or, in other words, if there exists a Lebesgue measurable function $f_t(\cdot)$ such that $u_t = f_t(y_1, y_2, \dots, y_t)$. Furthermore, system (1) is said to be *stabilizable* by adaptation if there exists a sequence of adaptive feedback $\{u_t\}$ such that the output process is bounded in the mean square sense, i.e., $\lim_{t \rightarrow \infty} Ey_t^2 < \infty, \forall y_1 \in R^1$.

Our main objective is to clarify which cases in (1) can be controlled by adaptive feedback and which cases cannot, or in other words, to understand quantitatively the capability and limitations of adaptation in the presence of information uncertainty and structural complexity. For this purpose, we introduce a matrix function $C(\cdot): R^N \rightarrow R^{N \times N}$, defined as follows for any $V = (v_1, v_2, \dots, v_N)^T \in R^N$:

$$C(V) \triangleq \{(a_j - v_i)^2 p_{ij}\}_{N \times N}. \quad (2)$$

Let the minimum spectrum radius of $C(V)$ over R^N be denoted by ρ^* , i.e.,

$$\rho^* \triangleq \min_{V \in R^N} \rho(C(V)), \quad (3)$$

where $\rho(A)$ denotes the spectrum radius of a matrix A , i.e. $\rho(A) = \max_i |\rho_i(A)|$ with $\rho_i(A)$'s being the eigenvalues of A . The following theorem shows how important the value of ρ^* is in characterizing the capability of adaptation.

¹ The traditional on-line identification-based certainty equivalence principle is only an example of such philosophy.

Theorem 2.1. Consider the control system (1). Assume that assumptions (A1)–(A3) hold and that the probability transition matrix satisfies $P \succ 0$.² Then, the necessary and sufficient condition for system (1) to be stabilizable by adaptation is $\rho^* < 1$.

Remark 2.1. The proof of Theorem 2.1 given in the next section is constructive in the case $\rho^* < 1$, where we can see that a recursive stabilizing adaptation law can be constructed in this case.

By (2) and (3), it appears that the quantity ρ^* depends on $\{a_i\}$ and $\{p_{ij}\}$ in a rather complicated way. To make it more clear in understanding how the capability of adaptation depends on both the complexity and uncertainty of the system, measured by $\{a_i - a_j\}$ and $\{p_{ij}\}$, respectively, we present the following corollary whose proof is also given in the next section.

Corollary 2.1. Under the conditions of Theorem 2.1, we have

- (i) System (1) is stabilizable by adaptation, if $\sum_{1 \leq k < j \leq N} (a_j - a_k)^2 p_{ij} p_{ik} < 1, \forall i: 1 \leq i \leq N$.
- (ii) System (1) is not stabilizable by adaptation, if $\sum_{1 \leq k < j \leq N} (a_j - a_k)^2 p_{ij} p_{ik} \geq 1, \forall i: 1 \leq i \leq N$.

To further understand why the capability of adaptation may not be a monotonic function of the rate of parameter changes in general, we consider the following example:

Example 2.1. Let the Markov chain $\{\theta_i\}$ have two states $\{1, 2\}$ only and let $p_{12} = p_{21}$. Obviously, the rate of parameter changes can be simply described by $p_{12} \in [0, 1]$. Now, by the fact that $N = 2, p_{12} = p_{21}$, and $p_{i1} + p_{i2} = 1, i = 1, 2$, it follows from Corollary 2.1 that the system is stabilizable if and only if $(a_1 - a_2)^2 (1 - p_{12}) p_{12} < 1$. Let us denote $I(p_{12})$ as

$$I(p_{12}) \triangleq 1 - (a_2 - a_1)^2 (1 - p_{12}) p_{12}, \tag{4}$$

which may be regarded as a measure of the capability of adaptation, and may be further represented by

$$I(p_{12}) = 1 - CU, \tag{5}$$

where $C \triangleq (a_2 - a_1)^2$ and $U \triangleq (1 - p_{12}) p_{12}$ can be interpreted as measures of the structural complexity (degree of dispersion) and the information uncertainty of the system, respectively. Obviously, the system is stabilizable $\Leftrightarrow I(p_{12}) > 0$. Now, the following two facts are most intriguing to us:

² To avoid conflicts with the traditional notation for positive definite matrices, throughout this paper, $A \succ 0$ means that all the elements of the matrix A are positive. Likewise, we can define $A \succcurlyeq 0$. Furthermore, $A \succ B$ means $A - B \succ 0$ and $A \succcurlyeq B$ means $A - B \succcurlyeq 0$.

Fact 1. The capability of adaptation $I(p_{12})$ is a monotonic function of the uncertainty U but is not monotonic in the rate of transition p_{12} . Furthermore, there does not exist a critical rate of transition $p_{12}^* \in (0, 1)$ such that the system is stabilizable or $I(p_{12}) > 0 \Leftrightarrow p_{12} \in [0, p_{12}^*)$. Moreover, the capability of adaptation $I(p_{12})$ achieves its maximum when the uncertainty U reaches its minimum. However, U reaches its minimum in both cases where the changing rate p_{12} is “slow” ($p_{12} \approx 0$) and “fast” ($p_{12} \approx 1$). The minimum capability of adaptation is reached at $p_{12} = \frac{1}{2}$, corresponding to the case where the uncertainty U reaches its maximum.

Fact 2. Our uncertainty measure U is closely related to the well-known Shannon information entropy, which is a measure of information uncertainty defined by $H = -\sum_{i=1}^2 p_{1i} \log p_{1i}$ in the current case (see e.g. Cover & Thomas, 1991). Note that H can be rewritten as $H = -(1 - p_{12}) \log(1 - p_{12}) - p_{12} \log p_{12}$ and that the dependence of H on $p_{12} \in (0, 1)$ is completely similar to that of U as defined above. Hence, it is not difficult to see that there exists a monotonically increasing function $m(\cdot)$ such that $U = m(H)$. This fact justifies why we refer to U as the measure of information uncertainty here. Now, by (5) we have $I = 1 - Cm(H)$, which implies that the capability of adaptation is also a monotonically decreasing function of the Shannon information entropy H .

Finally, we present a theorem on stabilizability in the “worst case” framework where the transition probability $\{p_{ij}\}$ is not known and is treated as arbitrary.

Theorem 2.2. The necessary and sufficient condition for the existence of a robust stabilizing control law for system (1) with arbitrary unknown transition matrix P is $\max_{1 \leq i, j \leq N} |a_i - a_j| < 2$.

Remark 2.2. Comparing the conditions imposed on $\{a_i\}$ in Corollary 2.1 with those in Theorem 2.2, one may find how conservative the “worst case” framework can be. Besides, we remark that the noise variance σ_w^2 does not affect the stabilizability in both Theorems 2.1 and 2.2; however, it does affect the upper bound of the output as can be seen from the proofs.

3. Proofs of the main results

3.1. Some preliminaries

For simplicity of presentation, we introduce the following notations:

$$\mathcal{F}_t \triangleq \sigma(\theta_{i-1}, w_i; 1 \leq i \leq t), \quad \mathcal{F}_t^y \triangleq \sigma(y_i, i \leq t), \tag{6}$$

$$\pi_t \triangleq [p(\theta_t = 1), \dots, p(\theta_t = N)]^t, \quad e \triangleq (1, \dots, 1)^t. \tag{7}$$

Moreover, without causing any confusions with (2), for a random variable v taking values in R^1 , we denote

$$C(v) \triangleq \{(a_j - v)^2 p_{ij}\}_{N \times N}. \tag{8}$$

To prove the necessity part of Theorem 2.1, we need the following four lemmas whose proofs are given in Appendix A.

