Communications in Information and Systems
Vol. 6, No. 4, pp. 299-320, 2006

003

ADAPTIVE CONTROL OF LINEAR TIME INVARIANT SYSTEMS: THE "BET ON THE BEST" PRINCIPLE*

S. BITTANTI† AND M. C. CAMPI‡

Abstract. Over the last three decades, the certainty equivalence principle has been the fundamental paradigm in the design of adaptive control laws. It is well known, however, that for general control criterions the performance achieved through its use is strictly suboptimal. In order to overcome this difficulty, two different approaches have been proposed: i) the use of cost-biased parameter estimators; and ii) the injection of probing signals into the system so as to enforce consistency in the parameter estimate.

This paper presents an overview of the cost-biased approach. New insight is achieved in this paper by the formalization of a general cost-biased principle named "Bet On the Best"-BOB. BOB may work in situations in which more standard implementations of the cost-biasing idea may fail to achieve optimality.

Key words: adaptive control; stochastic systems; certainty equivalence principle; long-term average cost; optimality.

1. Introduction: an overview of adaptation as a means to achieve an "ideal" control objective. An adaptive control problem is a control problem in which some parameter describing the system is known with uncertainty. During the operation of the control system, the controller collects information on the system behavior, thereby reducing the level of uncertainty regarding the value of the parameter. In turn, as the level of uncertainty is reduced, the controller is tuned more accurately on the system parameter so as to obtain a better control result. In this procedure it is essential that the controller chooses the control actions so as to minimize the performance index, as well as probe the system so that uncertainty is reduced to better select future control actions.

In this paper we consider adaptive long-term average optimal control problems. In adaptive control, due to the uncertainty affecting the true value of the system parameter, the control law cannot be expected to be optimal in finite time. When the cost criterion is of the long-term average type, however, the control performance in finite time does not affect the asymptotic value of the control cost. Hence, even in an adaptive context there is a hope to achieve optimality, i.e. to drive the long-term

*The authors would like to acknowledge the financial support of the National Interest Research Project "Innovative techniques for identification and adaptive control of industrial systems".

†Dept. of Electrical Engineering and Information, Politecnico di Milano, piazza L. da Vinci 38, 20123, Milano, Italy. E-Mail: bittanti@elet.polimi.it

‡Dept. of Electrical Engineering and Automation, University of Brescia, via Branze 38, 25123 Brescia, Italy. E-Mail: marco.campi@ing.unibs.it

average cost to the value which would have been obtained under complete knowledge of the system. When this happens, we say that the adaptive control law meets the *ideal objective*.

A control problem with an unknown system parameter is equivalent to a control problem with complete system knowledge in which the state comprises the set of all probability distributions on the unknown parameter, Striebel (1965). This theoretical result, however, does not translate into practical solution methods due to the complexity involved in handling the corresponding infinite dimensional problem. To make the problem tractable, it is common practice to resort to special solution methods able to abate the computational complexity.

The most common special solution methods rely on the so-called *certainty equivalence principle*, Bar-Shalom and Tse (1974), Bar-Shalom and Wall (1974). The unknown parameter is estimated via some estimation method and the estimate is used as if it were the true value of the unknown parameter. In this approach, the distribution of the unknown parameter is simply substituted by a single estimate representing, in some sense, the most probable value of it.

Certainty equivalent adaptive control schemes have been studied by many authors. Goodwin et al. (1981) prove that a certainty equivalent controller based on the stochastic approximation algorithm achieves the ideal objective for minimum output variance costs. This result has been extended to least squares minimum output variance adaptive control in Sin and Goodwin (1982), Bittanti et al. (1990), Campi (1991), and Bittanti and Campi (1996). A complete analysis of a minimum output variance self-tuning regulator equipped with the extended least squares algorithm can be found in Guo and Chen (1991). Again, the main result is that this adaptive scheme achieves the ideal objective.

The fact that the ideal objective is met in the situations described in the above mentioned papers is due to the special properties of the minimum output variance cost criterion. On the other hand, it is well known that the certainty equivalence principle suffers from a general identifiability problem, namely the parameter estimate can converge with positive probability to a false value, e.g. Aström and Wittenmark (1973), Becker et al. (1985), Campi (1996), Campi and Kumar (1998). When a cost criterion other than the output variance is considered, this identifiability problem leads to a strictly suboptimal performance. See e.g. Lin et al. (1985), Polderman (1986a,b), and van Schuppen (1994) for a discussion on this problem in different contexts.

In order to overcome this problem, two approaches have been proposed in the literature. The first one consists in adding a dither noise to the control input so as to improve the excitation characteristics of the signals, Caines and Lafortune (1984). As a consequence, standard parameter estimators are then able to provide consistent estimates and the mentioned identifiability problem automatically disappears. However, as noted by Chen and Guo (1987a), this may result in a degradation of the control system performance. Asymptotic optimality is recovered by letting the additive noise vanish in the long run (attenuating excitation). Many optimality results have been established along this line, Chen and Guo (1986, 1987a,b, 1988, 1991), Guo and Chen (1991), Guo (1996), Duncan et al. (1999), while persistence of excitation conditions of different types have been used in Duncan and Pasik-Duncan (1986,1991), Caines (1992).

The second approach has its origins in the work by Kumar and Becker (1982). It consists in the employment of a cost-biased parameter estimator, and does not require the use of any extra probing signal. The basic idea is as follows. Consider a standard (i.e. without biasing) estimator operating in a closed-loop adaptive control system. It is natural to expect that this estimator is able to correctly describe the closed-loop behavior of the system. Thus, one expects that the asymptotic behavior of the true system with the loop closed by the adaptively chosen controller will be the same as the behavior of the estimated system with the loop closed by the same controller. This implies that the long-term average cost associated with these two control systems will be the same. Since the adaptive controller is selected to be optimal for the estimated system, this also means that the adaptively controlled true system attains the optimal performance for the estimated system. On the other hand, the fact that the estimator is able to describe the closed-loop behavior of the true system by no means implies that the true system has been correctly estimated. As a matter of fact, it is possible that the estimated system and the true system share the same behavior in the actual closed-loop conditions, while they would behave differently in other situations. Even more so, it can be the case that if one knew the true system at the start, an optimal controller for it could be designed that outdoes the performance obtained by the adaptively chosen controller. These observations carry two consequences. First, the adaptive controller can be strictly suboptimal. Second, if this is the case, then the asymptotically estimated system has associated an optimal cost which is strictly larger than the optimal cost for the true system. In this way, we come to the conclusion that the standard parameter estimator has a natural tendency to return estimates with an optimal cost larger than or equal to that of the true system and, if it is strictly larger, this leads to a strictly suboptimal performance.