Lemma 3.1. Let $P > 0$, $J_0^* \triangleq e$, and define $J_{k+1}^* \triangleq [J_{k+1,1}^*, \dots, J_{k+1,N}^*]^T$, $k \geq 0$ with $J_{k+1,i}$ recursively defined by ($1 \leq i \leq N$)

$$J_{k+1,i}^* = \min_{x_{k,i} \in R^1} \left[\sum_{j=1}^N (a_j - x_{k,i})^2 p_{ij} J_{k,j}^* \right]. \tag{9}$$

Then, we have (1) $J_k^* > 0$, and for $1 \leq i \leq N$

$$J_{k+1,i}^* = \frac{\sum_{j>l} (a_j - a_l)^2 p_{1j} p_{1l} J_{k,j}^* \cdot J_{k,l}^*}{p_{11} J_{k,1}^* + p_{12} J_{k,2}^* + \dots + p_{1N} J_{k,N}^*}, \tag{10}$$

where the minimum value is attained at $X_k^* \triangleq (x_{k,1}^*, \dots, x_{k,N}^*)^T$ with $x_{k,i}^*$ expressed as $1 \leq i \leq N$,

$$x_{k,i}^* = \frac{a_1 p_{i1} J_{k,1}^* + a_2 p_{i2} J_{k,2}^* + \dots + a_N p_{iN} J_{k,N}^*}{p_{i1} J_{k,1}^* + p_{i2} J_{k,2}^* + \dots + p_{iN} J_{k,N}^*}, \tag{11}$$

i.e. $J_{k+1}^* = C(X_k^*) J_k^*$.

(2) For any $V_1, V_2, \dots, V_t \in R^N$ and $t \geq 1$, we have $C(V_1) \cdot C(V_2) \cdots C(V_t) \cdot e \succ J_t^*$.

Lemma 3.2. Let $P > 0$ and let us define

$$\lambda_{k+1} = (\lambda_{k+1,1}, \dots, \lambda_{k+1,N})^T \triangleq \left(\frac{J_{k+1,1}^*}{J_{k,1}^*}, \dots, \frac{J_{k+1,N}^*}{J_{k,N}^*} \right)^T, \tag{12}$$

where J_k^* is as defined in Lemma 3.1. Then there exists a vector $X_\infty^* \in R^N$ and positive numbers $J_{\infty,i} \in R^1$, $i = 1, \dots, N$, such that (i) $\lambda_{t,i} \rightarrow \rho^* > 0$, $i = 1, \dots, N$; (ii)

$$J_{t,i}^* / \rho^{*t} \rightarrow J_{\infty,i} > 0; \text{ (iii) } X_k^* \xrightarrow{k \rightarrow \infty} X_\infty^*, \text{ and } \rho(C(X_\infty^*)) = \rho^*,$$

where ρ^* is defined in (3), and X_k^* is defined as in Lemma 3.1.

Lemma 3.3. For any $t \geq 1$, there exists a deterministic vector $V_t \in R^N$ satisfying

$$\min_{v_t \in \mathcal{F}_t} E[(a(\theta_t) - v_t)^2 | \mathcal{F}_t] = e^T \cdot C^t(V_t) \cdot [I_{(\theta_{t-1}=1)}, \dots, I_{(\theta_{t-1}=N)}]^T \quad \text{a.s.},$$

where $C(V_t)$ is defined in (2).

Lemma 3.4. If $\beta_k > 0$ is an N -dimensional deterministic vector, v_k is a random variable, then

$$\begin{aligned} \min_{v_k \in \mathcal{F}_k} \beta_k^T \cdot EC^t(v_k) \cdot [I_{(\theta_{k-1}=1)} y_k^2, \dots, I_{(\theta_{k-1}=N)} y_k^2]^T \\ = \min_{v_k \in \sigma(\theta_{k-1})} \beta_k^T \cdot EC^t(v_k) \cdot [I_{(\theta_{k-1}=1)} \cdots I_{(\theta_{k-1}=N)}]^T y_k^2. \end{aligned}$$

To prove the sufficiency part we need to construct a feedback law first. Without loss of generality, we assume that the states of the Markov chain satisfy $a_1 < a_2 < \dots < a_N$, and we denote $d \triangleq \min_{i \neq j} |a_i - a_j|$. Introduce a function $\psi(x)$:

$$\psi(x) = \begin{cases} 1, & -\infty < x \leq a_1 + \frac{1}{2}d, \\ 2, & a_1 + \frac{1}{2}d < x \leq a_2 + \frac{1}{2}d, \\ \vdots & \\ N-1, & a_{N-2} + \frac{1}{2}d < x \leq a_{N-1} + \frac{1}{2}d, \\ N, & a_{N-1} + \frac{1}{2}d < x < +\infty. \end{cases} \tag{13}$$

Let $J_0^* \triangleq e$. For $t \geq 1$, recursively define $J_t^* \triangleq (J_{t,1}^*, \dots, J_{t,N}^*)^T$ by (10), and denote V_t as $V_t = X_{t-1}^*$ where X_t^* is as in Lemma 3.1. Now, define the control law by

$$\begin{aligned} u_t = -V_t^T \cdot \begin{pmatrix} I_{(\psi((y_t - u_{t-1})/y_{t-1})=1)} \\ \vdots \\ I_{(\psi((y_t - u_{t-1})/y_{t-1})=N)} \end{pmatrix} \cdot I_{(y_{t-1} \neq 0)} \cdot y_t \\ \triangleq -v_t \cdot I_{(y_{t-1} \neq 0)} y_t, \end{aligned} \tag{14}$$

where $I_{(\cdot)}$ is the indicator function, and v_t is defined in an obvious way.

To show that the above defined adaptation law is indeed stabilizing, we need the following lemmas whose proofs are given in Appendix B.

Lemma 3.5. Let $P > 0$ and $\rho^* < 1$. Then there exist $t_0 > 0$ and $N \times N$ matrix C_* satisfying $\rho(C_*) \triangleq \rho_* < 1$ such that $C(V_t)^T \leq C_*^t, \forall t \geq t_0$, where V_t is defined as above.

Lemma 3.6. For any $\lambda > 0$ and $d > 0$, there exists a constant $M_\lambda > 0$ such that system (1) under control (14) satisfies ($\forall t \geq 2$)

$$E y_{t+1}^2 I_{(|w_t/y_{t-1}| \geq d/2)} \cdot I_{(y_{t-1} \neq 0)} \leq M_\lambda + \lambda \cdot E y_{t-1}^2. \tag{15}$$

Lemma 3.7. Let $P > 0$ and $\rho^* < 1$, and let us denote

$$\begin{aligned} \gamma_{t+1} \triangleq E [I_{(\theta_t=1)} \cdots I_{(\theta_t=N)}]^T y_{t+1}^2 \\ \times I_{(|w_t/y_{t-1}| < d/2, y_{t-1} \neq 0)}. \end{aligned} \tag{16}$$

Then for any $\lambda > 0$ and $t \geq t_0$, there exists a constant M^* such that system (1) under control (14) satisfies

$$\gamma_{t+1} \leq C_*^t \cdot \gamma_t + (M^* + \lambda E y_{t-2}^2) \cdot C_*^t e + \sigma_w^2 e, \tag{17}$$

where t_0 and C_* are defined as in Lemma 3.5.

Corollary 3.1. Under the assumptions of Lemma 3.7, there exist constants $c_0 > N$, $\rho \in (0,1)$ and $M_1 = M_1(t_0)$ such that $\forall t \geq t_0$

$$\begin{aligned} e^t \gamma_{t+1} \leq M_1 + \lambda E y_{t_0-2}^2 \cdot c_0 \rho^{t+1-t_0} + \dots \\ + \lambda E y_{t-2}^2 \cdot c_0 \rho. \end{aligned} \tag{18}$$

Lemma 3.8. Let $\{b_k\}$ be a sequence of positive numbers satisfying the following inequality:

$$b_{k+1} \leq M + \eta \cdot \max_{k_0 \leq i \leq k} b_i, \quad \forall k \geq k_0, \quad (19)$$

where $M > 0$ and $\eta \in (0,1)$ are constants and $k_0 > 0$ is an integer. Then

$$b_k \leq \max \left\{ b_{k_0}, b_{k_0+1}, \frac{M}{1-\eta} \right\}, \quad \forall k \geq k_0.$$

3.2. The proof of Theorem 2.1

(i) *Necessity:* We will show that if $\rho^* \geq 1$, then no matter how the feedback control $u_t \in \sigma\{y_1, \dots, y_t\}$ is constructed, the output sequence is always divergent in the sense that $\lim_{t \rightarrow \infty} E y_t^2 = \infty, \forall y_1 \in R^1$. Now, for any feedback sequence $\{u_t\}$ with $u_t \in \sigma\{y_1, \dots, y_t\}$, let us denote $\xi_k \triangleq -(u_k/y_k)I_{(y_k \neq 0)}$. Then

$$\begin{aligned} E y_{t+1}^2 &= E(a(\theta_t)y_t + u_t)^2 + \sigma_w^2 \\ &\geq \sigma_w^2 + E(a(\theta_t) - \xi_t)^2 y_t^2 \\ &= \sigma_w^2 + E\{E((a(\theta_t) - \xi_t)^2 | \mathcal{F}_t) \cdot y_t^2\} \\ &\geq \sigma_w^2 + E\left\{ \min_{\xi_t \in \mathcal{F}_t} E((a(\theta_t) - \xi_t)^2 | \mathcal{F}_t) \cdot y_t^2 \right\} \\ &= \sigma_w^2 + e^t \cdot C^t(V_t) \cdot E[I_{(\theta_{t-1}=1)}, \dots, I_{(\theta_{t-1}=N)}]^t y_1^2, \end{aligned} \quad (20)$$

where the last relationship follows from Lemma 3.3 with $V_t \in R^N$ being a deterministic vector. Next, for any $j \in \{1, \dots, N\}$, we have

$$\begin{aligned} E y_t^2 I_{(\theta_{t-1}=j)} &= E[a(\theta_{t-1})y_{t-1} + u_{t-1}]^2 \cdot I_{(\theta_{t-1}=j)} \\ &\quad + \sigma_w^2 E I_{(\theta_{t-1}=j)} \\ &\geq E[a(\theta_{t-1}) - \xi_{t-1}]^2 y_{t-1}^2 \cdot I_{(\theta_{t-1}=j)} \\ &\quad + \sigma_w^2 \cdot E I_{(\theta_{t-1}=j)} \\ &= E\{E[(a(\theta_{t-1}) - \xi_{t-1})^2 \cdot I_{(\theta_{t-1}=j)} | \mathcal{F}_{t-1}] \cdot y_{t-1}^2\} \\ &\quad + \sigma_w^2 \cdot E I_{(\theta_{t-1}=j)}. \end{aligned}$$

Note that ξ_{t-1} is \mathcal{F}_{t-1} measurable, and $\{\theta_t\}$ is a Markov chain, we then have

$$\begin{aligned} E y_t^2 I_{(\theta_{t-1}=j)} &\geq E \left\{ \sum_{i=1}^N (a_j - \xi_{t-1})^2 p_{ij} I_{(\theta_{t-2}=i)} \cdot y_{t-1}^2 \right\} \\ &\quad + \sigma_w^2 \cdot E I_{(\theta_{t-1}=j)}. \end{aligned}$$

Consequently, by denoting

$$x_t \triangleq [I_{(\theta_{t-1}=1)} y_t^2 \cdots I_{(\theta_{t-1}=N)} y_t^2]^T$$

and recalling the notations in (6)–(8), we then have for any $\beta \in R^N$ with $\beta > 0$,

$$\begin{aligned} \beta^t E x_t &\geq \beta^t \cdot E \left[\begin{array}{c} \sum_{i=1}^N (a_1 - \xi_{t-1})^2 p_{i1} I_{(\theta_{t-2}=i)} \cdot y_{t-1}^2 \\ \vdots \\ \sum_{i=1}^N (a_N - \xi_{t-1})^2 p_{iN} I_{(\theta_{t-2}=i)} \cdot y_{t-1}^2 \end{array} \right] \\ &\quad + \sigma_w^2 \beta^t \cdot \left[\begin{array}{c} E I_{(\theta_{t-1}=1)} \\ \vdots \\ E I_{(\theta_{t-1}=N)} \end{array} \right] \\ &= \beta^t \cdot E C^t(\xi_{t-1}) x_{t-1} + \sigma_w^2 \beta^t \pi_{t-1} \\ &\geq \min_{\mu_{t-1} \in \mathcal{F}_{t-1}} \beta^t E C^t(\mu_{t-1}) x_{t-1} + \sigma_w^2 \beta^t \pi_{t-1} \\ &= \min_{\mu_{t-1} \in \sigma(\theta_{t-2})} \beta^t \cdot E C^t(\mu_{t-1}) x_{t-1} + \sigma_w^2 \beta^t \pi_{t-1}, \end{aligned} \quad (21)$$

where for the last equality we have used Lemma 3.4. Since $\mu_{t-1} \in \sigma(\theta_{t-2})$, we may assume that the minimum in (21) is reached at $\mu_{t-1} = \sum_{i=1}^N v_{t-1,i} I_{(\theta_{t-2}=i)}$. Let us denote $V_{t-1} = (v_{t-1,1}, \dots, v_{t-1,N})^T$. Then by (21), we know that for any $t \geq 2$ and $\beta > 0$, there is $V_{t-1} \in R^N$ which may depend on β such that

$$\beta^t E x_t \geq \beta^t C^t(V_{t-1}) E x_{t-1} + \sigma_w^2 \beta^t \pi_{t-1}. \quad (22)$$

Let us set $\beta_1 = C(V_t)e$ and take $\beta_2 \triangleq C(V_{t-1})\beta_1$. Since $e > 0$ and $P > 0$, it is easy to verify that $\beta_1 > 0$. Similarly, we have $\beta_2 > 0$. Hence, by (22) with $\beta = \beta_1$ we have

$$\beta_1^t E x_t \geq \beta_2^t E x_{t-1} + \sigma_w^2 \beta_1^t \pi_{t-1}. \quad (23)$$

Now, the term $\beta_2^t E x_{t-1}$ can be analyzed by applying (22) again with $\beta = \beta_2$ and with t replaced by $t - 1$. Hence, there exists $\beta_3^t \triangleq \beta_2^t C^t(V_{t-2})$ such that $\beta_2^t E x_{t-1} \geq \beta_3^t E x_{t-2} + \sigma_w^2 \beta_2^t \pi_{t-2}$. Substituting this into (23) gives $\beta_1^t E x_t \geq \beta_3^t E x_{t-2} + \sigma_w^2 \beta_2^t \pi_{t-2} + \sigma_w^2 \beta_1^t \pi_{t-1}$. Continuing this procedure, we finally get

$$\beta_1^t E x_t \geq \beta_i^t E x_1 + \sigma_w^2 [\beta_{t-1}^t \pi_1 + \dots + \beta_1^t \pi_{t-1}], \quad (24)$$

where $\beta_i^t = e^t C^t(V_t) C^t(V_{t-1}) \cdots C^t(V_{t-i+1})$. By Lemma 3.1, we know that $\beta_i \geq J_i^*$. Hence, by (20) and (24), $E y_{t+1}^2 \geq \sigma_w^2 + \beta_1^t E x_t \geq \sigma_w^2 + \beta_i^t E x_1 + \sigma_w^2 [J_{t-1}^{*t} \pi_1 + J_{t-2}^{*t} \pi_2 + \dots + J_1^{*t} \pi_{t-1}]$. By Lemma 3.2, there exists $c_0 > 0$ such that $J_i^{*t} \geq c_0 \rho^{*t} e, \forall t \geq 0$. Also, by the definitions of e and π_i , it follows that $\pi_i^t e = 1$. So, by $\rho^* \geq 1$ we have

$$\begin{aligned} E y_{t+1}^2 &\geq \sigma_w^2 + \beta_i^t E x_1 \\ &\quad + \sigma_w^2 [c_0 \rho^{*t-1} \cdot e^t \pi_1 + \dots + c_0 \rho^{*t} \cdot e^t \pi_{t-1}] \\ &= \sigma_w^2 + \beta_i^t E x_1 + c_0 \sigma_w^2 \cdot (\rho^{*t-1} + \dots + \rho^{*t}) \\ &\rightarrow \infty. \\ &\quad t \rightarrow \infty \end{aligned}$$

(ii) *Sufficiency:* We will show that if $\rho^* < 1$, then the adaptation law defined by (14) is stabilizing for any initial

condition $y_1 \in R^1$. For any $t \geq t_0$, by (16) we have

$$\begin{aligned} E y_{t+1}^2 &= E y_{t+1}^2 [I_{(|w_t/y_{t-1}| < d/2)} + I_{(|w_t/y_{t-1}| \geq d/2)}] I_{(y_{t-1} \neq 0)} \\ &\quad + E y_{t+1}^2 I_{(y_{t-1} = 0)} \\ &= e^{\tau} \gamma_{t+1} + E y_{t+1}^2 I_{(|w_t/y_{t-1}| \geq d/2)} \cdot I_{(y_{t-1} \neq 0)} \\ &\quad + E y_{t+1}^2 I_{(y_{t-1} = 0)}. \end{aligned} \tag{25}$$