Motivated by this observation, Kumar and Becker (1982) conceived of introducing a cost-biasing term in the parameter estimator that favors those parameter estimates corresponding to a smaller optimal cost. The cost-bias must be strong enough such that the estimator can never stick at a parameter estimate with an optimal cost larger than the true one. At the same time, however, it must be delicate so that the ability of identifying the closed-loop dynamics is not destroyed. If these two objectives are met simultaneously, then the performance of the true closed-loop system will be the same as the one of the estimated closed-loop system. Moreover, the latter is no worse than the optimal performance for the true system due to the cost-biasing and, therefore, optimality is achieved. This approach has been investigated in different contexts in the following papers, Kumar and Becker (1982), Kumar (1983a,b), Milito and Cruz (1987), Borkar (1993), Campi and Kumar (1998), Prandini and Campi (2001).

Objective of this paper

This paper is primarily an overview of cost-biased adaptation as a means to achieve optimality. Additionally, the cost-biased idea is here cast into a novel and fruitful viewpoint via the introduction of a new principle named "Bet On the Best" - BOB. BOB bears a promise of more general applicability than standard implementations of the cost-biasing idea.

Structure of the paper

The structure of the paper is as follows. The BOB principle is presented in Sections 2 and 3. As an example, in Section 4 the BOB principle is applied to a scalar adaptive linear quadratic Gaussian (LQG) control problem.

2. The adaptive control setting. This section serves the purpose of introducing the general control set-up and that of fixing notations. Explicit assumptions on the stochastic nature of signals are delayed to subsequent sections. Measurability conditions are assumed for granted throughout.

Consider a linear time invariant system described as

$$x_{t+1} = A(\theta^\circ)x_t + B(\theta^\circ)u_t + w_{t+1}^{(1)}, \tag{1}$$

$$y_t = C(\theta^\circ)x_t + w_{t+1}^{(2)}, \tag{2}$$

where $x_t \in \mathbf{R}^n$ is the state, $u_t \in \mathbf{R}^1$ is the control variable, $y_t \in \mathbf{R}^1$ is the system output, $w_{t+1}^{(1)}$ and $w_{t+1}^{(2)}$ are noise processes. θ° is an unknown true parameter belonging to a given parameter set Θ.

The adaptive control process takes place as follows. At time t the adaptive controller has access to the observations $o_t = \{u_1, u_2, \ldots, u_{t-1}, y_1, y_2, \ldots, y_t\}$. Based on this, it selects the control input u_t. As a consequence of this control action, the state transits from x_t to x_{t+1} according to equation (1), a new output y_{t+1} generated according to equation (2) becomes available and a cost $c(u_t, y_t)$ is paid. Then, the observation set is updated to $o_{t+1} = o_t \cup \{u_t, y_{t+1}\}$ and the controller selects the subsequent control input. A control law is a sequence of functions $l_t : \mathbf{R}^{t-1} \times \mathbf{R}^t \to \mathbf{R}^1$, and $l_t(o_t)$ is the corresponding control input after we have observed $o_t = \{u_1, u_2, \ldots, u_{t-1}, y_1, y_2, \ldots, y_t\}$.

The control objective is to minimize the long-term average cost criterion

$$\mathcal{J} = \limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} c(u_t, y_t).$$

For any $\theta \in \Theta$ and for a system as in (1) and (2) with θ in place of θ°, we assume that, for any control law l_t, $\mathcal{J} \geq \mathcal{J}_\theta^\star$ *a.s.* (almost surely), where $\mathcal{J}_\theta^\star$ is a deterministic quantity, and that $\mathcal{J}_\theta^\star$ is achieved a.s. by applying a control law $l_{\theta,t}^\star$. $l_{\theta,t}^\star$ is named an optimal control law. Our objective is that of driving $\mathcal{J}$ to the optimal value $\mathcal{J}_{\theta^\circ}^\star$ for the true system with parameter θ°. In the actual implementation of a control action, however, θ° is not known and, therefore, information regarding its value must be accrued through time via the observations u_t and y_t (*adaptive control problem*).

3. The "Bet On the Best" (BOB) principle. The BOB principle has first appeared in the conference paper Campi (1997). This is the first time this principle is discussed in a journal paper.

We start with an example in which the certainty equivalence principle leads to a control cost which is strictly suboptimal. A similar example is also provided in Kumar (1983b), where, differently from the present case, a finite parameter set is considered. This example will serve as a start for the subsequent discussion where we first summarize some well-recognized facts regarding the certainty equivalence approach. The discussion will then culminate in the formulation of the BOB principle.

EXAMPLE 1. Consider the system

$$x_{t+1} = a^\circ x_t + b^\circ u_t + w_{t+1},$$

where w_t is an i.i.d. $N(0,1)$ noise process and state x_t is accessible: $y_t = x_t$. Vector $[\ a^\circ \ \ b^\circ\ ]$ is unknown but we know that it belongs to a compact set $\Theta = \{[\ a \ \ b\ ] : b = 8a/5 - 3/5, a \in [0,1]\}$. Our objective is to minimize the long-term average cost $\limsup_{N\to\infty} 1/N \sum_{t=1}^{N} [q x_t^2 + u_t^2]$, where $q = 25/24$.

In order to determine an estimate of $[\ a^\circ \ \ b^\circ\ ]$ the standard least squares algorithm is used. This amounts to selecting at time t the vector $[\ a_t^{LS} \ \ b_t^{LS}\ ]$ which minimizes the index $\sum_{k=1}^{t-1} (x_{k+1} - a x_k - b u_k)^2$. Once estimate $[\ a_t^{LS} \ \ b_t^{LS}\ ]$ has been determined, according to the certainty equivalence principle the optimal control law for parameter $[\ a_t^{LS} \ \ b_t^{LS}\ ]$ is applied.

Suppose now that at a certain instant point t the least squares estimate is $[\ a_t^{LS} \ \ b_t^{LS}\ ] = [\ 1 \ \ 1\ ]$. Since the corresponding optimal control law is given by $u_t = -5/8\, x_t$ (see e.g. Bertsekas (1987)), the squared error at time $t+1$ turns out to be

$(x_{t+1}-ax_t-bu_t)^2 = (x_{t+1}-ax_t-(8a/5-3/5)(-5/8\,x_t))^2 = (x_{t+1}-3/8\,x_t)^2, \forall [\,a\ \ b\,] \in \Theta$. The important feature of this last expression is that it is independent of parameter $[\,a\ \ b\,] \in \Theta$. Hence, the term added at time $t+1$ to the least squares index does not influence the location of its minimizer and the least squares estimate remains unchanged at time $t+1$: $[\,a^{LS}_{t+1}\ \ b^{LS}_{t+1}\,] = [\,a^{LS}_t\ \ b^{LS}_t\,] = [\,1\ \ 1\,]$. As the same rationale can be repeated in the subsequent instant points, we can conclude that the estimate sticks at $[\,1\ \ 1\,]$.

Now, the important fact is that the least squares estimates can in fact take value $[\,1\ \ 1\,]$ with positive probability, even when the true parameter is different from $[\,1\ \ 1\,]$. Moreover, the optimal cost for the true parameter may be strictly lower than the incurred cost obtained by applying the optimal control law for parameter $[\,1\ \ 1\,]$. To see that this is the case, suppose that $[\,a^\circ\ \ b^\circ\,] = [\,0\ \ -3/5\,]$ and assume that the system is initialized with $x_1 = 1$ and $u_1 = 0$. Then, at time $t = 2$ the least squares estimate minimizes the cost $(x_2 - a)^2 = (w_2 - a)^2$. Thus, $[\,a^{LS}_2\ \ b^{LS}_2\,] = [\,1\ \ 1\,]$ whenever $w_2 > 1$, which happens with positive probability. In addition, it is easily seen that the cost associated with the optimal control law for parameter $[\,1\ \ 1\,]$ is $5/3$ whereas the optimal cost for the true parameter $[\,a^\circ\ \ b^\circ\,] = [\,0\ \ -3/5\,]$ is $25/24$. $\square$

A careful analysis of the example above reveals where the trouble comes from with a straightforward use of the certainty equivalence principle. When the suboptimal control $u_t = -5/8\,x_t$ is selected based on the current estimate $[\,a^{LS}_t\ \ b^{LS}_t\,] = [\,1\ \ 1\,]$, the resulting observation is $y_{t+1} = x_{t+1} = 3/8\,x_t + w_{t+1}$. This observation is in perfect agreement with the one which would have been obtained if $[\,a^{LS}_t\ \ b^{LS}_t\,] = [\,1\ \ 1\,]$ were the true parameter. Therefore, there is no reason for having doubts as to the correctness of the estimate $[\,a^{LS}_t\ \ b^{LS}_t\,]$ and thus this estimate is kept unchanged at the next time point.

This is just a single example of a general estimability problem arising in adaptive control problems. This general estimability problem can be described as follows:

- applying to the true system a control which is optimal for the estimated system may result in observations which concur with those that would have been obtained if the estimated system were the true system;

if the estimation method drives the estimate to a value such that the above happens, then

- there is no clue that the system is incorrectly estimated and, consequently, the estimate remains unchanged;

however,

- the adopted control law is optimal for the estimated system, while it may be *strictly* suboptimal for the true system.

A way out of this pernicious mechanism is to employ a more fine grained estimation method based on the optimal long-term average cost for the different systems with parameters $\theta \in \Theta$. Developing this idea will lead us to the formulation of the "Bet On the Best principle".

We start by observing the following elementary fact:

- suppose we apply to the true system a control law which is optimal for another system. If the long-term average cost we pay is different from the optimal cost for this second system, then this system is falsified by the observations and it can be dropped from the set of possible true systems.

Suppose now that at a certain instant point, we select among the systems which are still unfalsified the one with lower optimal cost. Then,

- if we pay a cost different from the expected one, we can falsify this system. In the opposite, we cannot falsify it, but then we are paying a cost which is minimal over the set of possible true systems. Indeed, this implies that we are actually paying the optimal cost for the true system.

These considerations can be summarized as follows: selecting a control law which is optimal for the best unfalsified system (i.e. the system with lower optimal cost among those that are as far unfalsified by the observations) may lead to an estimability problem only when we are achieving optimality. This is in contrast with what happens with the straightforward certainty equivalence principle, where an estimability problem may arise and, yet, the incurred cost may be strictly suboptimal.

The above observations suggest that a very natural way to overcome the estimability problem posed by the certainty equivalence principle is simply to iteratively select among the unfalsified systems the one with minimal optimal cost and then apply the optimal control law for it. We then arrive at formulating the following procedure of general validity:

The "Bet On the Best" (BOB) principle

At the generic instant point t, do the following:

1. determine the set of unfalsified systems;
2. select the system in the unfalsified set with lowest optimal cost;
3. apply the decision which is optimal for the selected system. □

3.1. Mathematical formalization of the BOB-principle. In this section we more precisely formalize the concept of unfalsified system and exhibit in mathematical terms the properties of the unfalsified set such that applying the BOB-principle leads to optimality.

Let $\mathcal{U}_t$ denote the unfalsified set at time t. Clearly, this set will depend on the observations $o_t = \{u_1, u_2, \ldots, u_{t-1}, y_1, y_2, \ldots, y_t\}$ available at time t, and so it is in fact a stochastic set. Moreover, we note that set $\mathcal{U}_t$ depends through $u_1, u_2, \ldots, u_{t-1}$ on the control law l_k applied from time $k = 1$ to time $k = t - 1$. Once the control law has been fixed, processes u_t and y_t are completely determined and so is the sequence of unfalsified sets $\mathcal{U}_t$. The question that we need now to address is: what are the mathematical conditions $\mathcal{U}_t$ has to satisfy so that application of the BOB principle leads to optimality? This question is answered in this section.

Assume that $\arg\min_{\theta \in \mathcal{U}_t} \mathcal{J}_\theta^\star$ exists and call it θ_t^{min}. Select the optimal control action for θ_t^{min}: $u_t = l^\star_{\theta_t^{min},t}(o_t)$. Then,

Condition i)

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} c(u_t, y_t) = \limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{J}^\star_{\theta_t^{min}} \quad a.s.$$

Condition ii)

$$\theta^\circ \in \cup_t \cap_{k>t} \mathcal{U}_k \quad a.s.$$

Securing condition i) appears a doable objective under general circumstances. In fact, if the long-term average cost paid by applying the optimal control law for θ_t^{min} were different from the expected average cost, there would be evidence that such a θ_t^{min} has to be falsified (and, therefore, θ_t^{min} should not be in $\mathcal{U}_t$ for some t).

Condition ii) simply says that the falsification procedure must not be overselective so that it also falsifies the true system (note that considering $\cup_t \cap_{k>t} \mathcal{U}_k$ rather than the straightforward $\mathcal{U}_t$ allows for transient phenomena due to stochastic fluctuations).