By (14), we have $u_{t-1} = 0$ and $u_t = 0$ on the set $\{y_{t-1} = 0\}$. So, by (1), $E y_{t+1}^2 I_{(y_{t-1} = 0)} = E[a(\theta_t)w_t + w_{t+1}]^2 I_{(y_{t-1} = 0)} \leq 2 \max_{1 \leq k \leq N} |a_k| \cdot \sigma_w^2 + 2\sigma_w^2$. Hence, there is a constant M_0 such that $E y_{t+1}^2 I_{(y_{t-1} = 0)} \leq M_0$. Substituting this into (25) and applying Lemma 3.6, we have

$$E y_{t+1}^2 \leq e^{\tau} \gamma_{t+1} + M_\lambda + \lambda E y_{t-1}^2 + M_0, \tag{26}$$

where $\lambda > 0$ is an arbitrary constant. Now, let us take $\lambda > 0$ to satisfy $\lambda c_0 / (1 - \rho) < 1$, where c_0 and ρ are defined as in Corollary 3.1. By Corollary 3.1, it follows from (26) that

$$\begin{aligned} E y_{t+1}^2 &\leq M_1 + \lambda E y_{t_0-2}^2 \cdot c_0 \rho^{t+1-t_0} + \dots \\ &\quad + \lambda E y_{t-2}^2 \cdot c_0 \rho + M_\lambda + \lambda E y_{t-1}^2 + M_0 \\ &\leq M_2 + \max_{t_0 \leq k \leq t-1} E y_k^2 \cdot [c_0 \rho^{t-1-t_0} + \dots \\ &\quad + c_0 \rho + 1] \lambda \\ &\leq M_2 + \frac{\lambda c_0}{1 - \rho} \max_{t_0 \leq k \leq t} E y_k^2, \end{aligned}$$

where we have used the fact that $c_0 > 1$, and where $M_2 \triangleq M_1 + \lambda E y_{t_0-2}^2 \cdot c_0 + \lambda E y_{t_0-1}^2 \cdot c_0 + M_\lambda + M_0$. Hence, by Lemma 3.8, we have $E y_t^2 \leq \max\{E y_{t_0}^2, E y_{t_0+1}^2, M_2 / (1 - \lambda c_0 / (1 - \rho))\}$. This completes the proof of Theorem 2.1. \square

3.3. The proof of Corollary 2.1

By the definition of λ_{1i} in (12), the expression of $J_{1,i}^*$ in (10) and the fact that $J_{0i}^* = 1$, we know that $\lambda_{1i} = \sum_{1 \leq k < j \leq N} (a_j - a_k)^2 p_{ij} p_{ik}$, $i = 1, \dots, N$. Moreover, by Lemmas 3.2(i) and A.1, it is clear that $\min_i \{\lambda_{1i}\} \leq \rho^* \leq \max_i \{\lambda_{1i}\}$. Hence, if $\sum_{1 \leq k < j \leq N} (a_j - a_k)^2 p_{ij} p_{ik} \geq 1$ for any $i: 1 \leq i \leq N$, then we have $\rho^* \geq 1$. Therefore, by Theorem 2.1, we know that the system is not stabilizable. Similarly, if $\sum_{1 \leq k < j \leq N} (a_j - a_k)^2 p_{ij} p_{ik} < 1$ for any $i: 1 \leq i \leq N$, then we have $\rho^* < 1$, and hence by Theorem 2.1, we know that the system is stabilizable. \square

3.4. The proof of Theorem 2.2

Without loss of generality, we assume that $a_1 < a_2 < \dots < a_N$.

(i) *Sufficiency*: Let $u_t = -[(a_N - a_1)/2]y_t$. Then substituting it into (1), we have $y_{t+1} = [a(\theta_t) -$

$(a_N - a_1)/2]y_t + w_{t+1}$. It is quite obvious that $|a(\theta_t) - (a_N - a_1)/2| \leq (a_N - a_1)/2$, so $E y_{t+1}^2 = E[a(\theta_t) - (a_N - a_1)/2]^2 y_t^2 + \sigma_w^2 < ((a_N - a_1)/2)^2 E y_t^2 + \sigma_w^2$. Since by assumption $0 < (a_N - a_1)/2 < 1$, we have for any $t \geq 1$, $E y_{t+1}^2 \leq E y_1^2 + \sigma_w^2 / (1 - ((a_N - a_1)/2)^2)$, and hence the desired result is true.

(ii) *Necessity*: Let us assume that the system is stabilizable for any $\{p_{ij}\}$ but $a_N - a_1 \geq 2$. Then we take a special transition matrix as $p_{i1} = p_{iN} = \frac{1}{2}$ and $p_{ij} = 0$, otherwise, which means that the Markov chain reduces to a two-state one within the first two steps, with a simplified transition matrix $\{p_{ij}\}_{2 \times 2} = \{\frac{1}{2}\}_{2 \times 2}$. Now, by the definition of $C(V)$, we know that $C(V) = \frac{1}{2}\{(a_j - v_i)^2\}_{2 \times 2}$. Hence, we get

$$C(V) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \succcurlyeq \frac{1}{2} \begin{pmatrix} (a_1 - a_2)^2 \\ 2 \\ (a_1 - a_2)^2 \\ 2 \end{pmatrix} \succcurlyeq \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \forall V \in R^2.$$

So, by the theory of nonnegative matrix (Horn & Johnson, 1985, p. 493), we have $\min_{V \in R^2} \rho(C(V)) \geq 1$. Hence by Theorem 2.1, we know that the system is not stabilizable. This contradicts our assumption, and thus the proof is completed. \square

4. Concluding remarks

A fundamental issue in adaptive theory, is to understand the capability, and limitations, of adaptation for time-varying systems. As a starting point towards this investigation, we have in this paper studied a first-order linear control system with time-varying parameters modeled by a hidden Markov chain. The basic insights that we have obtained into the capacity of adaptation and its inextricable link to structure complexity, information uncertainty and the rate of parameter changes should have meaningful implications for more general time-varying systems. In particular, the demonstrated fact that ‘‘uncertainty’’ is more pertinent than the ‘‘rate of parameter changes’’ in characterizing the capability of adaptation has furthered the existing understanding of adaptation. For further investigation, it would be of interest to characterize the capability of adaptation for more general high dimensional time-varying systems, by finding an explicit necessary and sufficient condition connected to structural complexity, and information uncertainty, as was done in Theorem 2.1 and Example 2.1.

Acknowledgements

The authors would like to thank the anonymous referees for their helpful comments which improved the presentation of this paper.

Appendix A. Proofs of Lemmas 3.1–3.4

Proof of Lemma 3.1. (1) We first prove that $J_k^* > 0$. By the assumption that $P > 0$ and $a_i \neq a_j, i \neq j$, it follows that if $J_k^* > 0$, then $(a_1 - x_{ki})^2 p_{i1} J_{k1}^* + \dots + (a_N - x_{ki})^2 p_{iN} J_{kN}^* > 0, \forall x_{ki} \in R^1, i = 1, \dots, N$. Since the optimization problem in (9) is solved essentially within a closed set of R^1 , it is clear that the optimal value is greater than 0. Hence, by induction we have $J_k^* > 0$, as long as $J_0 > 0$. Note that (9) is simply an optimization problem on nonnegative quadratic polynomials, so the optimal value is simply (10) and the optimal solution is (11).

(2) We use induction to prove the second conclusion. First, for $t = 1$, it is clear that

$$C(V_1)e \succcurlyeq \begin{pmatrix} \min_{x_{1,1} \in R^1} \sum_{i=1}^N (a_i - x_{1,1})^2 p_{1i} \\ \vdots \\ \min_{x_{1,N} \in R^1} \sum_{i=1}^N (a_i - x_{1,N})^2 p_{Ni} \end{pmatrix} = J_1^*.$$

Hence, the conclusion is valid at $t = 1$. Suppose that the conclusion is true for $t = k$, i.e., for J_k^* defined by (10), $C(V_1) \cdot C(V_2) \cdots C(V_k) \cdot e \succcurlyeq J_k^*, \forall V_1, \dots, V_k \in R^N$. By this induction assumption and the fact that all the elements of $C(V)$ are nonnegative, we have for $t = k + 1$, $C(V_1) \cdot C(V_2) \cdots C(V_{k+1}) \cdot e \succcurlyeq C(V_1) \cdot J_k^* \succcurlyeq J_{k+1}^*$. Thus, by induction we have completed the proof. \square

To prove Lemma 3.2, we need the following two auxiliary lemmas.