The following simple theorem points out the effectiveness of the BOB-principle when conditions i) and ii) are met.

Theorem 1. *Under conditions i) and ii), the BOB-procedure achieves the ideal objective, i.e.*

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} c(u_t, y_t) = \mathcal{J}^\star_{\theta^\circ} \quad a.s.$$

Proof. Condition ii) implies that $\theta^\circ \in \mathcal{U}_t, \forall t \geq \bar{t}$, where $\bar{t}$ is a suitable instant point, a.s. From this, $\inf_{\theta \in \mathcal{U}_t} \mathcal{J}^\star_\theta \leq \mathcal{J}^\star_{\theta^\circ}, \forall t \geq \bar{t}$, a.s. Since, according to the BOB-procedure, at each instant point t we select in $\mathcal{U}_t$ the parameter θ_t^{min} with lower optimal cost $\mathcal{J}^\star_{\theta_t^{min}}$, we obtain $\limsup_{N\to\infty} 1/N \sum_{t=1}^{N} \mathcal{J}^\star_{\theta_t^{min}} \leq \limsup_{N\to\infty} 1/N \sum_{t=1}^{N} \mathcal{J}^\star_{\theta^\circ} = \mathcal{J}^\star_{\theta^\circ}$. Thus, applying condition i) yields

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} c(u_t, y_t) = \limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{J}^\star_{\theta_t^{min}} \leq \mathcal{J}^\star_{\theta^\circ}, \quad a.s.,$$

so concluding that the incurred cost is optimal. $\square$

3.2. Discussion. The BOB principle bears a promise of more general applicability than other formulations of the cost-biasing idea.

In previous contributions such as Kumar (1993b) and Campi and Kumar (1998), identification was based on a delicate two-term estimation criterion, where the first term was the standard maximum likelihood and the second term was a cost-biasing term. A correct estimation of the closed-loop dynamics was relying on the presence of the first term, which, so to say, had not to be bogged down by the second term that was pushing the estimate towards parameter locations corresponding to lower optimal costs. In turn, this was calling for the presence of system noise that could suitably excite the true system. This delicate balance is automatically overcome with the BOB philosophy: if the estimate is biased towards parameters that correspond to a "superoptimal" cost, in the long run an average cost larger than the expected superoptimal cost will certainly be obtained and, therefore, according to condition i) in Section 3.1 this parameter will be discarded.

For a practical implementation of the BOB procedure, what remains to determine is the actual falsification rule. This determination is dependent on the specific control set-up. To make things concrete, in the next section we present an application to a scalar adaptive control problem. The recent interesting work by Levanony and Caines (2005) can be seen as another application of this same BOB principle. In this latter paper, the analysis is carried out for systems with multivariate state thanks to the observation that optimizing the LQG cost restricted to a region that shrinks around where closed-loop identification holds necessarily leads to a consistent estimate. See also Levanony and Caines (2001a,b) for a recursive implementation of the algorithm.

4. An application of the BOB principle: scalar adaptive LQG control.

4.1. Problem position. Consider the scalar system

$$x_{t+1} = a^\circ x_t + b^\circ u_t + w_{t+1}, \tag{3}$$

where w_t is a noise process described as an i.i.d. Gaussian sequence with zero mean and unitary variance. The true parameter $\theta^\circ = [\ a^\circ \ \ b^\circ\ ]$ is unknown and belongs to a known compact set $\Theta \subset \mathbf{R}^2$ such that $b \neq 0, \forall [\ a \ \ b\ ] \in \Theta$ (controllability condition). The system state is observed without noise, i.e. $y_t = x_t$. Finally, the long-term cost criterion is given by

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} [q x_t^2 + u_t^2], \qquad q > 0. \tag{4}$$

In the case in which the true parameter θ° is known, it is a standard matter to compute the optimal control law that minimizes criterion (4) (see e.g. Bertsekas

(1987)). Letting $p(a^\circ, b^\circ)$ be the positive solution to the scalar Riccati equation

$$p = \frac{(a^\circ)^2 p}{(b^\circ)^2 p + 1} + q,$$

the control input at time t is computed as

$$u_t = K(a^\circ, b^\circ) x_t,$$

where gain $K(a^\circ, b^\circ)$ is given by

$$K(a^\circ, b^\circ) = -\frac{a^\circ b^\circ p(a^\circ, b^\circ)}{(b^\circ)^2 p(a^\circ, b^\circ) + 1}.$$

The corresponding optimal cost is simply $\mathcal{J}^\star_{(a^\circ, b^\circ)} = p(a^\circ, b^\circ)$.

In the adaptive case where θ° is not known, we set the following

Adaptive control problem

Find a control law l_t such that, with the position $u_t = l_t(o_t)$, we achieve the ideal objective, i.e. $\limsup_{N\to\infty} 1/N \sum_{t=1}^N [q x_t^2 + u_t^2] = \mathcal{J}^\star_{(a^\circ, b^\circ)}$ a.s., $\forall [\ a^\circ \ \ b^\circ\ ] \in \Theta$. $\square$

4.2. Solving the adaptive control problem via the BOB-principle. To attack the adaptive control problem with the BOB-principle we need to find a suitable falsification criterion. The resulting unfalsified sets should satisfy conditions i) and ii) in Theorem 1.

A hint on how to select the unfalsified sets so as to satisfy condition ii) is provided by Lemma 1 below.

Name $[\ a_t^{LS} \ \ b_t^{LS}\ ]$ the least squares estimate of $[\ a^\circ \ \ b^\circ\ ]$:

$$[\ a_t^{LS} \ \ b_t^{LS}\ ] := \arg\min_{[a \ \ b] \in \mathbf{R}^2} \sum_{k=1}^{t-1} (x_{k+1} - a x_k - b u_k)^2,$$

and define $\phi_k := [\ x_k \ \ u_k\ ]$, and $V_t := \sum_{k=1}^{t-1} \phi_k^T \phi_k$.

Lemma 1. *Choose a function μ_t such that $\log \sum_{k=1}^{t-1} x_k^2 = o(\mu_t)$ and define the unfalsified set sequence through equation*

(5)

$$\mathcal{U}_t := \left\{ [\ a \ \ b\ ] \in \Theta : \quad ([\ a \ \ b\ ] - [\ a_t^{LS} \ \ b_t^{LS}\ ]) V_t ([\ a \ \ b\ ] - [\ a_t^{LS} \ \ b_t^{LS}\ ])^T \le \mu_t \right\}.$$

Then,

$$[\ a^\circ \ \ b^\circ\ ] \in \cup_t \cap_{k>t} \mathcal{U}_k \quad a.s.$$

Proof. The least squares estimate $[\ a_t^{LS} \ \ b_t^{LS}\ ]$ writes:

$$
\begin{aligned}
[\ a_t^{LS} \quad b_t^{LS}\ ] &= \left(\sum_{k=1}^{t-1} \phi_k x_{k+1}\right)\left(\sum_{k=1}^{t-1} \phi_k^T \phi_k\right)^{-1} \\
&= [\ a^\circ \quad b^\circ\ ] + \left(\sum_{k=1}^{t-1} \phi_k w_{k+1}\right)\left(\sum_{k=1}^{t-1} \phi_k^T \phi_k\right)^{-1}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&([\ a^\circ \quad b^\circ\ ] - [\ a_t^{LS} \quad b_t^{LS}\ ])V_t([\ a^\circ \quad b^\circ\ ] - [\ a_t^{LS} \quad b_t^{LS}\ ])^T \\
&= \left(\sum_{k=1}^{t-1} \phi_k w_{k+1}\right)\left(\sum_{k=1}^{t-1} \phi_k^T \phi_k\right)^{-1}\left(\sum_{k=1}^{t-1} \phi_k^T w_{k+1}\right).
\end{aligned}
$$

By applying result iii) in Lemma 1 of Lai and Wei (1982) we know that the right hand side of this last equation is $O(\log \sum_{k=1}^{t-1} \|\phi_k\|^2)$ *a.s.*. Moreover, by equation $|u_k| \leq \sup_{[a\ \ b]\in\Theta} |K(a,b)|\ |x_k|$ we know that $\log \sum_{k=1}^{t-1} \|\phi_k\|^2 = O(\log \sum_{k=1}^{t-1} x_k^2)$. In conclusion,

$$
([\ a^\circ \quad b^\circ\ ] - [\ a_t^{LS} \quad b_t^{LS}\ ])V_t([\ a^\circ \quad b^\circ\ ] - [\ a_t^{LS} \quad b_t^{LS}\ ])^T = O\left(\log \sum_{k=1}^{t-1} x_k^2\right) \quad a.s.
$$

The thesis immediately follows from this last equation by recalling that

$$
O(\log \sum_{k=1}^{t-1} x_k^2) = o(\mu_t).
$$

□

Lemma 1 delivers a lower bound for μ_t, the fulfillment of which implies that condition ii) is satisfied. Next, we need to determine a condition such that condition i) is satisfied as well. This will lead us to introduce an upper bound for μ_t.

We start by proving the following stability result.

THEOREM 2. *Choose a function μ_t such that $\mu_t = o(\log^2 \sum_{k=1}^{t-1} x_k^2)$ and set $u_t = K(a_t, b_t)x_t$, where $[\ a_t \quad b_t\ ]$ belongs almost surely to set $\mathcal{U}_t$ defined through equation (5). Then,*

$$
\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} [|x_t|^r + |u_t|^r] < \infty \quad a.s., \qquad \forall r.
$$

The proof of Theorem 2 is based on the following auxiliary lemma, the technical proof of which is given in Appendix A.

LEMMA 2. *Under the same assumptions as in Theorem 2 we have*

$$\sum_{t=1, t\notin \mathcal{T}_N}^{N} |(a^\circ - a_t)x_t + (b^\circ - b_t)u_t|^r = o(\sum_{t=1}^{N} |x_t|^r) \quad a.s., \qquad \forall r \geq 2.$$

where $\mathcal{T}_N$ is a set of time instant points depending on N such that $|\mathcal{T}_N| \leq 2$, $\forall N$ ($|\cdot|$ stands for cardinality).

Proof of Theorem 2.

Fix an integer N. For $t \in [1, N]$, rewrite system (3) as follows

$$x_{t+1} = \begin{cases} (a_t + b_t K(a_t, b_t))x_t + p_t + w_{t+1}, & t \notin \mathcal{T}_N \\ (a^\circ + b^\circ K(a_t, b_t))x_t + w_{t+1}, & t \in \mathcal{T}_N \end{cases} \tag{6}$$

where p_t is a perturbation term defined as

$$p_t := (a^\circ - a_t)x_t + (b^\circ - b_t)u_t,$$

and $\mathcal{T}_N$ is the set of instant points mentioned in Lemma 2.

Set

$$\alpha := \sup_{[a\ \ b] \in \Theta} |a^\circ + b^\circ K(a, b)|,$$

$$\rho := \sup_{[a\ \ b] \in \Theta} |a + bK(a, b)|.$$

Since $K(a, b)$ is the optimal gain for system $x_{t+1} = ax_t + bu_t + w_{t+1}$, the closed-loop dynamical matrix $a + bK(a, b)$ is stable, i.e. $|a + bK(a, b)| < 1$. Since Θ is a compact set, we then have $\rho < 1$.

With these positions, state x_t generated by system (6) can be bounded as follows

$$|x_t| \leq \alpha^2 \sum_{k=1, k\notin \mathcal{T}_N}^{t-1} \rho^{(t-1-k)-2} |p_k| + \alpha^2 \sum_{k=1}^{t-1} \rho^{(t-1-k)-2} |w_{k+1}| + \alpha^2 \rho^{(t-1)-2} |x_1|.$$

Form this,

$$\sum_{t=1}^{N} |x_t|^r \leq c \sum_{t=1, t\notin \mathcal{T}_N}^{N-1} |p_t|^r + c \sum_{t=1}^{N-1} |w_{t+1}|^r + c, \tag{7}$$

where c is a suitable constant.

Note now that $\sum_{t=1}^{N-1} |w_{t+1}|^r = O(N)$ a.s.. Moreover, in view of Lemma 2, for any $r \geq 2$ we have: $\sum_{t=1, t\notin \mathcal{T}_N}^{N-1} |p_t|^r = o(\sum_{t=1}^{N} |x_t|^r)$ a.s.. By substituting these estimates in (7) we obtain

$$\frac{1}{N} \sum_{t=1}^{N} |x_t|^r = o\left(\frac{1}{N} \sum_{t=1}^{N} |x_t|^r\right) + O(1) \quad a.s., \qquad \forall r \geq 2,$$

from which the conclusion is immediately drawn that

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}|x_t|^r < \infty \quad a.s., \qquad \forall r \geq 2.$$

Result $\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}|u_t|^r < \infty$ a.s., $\forall r \geq 2$ also follows by noting that $|u_t| \leq \sup_{[a\ \ b]\in\Theta}|K(a,b)||x_t|$. In conclusion,

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}[|x_t|^r + |u_t|^r] < \infty \quad a.s., \qquad \forall r \geq 2.$$

Finally, we observe that the boundedness result for $r \geq 2$ obviously implies that a similar result holds true for any r, so that the stability result remains proven for any r. $\square$

The next lemma gives an upper bound for μ_t such that condition i) in Theorem 1 is satisfied. This result used in conjunction with Lemma 1 provides us with the conditions such that the BOB-principle can be successfully used.