Lemma A.1. Under the assumption and notations of Lemma 3.2, we have for any integer $k \geq 1$,

$$\begin{aligned} \min_{1 \leq i \leq N} \lambda_{k,i} &\leq \min_{1 \leq i \leq N} \lambda_{k+1,i} \\ &\leq \max_{1 \leq i \leq N} \lambda_{k+1,i} \leq \max_{1 \leq i \leq N} \lambda_{k,i}. \end{aligned} \tag{A.1}$$

Proof. By Lemma 3.1, we have for $1 \leq i \leq N$,

$$\begin{aligned} J_{k+1,i}^* &= \min_{u \in R^1} \left[\sum_{j=1}^N (a_j - u)^2 p_{ij} J_{kj}^* \right] \\ &= \min_{u \in R^1} \left[\sum_{j=1}^N (a_j - u)^2 p_{ij} J_{k-1,j}^* \lambda_{k,j} \right]. \end{aligned}$$

So, we have

$$\min_{1 \leq j \leq N} \lambda_{k,j} \cdot J_{k,i}^* \leq J_{k+1,i}^* \leq \max_{1 \leq j \leq N} \lambda_{k,j} \cdot J_{k,i}^*.$$

Hence by the definition of $\lambda_{k+1,i}$, we have $\lambda_{k+1,i} \in [\min_j \{\lambda_{k,j}\}, \max_j \{\lambda_{k,j}\}]$, which is tantamount to the desired result. \square

Lemma A.2. Let $P > 0$. Then there exist constants $\delta > 0, c_1 > 0$ and $\varepsilon_0 \in (0,1)$ such that

(1) $\min_i J_{k,i}^* / \max_i J_{k,i}^* \geq \delta, \forall k$; (2) $J_{k+1,i}^* / J_{k,i}^* \leq c_1, \forall k, i$; and (3) $(\lambda_{k+1}^+ - \lambda_{k+1}^-) \leq (1 - \varepsilon_0)(\lambda_k^+ - \lambda_k^-)$, where $\lambda_k^+ \triangleq \max_{1 \leq j \leq N} \lambda_{k,j}$, and $\lambda_k^- \triangleq \min_{1 \leq j \leq N} \lambda_{k,j}$.

Proof. (1) We need only to verify the result for $k \geq 1$. By Lemma 3.1, we have $J_k^* > 0$. Then by $p_{ij} \leq 1$, it follows that $J_{k,i}^* = \min_{u \in R^1} [\sum_{j=1}^N (a_j - u)^2 p_{ij} J_{k-1,j}^*] \leq \min_{u \in R^1} [\sum_{j=1}^N (a_j - u)^2 J_{k-1,j}^*] \triangleq \eta_k, k \geq 1$. On the other hand, if we denote $\delta \triangleq \min_{i,j} p_{ij} > 0$, then it can be easily verified that $J_{k,i}^* \geq \delta \eta_k$, hence, it follows that $\min J_{k,i}^* / \max J_{k,i}^* \geq \delta \eta_k / \eta_k = \delta > 0, k \geq 1$.

(2) By the following formula:

$$J_{k+1,i}^* = \min_{u \in R^1} \sum_{j=1}^N [(a_j - u)^2 p_{ij} J_{kj}^*],$$

we have by setting $u = 0$ and noting (1), $J_{k+1,i}^* \leq (\sum_{l=1}^N a_l^2) \cdot \max_{1 \leq h \leq N} J_{k,h}^* \leq (\sum_{l=1}^N a_l^2) (1/\delta) \cdot J_{k,i}^*$. So we just need to take $c_1 = (\sum_{l=1}^N a_l^2) 1/\delta$ for the desired result to hold.

(3) As before, let us define δ and d by $\delta \triangleq \min_{i,j} p_{ij} > 0$ and $d \triangleq \min_{i \neq j} |a_i - a_j| > 0$. We consider two cases separately.

Case 1: $N = 2$ (without loss of generality, we assume that $a_2 > a_1$). In this case,

$$J_{k+1,1}^* = \min_{u_k} [(a_1 - u_k)^2 p_{11} J_{k1}^* + (a_2 - u_k)^2 p_{12} J_{k2}^*],$$

$$J_{k+1,2}^* = \min_{v_k} [(a_1 - v_k)^2 p_{21} J_{k1}^* + (a_2 - v_k)^2 p_{22} J_{k2}^*].$$

It is easy to see that the corresponding minima are

$$u_k^* = \frac{a_1 p_{11} J_{k1}^* + a_2 p_{12} J_{k2}^*}{p_{11} J_{k1}^* + p_{12} J_{k2}^*},$$

$$v_k^* = \frac{a_1 p_{21} J_{k1}^* + a_2 p_{22} J_{k2}^*}{p_{21} J_{k1}^* + p_{22} J_{k2}^*},$$

By this and (1), we have

$$u_k^* - a_1 = \frac{(a_2 - a_1) p_{12} J_{k2}^*}{p_{11} J_{k1}^* + p_{12} J_{k2}^*} = \frac{(a_2 - a_1) p_{12}}{p_{11} \delta + p_{12}},$$

which implies that there exists a constant ε such that $|u_k^* - a_1| \geq \varepsilon > 0$. Similarly, it can be shown that there exists $\varepsilon' > 0$ such that $|a_2 - u_k^*| \geq \varepsilon'$ and $|a_i - v_k^*| \geq \varepsilon', i = 1, 2$.

Next, without loss of generality, we assume $\lambda_k^+ = \lambda_{k1}$ and $\lambda_k^- = \lambda_{k2}$. Then

$$\begin{aligned} J_{k+1,1}^* &= (a_1 - u_{k-1}^*)^2 \lambda_k^+ \cdot J_{k-1,1}^* p_{11} \\ &\quad + (a_2 - u_{k-1}^*)^2 \lambda_k^- \cdot J_{k-1,2}^* p_{12} \\ &\leq \lambda_k^+ \sum_{i=1}^2 (a_i - u_{k-1}^*)^2 \cdot J_{k-1,i}^* p_{1i} \\ &\quad - (\lambda_k^+ - \lambda_k^-) (a_2 - u_{k-1}^*)^2 \cdot J_{k-1,2}^* p_{12} \end{aligned}$$

$$\begin{aligned} &\leq \lambda_k^+ J_{k1}^* - (\lambda_k^+ - \lambda_k^-) \varepsilon'^2 \delta \cdot J_{k-1,2}^* \\ &\leq \left[\lambda_k^+ - \varepsilon'^2 \cdot \frac{1}{c_1} (\lambda_k^+ - \lambda_k^-) \right] J_{k,1}^*. \end{aligned}$$

Likewise, we have $J_{k+1,2}^* \leq [\lambda_k^+ - \varepsilon'^2 \cdot (1/c_1) (\lambda_k^+ - \lambda_k^-)] J_{k,2}^*$. So, $\lambda_{k+1}^+ \leq \lambda_k^+ - (\varepsilon'^2/c_1) \cdot (\lambda_k^+ - \lambda_k^-)$. By Lemma A.1, $\lambda_{k+1}^- \geq \lambda_k^-$, and consequently, $\lambda_{k+1}^+ - \lambda_{k+1}^- \leq (1 - \varepsilon'^2/c_1) \cdot (\lambda_k^+ - \lambda_k^-)$.