LEMMA 3. *Let* $\mu_t = \log^s \sum_{k=1}^{t-1} x_k^2$ *for some* $s < 2$ *and set* $u_t = K(a_t^{min}, b_t^{min})x_t$, *where* $[\ a_t^{min}\ \ b_t^{min}\ ] := \arg\min_{[\ a\ \ b\ ]\in\mathcal{U}_t} \mathcal{J}^\star_{[\ a\ \ b\ ]}$. *Then*

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}[qx_t^2 + u_t^2] = \limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\mathcal{J}^\star_{(a_t,b_t)} \quad a.s.$$

Proof. For notational simplicity, throughout we write a_t and b_t for a_t^{min} and b_t^{min}.

Let $\mathcal{F}_t := \sigma(w_1, w_2, \ldots, w_t)$.

The dynamic programming equation for model $x_{t+1} = a_t x_t + b_t u_t + w_{t+1}$ writes (Bertsekas (1987))

$$\begin{aligned}
&\mathcal{J}^\star_{(a_t,b_t)} + p(a_t,b_t)x_t^2 \\
&= qx_t^2 + u_t^2 + E[p(a_t,b_t)(a_t x_t + b_t u_t + w_{t+1})^2 \mid \mathcal{F}_t] \\
&= qx_t^2 + u_t^2 + E[p(a_t,b_t)x_{t+1}^2 \mid \mathcal{F}_t] + p(a_t,b_t)\left\{(a_t x_t + b_t u_t)^2 - (a^\circ x_t + b^\circ u_t)^2\right\},
\end{aligned}$$

from which

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N}\mathcal{J}^\star_{(a_t,b_t)} + \underbrace{\frac{1}{N}\sum_{t=1}^{N}\left\{p(a_t,b_t)x_t^2 - E[p(a_{t+1},b_{t+1})x_{t+1}^2 \mid \mathcal{F}_t]\right\}}_{A} \\
&= \frac{1}{N}\sum_{t=1}^{N}[qx_t^2 + u_t^2] + \underbrace{\frac{1}{N}\sum_{t=1}^{N}E[(p(a_t,b_t) - p(a_{t+1},b_{t+1}))x_{t+1}^2 \mid \mathcal{F}_t]}_{B} \\
&\quad + \underbrace{\frac{1}{N}\sum_{t=1}^{N}p(a_t,b_t)\left\{(a_t x_t + b_t u_t)^2 - (a^\circ x_t + b^\circ u_t)^2\right\}}_{C}. \qquad (8)
\end{aligned}$$

Let us study separately the different terms appearing in this equation.

A) Term (A) can be rewritten as

$$\begin{aligned}&\frac{1}{N}p(a_1,b_1)x_1^2-\frac{1}{N}p(a_{N+1},b_{N+1})x_{N+1}^2\\&+\frac{1}{N}\sum_{t=1}^{N}\left\{p(a_{t+1},b_{t+1})x_{t+1}^2-E[p(a_{t+1},b_{t+1})x_{t+1}^2\mid\mathcal{F}_t]\right\}.\end{aligned}$$

The first term obviously tends to zero. As for the second term, note that if we assume that it does not tend to zero, then there exists a time sequence t_k and a real number $\alpha>0$ such that $|x_{t_k}|^2>\alpha t_k$, $\forall k$. From this, $\limsup_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}|x_t|^4\geq \limsup_{k\to\infty}\frac{1}{t_k}|x_{t_k}|^4>\limsup_{k\to\infty}\frac{1}{t_k}\alpha^2t_k^2=\infty$. This contradicts Theorem 2 and, so, the second term tends to zero as well. In the third term, $\alpha_{t+1}:=p(a_{t+1},b_{t+1})x_{t+1}^2-E[p(a_{t+1},b_{t+1})x_{t+1}^2\mid\mathcal{F}_t]$ is a martingale difference. Therefore, $\frac{1}{N}\sum_{t=1}^{N}\alpha_{t+1}\to 0$, provided that $\sum_{t=1}^{\infty}t^{-2}E[\alpha_{t+1}^2\mid\mathcal{F}_t]<\infty$ (see Hall and Heyde (1980), Theorem 2.18). Since $p(a_{t+1},b_{t+1})$ is bounded, it is easily seen that this last condition is implied by $\sum_{t=1}^{\infty}t^{-2}[|x_t|^4+|u_t|^4]<\infty$. Again, this conclusion can be drawn by contradiction from Theorem 2. In fact, if this conclusion were false, sequence $t^{-1/2}[|x_t|^4+|u_t|^4]$ would be unbounded, and therefore there would exist a sequence of instant points t_k such that $[|x_{t_k}|^4+|u_{t_k}|^4]>t_k^{1/2}$, $\forall k$. From this, $\limsup_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}[|x_t|^4+|u_t|^4]^4\geq \limsup_{N\to\infty}\frac{1}{t_k}[|x_{t_k}|^4+|u_{t_k}|^4]^4>\limsup_{k\to\infty}\frac{1}{t_k}t_k^2=\infty$ and this is in contradiction with Theorem 2.

In conclusion, $A\to 0$ a.s..

B) Notice first that, by Schwarz inequality,

$$\begin{aligned}&\left|\frac{1}{N}\sum_{t=1}^{N}(p(a_t,b_t)-p(a_{t+1},b_{t+1}))x_{t+1}^2\right|\\&\leq\frac{1}{N}\sum_{t=1}^{N}|(p(a_t,b_t)-p(a_{t+1},b_{t+1}))|x_{t+1}^2\\&\leq\left(\frac{1}{N}\sum_{t=1}^{N}(p(a_t,b_t)-p(a_{t+1},b_{t+1}))^2\right)^{1/2}\left(\frac{1}{N}\sum_{t=1}^{N}x_{t+1}^4\right)^{1/2}.\end{aligned}$$

In this last expression, the second term remains bounded by Theorem 2, while the first term tends to zero (see Appendix B for the proof of this fact). Thus,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}(p(a_t,b_t)-p(a_{t+1},b_{t+1}))x_{t+1}^2=0\quad a.s. \tag{9}$$

Finally, conclusion

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[(p(a_t,b_t) - p(a_{t+1},b_{t+1}))x_{t+1}^2 \mid \mathcal{F}_t] \quad a.s.$$

is drawn from (9) by observing that

$$\beta_{t+1} := (p(a_t,b_t) - p(a_{t+1},b_{t+1}))x_{t+1}^2 - E[(p(a_t,b_t) - p(a_{t+1},b_{t+1}))x_{t+1}^2 \mid \mathcal{F}_t]$$

is a martingale difference for which, by calculations resembling those developed in point (A) for α_t, we have $\frac{1}{N}\sum_{t=1}^{N} \beta_{t+1} \to 0$.

C) By Schwarz inequality,

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{t=1}^{N} p(a_t,b_t) \left\{ (a_t x_t + b_t u_t)^2 - (a^\circ x_t + b^\circ u_t)^2 \right\} \right| \\
&\leq \sup_{[a\ \ b]\in\Theta} p(a,b) \left(\frac{1}{N} \sum_{t=1}^{N} ((a_t x_t + b_t u_t) - (a^\circ x_t + b^\circ u_t))^2 \right)^{1/2} \\
&\qquad \times \left(\frac{1}{N} \sum_{t=1}^{N} ((a_t x_t + b_t u_t) + (a^\circ x_t + b^\circ u_t))^2 \right)^{1/2}
\end{aligned}$$

Since $1/N\sum_{t=1}^{N}((a_t x_t + b_t u_t) + (a^\circ x_t + b^\circ u_t))^2$ remains bounded (see Theorem 2), to show that $C \to 0$ it suffices to prove that $1/N\sum_{t=1}^{N}((a_t x_t + b_t u_t) - (a^\circ x_t + b^\circ u_t))^2 \to 0$.

From Lemma 2 we have

$$\begin{aligned}
&\frac{1}{N} \sum_{t=1}^{N} ((a_t x_t + b_t u_t) - (a^\circ x_t + b^\circ u_t))^2 \\
&= o\left(\frac{1}{N} \sum_{t=1}^{N} x_t^2 \right) + \frac{1}{N} \sum_{t\in\mathcal{T}_N} ((a_t x_t + b_t u_t) - (a^\circ x_t + b^\circ u_t))^2.
\end{aligned}$$

The first term in the right hand side tends to zero because of the stability Theorem 2. As for the second term, by recalling that $|\mathcal{T}_N| \leq 2$, it is easy to prove that it tends to zero by arguments similar to those used in point (A) to show that $\frac{1}{N}p(a_{N+1},b_{N+1})x_{N+1}^2 \to 0$.

By inserting all the partial results in equation (8) the thesis is obtained. $\square$

By selecting the unfalsified set at time t as given in definition (5) with the bounds on μ_t as suggested by Lemma 1 and 3, the BOB procedure writes

Adaptive control method

At time t, do the following:

1. determine $\mathcal{U}_t$ as in definition (5) with $\mu_t = \log^s \sum_{k=1}^{t-1} x_k^2, \quad s \in (1,2)$;
2. compute $[\ a_t^{min} \quad b_t^{min}]$ as the minimizer of $\mathcal{J}^\star_{(a,b)}$ in $\mathcal{U}_t$:

$$[\ a_t^{min} \quad b_t^{min}] := \arg\min_{[a \quad b] \in \mathcal{U}_t} \mathcal{J}^\star_{(a,b)};$$

3. compute u_t by applying the optimal control law for $[\ a_t^{min} \quad b_t^{min}]$:

$$u_t = K(a_t^{min}, b_t^{min}) x_t.$$

□

The effectiveness of this adaptive control method is guaranteed by Theorem 1 due to that conditions i) and ii) follow from Lemma 1 and Lemma 3, so delivering the following optimality theorem.

THEOREM 3. *With the control law chosen according to the adaptive control method, we achieve the ideal objective, i.e.*

$$\limsup_{N\to\infty} 1/N \sum_{t=1}^{N} [q x_t^2 + u_t^2] = \mathcal{J}^\star_{(a^\circ, b^\circ)} \quad a.s., \ \forall [\ a^\circ \quad b^\circ\ ] \in \Theta.$$

Appendix A

Define

$$v_t := [\ (a^\circ - a_t) \quad (b^\circ - b_t)\ ].$$

Since $[\ a_t \quad b_t\ ] \in \Theta$, sequence v_t is bounded. Denote by $\bar{v}$ an upper bound for $\|v_t\|$: $\|v_t\| \leq \bar{v}, \forall t$.

We start by proving that (remember that $\phi_t = [\ x_t \quad u_t\ ]$):

$$\sum_{t=1}^{N-1} |\phi_t v_N^T|^r = o\left(\sum_{t=1}^{N-1} |x_t|^r\right) \quad a.s. \qquad \forall r \geq 2. \tag{10}$$

To this purpose, note first that

$$\sum_{t=1}^{N-1} x_t^2 = O\left(\sum_{t=1}^{N-1} |x_t|^r\right) \quad a.s. \tag{11}$$

Indeed,

$$\begin{aligned}
\sum_{t=1}^{N-1} x_t^2 &= N \left[\left(\frac{1}{N}\sum_{t=1}^{N-1} x_t^2\right)^{r/2}\right]^{2/r} \\
&\leq N \left[\frac{1}{N}\sum_{t=1}^{N-1} |x_t|^r\right]^{2/r} \qquad (using\ Jensen's\ inequality) \\
&= \sum_{t=1}^{N-1} |x_t|^r \left[\frac{N}{\sum_{t=1}^{N-1} |x_t|^r}\right]^{1-\frac{2}{r}},
\end{aligned}$$

and $\limsup_{N\to\infty} N/\sum_{t=1}^{N-1}|x_t|^r < \infty$ a.s. since process x_t is affected by the noise process w_t. Secondly, by observing that $[\ a_N \quad b_N\ ] \in \mathcal{U}_N$ and $[\ a^\circ \quad b^\circ\ ] \in \mathcal{U}_N$ too for N large enough (see Lemma 1),

$$\sum_{t=1}^{N-1}(\phi_t v_N^T)^2 = o\left(\log^2 \sum_{t=1}^{N-1} x_t^2\right) \quad a.s. \tag{12}$$

Equation (10) is easily derived from (11) and (12) as follows:

$$\begin{aligned}
\sum_{t=1}^{N-1}|\phi_t v_N^T|^r &\le \left|\sum_{t=1}^{N-1}(\phi_t v_N^T)^2)\right|^{r/2} \\
&= o\left(\log^r \sum_{t=1}^{N-1} x_t^2\right) && (using\ (12)) \\
&= o\left(\sum_{t=1}^{N-1} x_t^2\right) \\
&= o\left(\sum_{t=1}^{N-1} |x_t|^r\right). && (using\ (11))
\end{aligned}$$

Fix now a real number $\epsilon > 0$ and an integer N.