Case 2: $N \geq 3$. For $k \geq 1$, there are at least two elements λ_{k,j_1} and λ_{k,j_2} in the set $\{\lambda_{k1}, \lambda_{k2}, \dots, \lambda_{kN}\}$, both of which are either less than or equal to $(\lambda_k^+ + \lambda_k^-)/2$, or greater than or equal to $(\lambda_k^+ + \lambda_k^-)/2$. Without loss of generality, we assume that both λ_{k,j_1} and λ_{k,j_2} are less than or equal to $(\lambda_k^+ + \lambda_k^-)/2$. We may also assume that $j_1 = 1, j_2 = 2$. For any $i \in \{1, \dots, N\}$, let u_{k-1}^* be the solution of the minimization problem: $\min_{u_{k-1}} \sum_{j=1}^N (a_j - u_{k-1})^2 p_{ij} J_{k,j}^*$. Then

$$\begin{aligned} J_{k+1,i}^* &= \min_{u_k} \sum_{j=1}^N (a_j - u_k)^2 p_{ij} J_{k,j}^* \\ &\leq \sum_{j=1}^N (a_j - u_{k-1}^*)^2 p_{ij} \cdot \lambda_{k,j} J_{k-1,j}^* \\ &\leq \lambda_k^+ \left[\sum_{j=1}^N (a_j - u_{k-1}^*)^2 p_{ij} J_{k-1,j}^* \right] \\ &\quad - \frac{\lambda_k^+ - \lambda_k^-}{2} \cdot [(a_1 - u_{k-1}^*)^2 p_{i1} J_{k-1,1}^* \\ &\quad + (a_2 - u_{k-1}^*)^2 p_{i2} J_{k-1,2}^*] \\ &\leq \lambda_k^+ J_{k,i}^* - \frac{\lambda_k^+ - \lambda_k^-}{2} \cdot \delta \\ &\quad \times \min_{1 \leq j \leq N} J_{k-1,j}^* \cdot [(a_1 - u_{k-1}^*)^2 + (a_2 - u_{k-1}^*)^2] \\ &\leq \lambda_k^+ J_{k,i}^* - \frac{\lambda_k^+ - \lambda_k^-}{2} \cdot \delta \cdot \frac{d^2}{4} \cdot \min_{1 \leq j \leq N} J_{k-1,j}^*. \end{aligned} \tag{A.2}$$

Now, since by (2) and (1),

$$\min_{1 \leq j \leq N} J_{k-1,j}^* \geq \min_{1 \leq j \leq N} \frac{1}{c_1} \cdot J_{k,j}^* \geq \frac{1}{c_1} \cdot \delta \cdot J_{k,i}^*,$$

we have by (A.2),

$$\begin{aligned} J_{k+1,i}^* &\leq \lambda_k^+ J_{k,i}^* - \frac{1}{c_1} \delta^2 \cdot \frac{d^2}{4} \cdot \frac{\lambda_k^+ - \lambda_k^-}{2} \cdot J_{k,i}^* \\ &= \left[\lambda_k^+ - \frac{d^2 \delta^2}{8c_1} \cdot (\lambda_k^+ - \lambda_k^-) \right] \cdot J_{k,i}^*. \end{aligned}$$

By Lemma A.1, $\lambda_{k+1}^- \geq \lambda_k^-$, so we have $\lambda_{k+1}^+ - \lambda_{k+1}^- \leq (1 - d^2 \delta^2 / 8c_1) \cdot (\lambda_k^+ - \lambda_k^-)$. \square

Proof of Lemma 3.2. By Lemma A.2, we know that $\lambda_{t,i}$, $1 \leq i \leq N$, tend to the same limit $\rho_0 > 0$ as $t \rightarrow \infty$. Let

us write $\lambda_{t,i} = \rho_0 + \Delta_{t,i}$, then it is easy to see that

$$\begin{aligned} |\Delta_{t,i}| &\leq (\lambda_t^+ - \lambda_t^-) \leq (1 - \varepsilon_0)^{t-1} \cdot (\lambda_1^+ - \lambda_1^-) \\ &\triangleq (1 - \varepsilon_0)^{t-1} \cdot \Delta. \end{aligned} \tag{A.3}$$

Furthermore, by $J_{t,i}^* = \lambda_{t,i} \cdot \lambda_{t-1,i} \cdots \lambda_{1,i} \cdot J_{0i}^* = \prod_{k=1}^t (\rho_0 + \Delta_{t,i}) \cdot J_{0i}^*$, we have $J_{t,i}^* / \rho_0^t = \prod_{k=1}^t (1 + (1/\rho_0) \Delta_{t,i}) \cdot J_{0i}^*$, where the product converges to a positive number by (A.3). So there exists $J_{\infty,i} > 0$ satisfying:

$$\frac{J_{t,i}^*}{\rho_0^t} \rightarrow J_{\infty,i} > 0, \quad t \rightarrow \infty. \tag{A.4}$$

Next, we proceed to prove that $\rho^* = \rho_0$. By Lemma 3.1, $C(X_t^*) J_t^* = J_{t+1}^*$, hence we have $C(X_t^*) \cdot J_t^* / \rho_0^t = J_{t+1}^* / \rho_0^{t+1} \cdot \rho_0$. By (11) it is obvious that the value of X_t^* is related to the ratios between the components of J_t^* , so if J_t^* / ρ_0^t has a limit, then X_t^* has a limit X_∞^* too. Now, let $t \rightarrow \infty$, we get $C(X_\infty^*) J_\infty = \rho_0 J_\infty$. Since $C(X_\infty^*) \succcurlyeq 0$, by the theory of nonnegative matrix (Horn & Johnson, 1985, p. 493), we have $\rho(C(X_\infty^*)) = \rho_0$. So by the definition of ρ^* , we have $\rho_0 \geq \rho^*$. Hence, if $\rho_0 = \rho^*$ were not true, we would have $\rho_0 > \rho^*$. By (3) we may assume that C^* is the matrix satisfying $\rho(C^*) = \rho^*$. Therefore, by Lemma 3.1(ii) we see that $(C^*)^t e \succcurlyeq J_t^*$ or

$$\frac{(C^*)^t e}{\rho_0^t} \succcurlyeq \frac{J_t^*}{\rho_0^t}.$$

The right-hand side tends to a vector with nonzero components as $t \rightarrow \infty$ by (A.4), while the left-hand side tends to zero by $\rho_0 > \rho^*$. This is a contradiction, which shows that $\rho_0 = \rho^*$ is true. \square

Proof of Lemma 3.3. For $v_t \in \mathcal{F}_t$, we have

$$\begin{aligned} E[(a(\theta_t) - v_t)^2 | \mathcal{F}_t] &= E[(a(\theta_t) - x)^2 | \mathcal{F}_t]_{x=v_t} \\ &= E[(a(\theta_t) - x)^2 | \theta_{t-1}]_{x=v_t} \\ &= \sum_{i,j=1}^N (a_j - v_t)^2 p_{ij} I_{(\theta_{t-1}=i)}. \end{aligned}$$

From this we see that the coefficient of $I_{(\theta_{t-1}=i)}$, denoted as $\alpha_i(v_t)$, is a nonnegative quadratic function of v_t with deterministic coefficients. So on $\{\theta_{t-1} = i\}$, there exists a deterministic v_{ti} that minimizes $\alpha_i(v_t)$. Thus, defining $V_t \triangleq (v_{t1}, \dots, v_{t2})$, we get

$$\begin{aligned} \min_{v_t \in \mathcal{F}_t} E[(a(\theta_t) - v_t)^2 | \mathcal{F}_t] &= \sum_{i,j=1}^N (a_j - v_{ti})^2 p_{ij} I_{(\theta_{t-1}=i)} \\ &= e^T \cdot C^t(V_t) \cdot [I_{(\theta_{t-1}=1)}, \dots, I_{(\theta_{t-1}=N)}]^T \quad \text{a.s.} \quad \square \end{aligned}$$

Proof of Lemma 3.4. If we denote

$$J \triangleq \beta_k^T \cdot EC^t(v_k) \cdot [I_{(\theta_{k-1}=1)} y_k^2, \dots, I_{(\theta_{k-1}=N)} y_k^2]^T,$$

then J must have the form: $J = E[\alpha_1(v_k) \cdot I_{(\theta_{k-1}=1)} + \dots + \alpha_N(v_k) \cdot I_{(\theta_{k-1}=N)}] \cdot y_k^2$, where $\alpha_j(v_k), j = 1, \dots, N$ are nonnegative quadratic polynomials of v_k . So, there exists a unique deterministic $v_{kj} \in \mathbb{R}^1$ that makes $\alpha_j(v_k)$ reach its minimum for any j . Hence, a lower bound to J is $J \geq E[\alpha_1(v_{k1})I_{(\theta_{k-1}=1)} + \dots + \alpha_N(v_{kN})I_{(\theta_{k-1}=N)}]y_k^2, \forall v_k \in \mathcal{F}_k$. Now, if we set $v_k = \sum_{j=1}^N v_{kj} \cdot I_{(\theta_{k-1}=j)}$, then J can reach its minimum, and obviously, $v_k \in \sigma(\theta_{k-1})$. \square

Appendix B. Proofs of Lemmas 3.5–3.8 and Corollary 3.1

Proof of Lemma 3.5. By Lemma 3.2, we see V_t has a limit X_∞^* and so $\lim_{t \rightarrow \infty} C(V_t) = C(X_\infty^*)$. Also, by $\rho(C(X_\infty^*)) = \rho^*$, it is clear that we can select a $\delta > 0$ small enough such that $\rho^* < \rho(C(X_\infty) + \delta \cdot ee^\tau) < 1$. Thus, if we denote $C_* \triangleq C(X_\infty^*) + \delta \cdot ee^\tau$ and $\rho_* \triangleq \rho(C(X_\infty^*) + \delta \cdot ee^\tau)$, then there exists $t_0 > 0$ such that $C(V_t) \leq C_*, \forall t \geq t_0$. This completes the proof. \square

Proof of Lemma 3.6. By the definition of V_t , we know that $\|V_t\| \leq \sqrt{N} \max_{1 \leq k \leq N} |a_k|$, so there exists $M > 0$ such that $[a_t - v_t I_{(y_{t-1} \neq 0)}]^2 \leq M, \forall t > 0$, where v_t is defined in (14). Hence denoting $\lambda_t = I_{(|w_t/y_{t-1}| \geq d/2, y_{t-1} \neq 0)}$, we have from (1)

$$\begin{aligned} E y_{t+1}^2 \lambda_t &= E \{ [a_t - v_t I_{(y_{t-1} \neq 0)}] y_t + w_{t+1} \}^2 \lambda_t \\ &\leq M \cdot E y_t^2 \lambda_t + \sigma_w^2 \\ &\leq 2M^2 \cdot E y_{t-1}^2 \lambda_t + 2M\sigma_w^2 + \sigma_w^2. \end{aligned} \tag{B.1}$$

Now, notice that

$$\lambda_t \leq I_{(|y_{t-1}| \leq \sqrt{(2M^2\sigma_w^2/\lambda)} \cdot 2/d)} + I_{(w_t^2 > 2M^2\sigma_w^2/\lambda)}.$$

Substituting this into (B.1), we have

$$\begin{aligned} E y_{t+1}^2 \lambda_t &\leq 2M^2 \cdot E y_{t-1}^2 I_{(|y_{t-1}| \leq \sqrt{(2M^2\sigma_w^2/\lambda)} \cdot 2/d)} \\ &\quad + 2M^2 \cdot E y_{t-1}^2 I_{(w_t^2 > 2M^2\sigma_w^2/\lambda)} + (2M + 1)\sigma_w^2 \\ &\leq 2M^2 \cdot \frac{2M^2\sigma_w^2}{\lambda} \cdot \frac{4}{d^2} \\ &\quad + 2M^2 \cdot P\left(w_t^2 > \frac{2M^2\sigma_w^2}{\lambda}\right) \cdot E y_{t-1}^2 + (2M + 1)\sigma_w^2 \\ &= \left[\frac{16M^4\sigma_w^2}{\lambda d^2} + (2M + 1)\sigma_w^2 \right] + \lambda \cdot E y_{t-1}^2. \quad \square \end{aligned}$$

Proof of Lemma 3.7. For any $i \in \{1, \dots, N\}$, we have by denoting $\lambda_t = I_{(|w_t/y_{t-1}| < d/2, y_{t-1} \neq 0)}$,

$$\begin{aligned} E I_{(\theta_t=i)} y_{t+1}^2 \cdot \lambda_t &\leq E \{ [a_i - v_{t1} I_{(y_{t-1} \neq 0)}]^2 y_t^2 \cdot \lambda_t + \sigma_w^2 \} I_{(\theta_t=i)} \\ &= E \{ E[(a_i - v_{t1} I_{(y_{t-1} \neq 0)})^2 \cdot I_{(\theta_t=i)} \cdot \lambda_t | \mathcal{F}_t] y_t^2 \} \\ &\quad + \sigma_w^2 \cdot P(\theta_t = i). \end{aligned} \tag{B.2}$$

By (1) it can be seen that on the set $\{|w_t/y_{t-1}| < d/2, y_{t-1} \neq 0\}$, $(y_t - u_{t-1})/y_{t-1} = a(\theta_{t-1}) + w_t/y_{t-1}$. Hence $\psi((y_t - u_{t-1})/y_{t-1}) = \theta_{t-1}$, on the set $\{|w_t/y_{t-1}| < d/2, y_{t-1} \neq 0\}$. Substituting this into (14), we know that on the set $\{|w_t/y_{t-1}| < d/2, y_{t-1} \neq 0\}$, $v_t I_{(y_{t-1} \neq 0)} = v_{t1} I_{(\theta_{t-1}=1)} + \dots + v_{tN} I_{(\theta_{t-1}=N)}$. So, by (B.2),

$$\begin{aligned} E I_{(\theta_t=i)} y_{t+1}^2 \cdot \lambda_t &= E \left\{ E \left[\left(a_i - \sum_{j=1}^N v_{tj} I_{(\theta_{t-1}=j)} \right)^2 I_{(\theta_t=i)} \lambda_t | \mathcal{F}_t \right] y_t^2 \right\} \\ &\quad + \sigma_w^2 P(\theta_t = i) \\ &= E \{ E[(a_i - v_{t1} I_{(\theta_{t-1}=1)} - \dots - v_{tN} I_{(\theta_{t-1}=N)})^2 \cdot I_{(\theta_t=i)} | \theta_{t-1}] \cdot \lambda_t y_t^2 \} \\ &\quad + \sigma_w^2 \cdot P(\theta_t = i) \\ &= E \sum_{k=1}^N I_{(\theta_{t-1}=k)} (a_i - v_{tk})^2 p_{ki} \lambda_t y_t^2 \\ &\quad + \sigma_w^2 \cdot P(\theta_t = i) \\ &\leq \sum_{k=1}^N (a_i - v_{tk})^2 p_{ki} \cdot E I_{(\theta_{t-1}=k)} y_t^2 + \sigma_w^2 \cdot P(\theta_t = i). \end{aligned}$$

Hence, by this, (16) and (2), we have

$$\gamma_{t+1} \leq C(V_t)^\tau \cdot \begin{bmatrix} E I_{(\theta_{t-1}=1)} y_t^2 \\ \vdots \\ E I_{(\theta_{t-1}=N)} y_t^2 \end{bmatrix} + \sigma_w^2 \pi_t.$$

Now, it is easy to verify that if $0 \leq A_1 \leq B_1$, and $0 \leq A_2 \leq B_2$, then $A_1 A_2 \leq B_1 B_2$. Hence, by denoting $x_t \triangleq [I_{(\theta_{t-1}=1)} y_t^2 \dots I_{(\theta_{t-1}=N)} y_t^2]^\tau$, it follows from Lemma 3.5 that for $t \geq t_0$,

$$\begin{aligned} \gamma_{t+1} &\leq C_*^\tau E x_t + \sigma_w^2 \pi_t \\ &= C_*^\tau \gamma_t + C_*^\tau E x_t I_{(y_{t-2}=0)} + C_*^\tau E x_t \lambda_{t-1} + \sigma_w^2 \pi_t \\ &\leq C_*^\tau \gamma_t + C_*^\tau \cdot E y_t^2 I_{(y_{t-2}=0)} \cdot e \\ &\quad + C_*^\tau \cdot E y_t^2 \lambda_{t-1} \cdot e + \sigma_w^2 e. \end{aligned} \tag{B.3}$$

By (14), we have $u_{t-2} = 0$ and $u_{t-1} = 0$ on the set $\{y_{t-2} = 0\}$. So, by (1),

$$\begin{aligned} E y_t^2 I_{(y_{t-2}=0)} &= E [a(\theta_{t-1}) w_{t-1} + w_t]^2 I_{(y_{t-2}=0)} \\ &\leq E (a(\theta_{t-1}) w_{t-1} + w_t)^2 \\ &\leq 2 \max_{1 \leq k \leq N} |a_k| \cdot \sigma_w^2 + 2\sigma_w^2. \end{aligned}$$

Therefore, substituting this into (B.3) and by Lemma 3.6, we have

$$\begin{aligned} \gamma_{t+1} &\leq C_*^\tau \gamma_t + \left[2 \max_{1 \leq k \leq N} |a_k| \cdot \sigma_w^2 + 2\sigma_w^2 \right. \\ &\quad \left. + M_\lambda + \lambda E y_{t-2}^2 \right] C_*^\tau e + \sigma_w^2 e. \quad \square \end{aligned}$$

Proof of Corollary 3.1. By Lemma 3.7, we have

$$\begin{aligned} \gamma_{t+1} &\preceq C_*^\tau \gamma_t + M^* C_*^\tau e + \lambda E y_{t-2}^2 C_*^\tau e + \sigma_w^2 e \\ &\preceq (C_*^\tau)^2 \gamma_{t-1} + M^* (C_*^\tau)^2 e + \lambda E y_{t-3}^2 (C_*^\tau)^2 e \\ &\quad + \sigma_w^2 C_*^\tau e + M^* C_*^\tau e + \lambda E y_{t-2}^2 C_*^\tau e + \sigma_w^2 e \\ &\preceq \dots \preceq (C_*^\tau)^{t-t_0+1} \gamma_{t_0} + M^* (C_*^\tau)^{t-t_0+1} e \\ &\quad + \lambda E y_{t_0-2}^2 (C_*^\tau)^{t-t_0+1} e + \sigma_w^2 (C_*^\tau)^{t-t_0} e \\ &\quad + \dots + M^* C_*^\tau e + \lambda E y_{t-2}^2 C_*^\tau e + \sigma_w^2 e. \end{aligned}$$

Now, by Lemma 3.5 we know that there exist constants $c_0 > N$ and $\rho \in (\rho^*, 1)$ such that $e^\tau (C_*^\tau)^k e \leq c_0 \rho^k$, $\forall k \geq 0$. Hence, by the fact that $\gamma_{t_0} \preceq E y_{t_0}^2 \cdot e$, we have for $t \geq t_0$,

$$\begin{aligned} e^\tau \gamma_{t+1} &\leq e^\tau (C_*^\tau)^{t-t_0+1} \gamma_{t_0} + M^* e^\tau (C_*^\tau)^{t-t_0+1} e \\ &\quad + \lambda E y_{t_0-2}^2 e^\tau (C_*^\tau)^{t-t_0+1} e + \sigma_w^2 e^\tau (C_*^\tau)^{t-t_0} e \\ &\quad + \dots + M^* e^\tau C_*^\tau e + \lambda E y_{t-2}^2 e^\tau C_*^\tau e + \sigma_w^2 e^\tau e \\ &\leq c_0 e^\tau \rho^{t-t_0+1} e \cdot E y_{t_0}^2 + M^* \cdot c_0 (\rho^{t-t_0+1} + \dots + \rho) \\ &\quad + \lambda E y_{t_0-2}^2 c_0 \rho^{t-t_0+1} + \lambda E y_{t_0-1}^2 c_0 \rho^{t-t_0} + \dots \\ &\quad + \lambda E y_{t-2}^2 c_0 \rho + \sigma_w^2 c_0 \cdot (\rho^{t-t_0} + \dots + \rho + 1) \\ &\leq c_0 N \cdot E y_{t_0}^2 + \frac{M^* c_0}{1-\rho} + \lambda E y_{t_0-2}^2 c_0 \rho^{t-t_0+1} \\ &\quad + \lambda E y_{t_0-1}^2 c_0 \rho^{t-t_0} + \dots + \lambda E y_{t-2}^2 c_0 \rho + \frac{\sigma_w^2 c_0}{1-\rho}. \end{aligned}$$

Finally, by denoting $M_1 \triangleq c_0 N \cdot E y_{t_0}^2 + M^* c_0 / (1-\rho) + \sigma_w^2 c_0 / (1-\rho)$, we get the desired result. \square

Proof of Lemma 3.8. Obviously, when $k = k_0$ the inequality is valid. We proceed to prove the result by induction. Suppose that when $k \leq K$, we have $b_k \leq \max\{b_{k_0}, b_{k_0+1}, M/(1-\eta)\}$. Consequently, $M_K \triangleq \max_{k_0 \leq i \leq K} b_i \leq \max\{b_{k_0}, b_{k_0+1}, M/(1-\eta)\}$. If $M_K \leq M/(1-\eta)$, then by (19), $b_{K+1} \leq M + \eta \cdot M_K \leq M + M\eta/(1-\eta) = M/(1-\eta)$; Otherwise, if $M_K > M/(1-\eta)$, then $M + \eta M_K < M_K$. Hence by (19), $b_{K+1} \leq M + \eta \cdot M_K < M_K \leq \max\{b_{k_0}, b_{k_0+1}, M/(1-\eta)\}$. Hence, in any case we have $b_{K+1} \leq \max\{b_{k_0}, b_{k_0+1}, M/(1-\eta)\}$, and the proof is completed by the induction. \square

References

Aström, K. J., & Wittenmark, B. (1995). *Adaptive control*. Reading, MA: Addison-Wesley.
 Benveniste, A. (1987). Design of adaptive algorithms for the tracking of time-varying systems. *International Journal of Adaptive Control Signal Processing*, 1, 3–29.
 Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.
 Guo, L. (1990). On adaptive stabilization of time-varying stochastic systems. *SIAM Journal of Control and Optimization*, 28, 1432–1451.

Guo, L., & Ljung, L. (1995). Performance analysis of general tracking algorithms. *IEEE Transactions on Automatic Control*, 40, 1388–1402.
 Horn, R. A., & Johnson, C. R. (1985). *Matrix analysis*. Cambridge: Cambridge University Press.
 Ioannou, P., & Tsakalis, K. (1985). Robust discrete-time adaptive control. In: K. S. Narendra (Ed.), *Adaptive and learning systems: Theory and applications*. pp. 73–85. New York, Plenum Press.
 Meyn, S. P., & Brown, L. J. (1993). Model reference adaptive control of time varying and stochastic systems. *IEEE Transactions on Automatic Control*, 38, 1738–1753.
 Middleton, R. H., & Goodwin, G. C. (1988). Adaptive control of time-varying linear systems. *IEEE Transactions on Automatic Control*, 33, 150–157.
 Millnert, M. (1987). Identification of ARX models with Markovian parameters. *International Journal of Control*, 45(6), 2045–2058.
 Zames, G., Lin, L., & Wang, L. Y. (1994). Fast identification n-widths and uncertainty principles for LTI and slowly varying systems. *IEEE Transactions on Automatic Control*, 39, 1827–1838.



Lei Guo was born in China in 1961. He received the B.S. degree in mathematics from Shandong University in 1982, and the Ph.D. degree in control theory from the Chinese Academy of Sciences in 1987. He has been a postdoctoral fellow at the Australian National University (1987–1989), and a visiting professor at several universities in the US and Europe. Since 1992, he has been a professor of the Institute of Systems Science at the Chinese Academy of Sciences, where he is currently the director of the Institute. His research interests include identification and control of stochastic, uncertain, linear and nonlinear systems.

Dr. Guo is a Fellow of IEEE. He has twice received, in 1987 and 1997, the National Natural Science Prize of China. He is also the recipient of the 1993 IFAC World Congress Young Author Prize. He is currently serving as the chairman of the IFAC Technical Committee on Modeling, Identification and Signal Processing (1999–2001), is the Editor-in-Chief of the *Journal of Systems Science and Complexity*, and is on the editorial boards of several other technical journals in the areas of systems, control and mathematics.



Minyi Huang received the BS degree from Shandong University, in 1995, and the MS degree from Institute of Systems Science, Chinese Academy of Sciences in 1998. He is currently working towards his Ph.D. degree in systems and control at McGill University, Montreal, Canada. His research interests include hierarchical and complex systems, optimization of wireless communication networks, and stochastic adaptive control.



Feng Xue received the B.S. degree in Control Science from Shandong University, China, in 1997, and the M.S. degree in Control Theory and Operational Research from Institute of Systems Science, Chinese Academy of Sciences in 2000. He is currently a Ph.D. student at the Department of Electrical and Computer Engineering, University of Illinois at Urbana Champaign. His research interests include stochastic systems, sampled-data control and network information theory.