Define a sequence of subspaces $S_t, t = 1, 2, \ldots, N+1$ of $\mathbf{R}^2$ by the following backward recursive procedure:

for $t = N+1$, set $S_t = \emptyset$;
for $t = N, N-1, \ldots, 1$, set (the symbol $v_{t,S}$ stands for the projection of vector v_t onto subspace S)

$$S_t = \begin{cases} S_{t+1}, & \text{if } \|v_{t,S_{t+1}^\perp}\| \le \epsilon \\ S_{t+1} \oplus span\{v_t\}, & \text{otherwise.} \end{cases} \tag{13}$$

Denote by $\mathcal{T}_N$ the set of instant points at which subspace S_t expands: if $t \in \mathcal{T}_N$, then $S_t \supset S_{t+1}$ strictly. These instant pints are obviously at most two. Let denote them by t_1 and t_2 ($t_1 > t_2$). Moreover, let $i(t) := \max\{i : t_i \ge t\}$. Since $\|v_t\| \le \bar{v}, \forall t$, the angle between v_{t_1} and v_{t_2} may tend to zero only if $\epsilon \to 0$. Then, there exists a constant $c(\epsilon)$ dependent on ϵ, but independent of N, such that

$$\|\phi_{t,S_t}\|^r \le c(\epsilon)\sum_{i=1}^{i(t)}|\phi_t v_{t_i}^T|^r.$$

Thus, for each $t \in [1, N]$ we have

$$\begin{aligned}
|\phi_t v_t^T|^r &\le k|\phi_{t,S_t^\perp} v_{t,S_t^\perp}^T|^r + k|\phi_{t,S_t} v_{t,S_t}^T|^r \\
&\le k\epsilon^r\|\phi_t\|^r + k\bar{v}^r c(\epsilon)\sum_{i=1}^{i(t)}|\phi_t v_{t_i}^T|^r,
\end{aligned}$$

where k is a suitable constant depending on r.

So,

$$\begin{aligned}\sum_{t=1, t\notin \mathcal{T}_N}^{N} |\phi_t v_t^T|^r &\le k\epsilon^r \sum_{t=1}^{N} \|\phi_t\|^r + k\bar{v}^r c(\epsilon) \sum_{t=1, t\notin \mathcal{T}_N}^{N} \sum_{i=1}^{i(t)} |\phi_t v_{t_i}^T|^r \\ &\le k\epsilon^r \sum_{t=1}^{N} \|\phi_t\|^r + k\bar{v}^r c(\epsilon) \sum_{i=1}^{i(t)} \sum_{t=1}^{t_i-1} |\phi_t v_{t_i}^T|^r \\ &\le k\epsilon^r \sum_{t=1}^{N} \|\phi_t\|^r + k\bar{v}^r c(\epsilon) 2o\left(\sum_{t=1}^{N-1} |x_t|^r\right). \qquad (\mathit{using}\ (10))\end{aligned}$$

Since $\|\phi_t\|^r \le c|x_t|^r$, where c is a suitable constant, from this last inequality we conclude that

$$\limsup_{N\to\infty} \frac{\sum_{t=1, t\notin \mathcal{T}_N}^{N} |\phi_t v_t^T|^r}{\sum_{t=1}^{N} |x_t|^r} \le k\epsilon^r c.$$

Due to the arbitrariness of ϵ, this completes the proof of the lemma. □

Appendix B

Note first that

$$\mu_{t+1} - \mu_t \to 0 \quad a.s. \tag{14}$$

Indeed,

$$\begin{aligned}\mu_{t+1} - \mu_t &= \log^s \sum_{k=1}^{t} x_k^2 - \log^s \sum_{k=1}^{t-1} x_k^2 \\ &\le \log^2 \sum_{k=1}^{t} x_k^2 - \log^2 \sum_{k=1}^{t-1} x_k^2 \qquad \left(\mathit{when}\ \log^2 \sum_{k=1}^{t-1} x_k^2 > 1,\ \mathit{since}\ s < 2\right) \\ &= \log \frac{\sum_{k=1}^{t} x_k^2}{\sum_{k=1}^{t-1} x_k^2} \left(\log \sum_{k=1}^{t} x_k^2 + \log \sum_{k=1}^{t-1} x_k^2\right) \\ &\le \frac{x_t^2}{\sum_{k=1}^{t-1} x_k^2} 2 \log \sum_{k=1}^{t} x_k^2. \qquad (\mathit{using\ relation}\ \log(1+x) \le x) \qquad (15)\end{aligned}$$

In this last expression, $\sum_{k=1}^{t} x_k^2$ grows linearly (in fact, $\sum_{k=1}^{t} x_k^2$ does not grow less than linearly because of the presence of noise w_t affecting the system equation (3) and does not grow faster than linearly because of the stability Theorem 2). Moreover, by similar arguments as those used in point (A) of the proof of Lemma 3, $x_t^2 = o(t^{1/2})$. Using these estimates in (15) we obtain

$$\mu_{t+1} - \mu_t \le o\left(\frac{\log t}{t^{1/2}}\right) \quad a.s.,$$

which implies (14).

Consider now definition (5) of set $\mathcal{U}_t$. In the light of equations (14) and also considering that kernel V_t is increasing and that $\|[\, a_t^{LS} \;\; b_t^{LS}\,] - [\, a_{t+1}^{LS} \;\; b_{t+1}^{LS}\,]\| \to 0$ *a.s.*, we can conclude that any parameter $[a\ b] \in \mathcal{U}_{t+1} - \mathcal{U}_t$ has a distance from $\mathcal{U}_t$ that tends to zero as $t \to \infty$, namely $\sup_{[ab] \in \mathcal{U}_{t+1} - \mathcal{U}_t} \inf_{[a'b'] \in \mathcal{U}_t} \|[a\ b] - [a'\ b']\| \to 0,\ t \to \infty$. Since $\mathcal{J}^\star_{(\cdot,\cdot)} = p(\cdot,\cdot)$ is a continuous function in Θ we then have that there exist a vanishing function ϵ_t ($\epsilon_t \to 0$) such that

$$p(a_t, b_t) - p(a_{t+1}, b_{t+1}) = \mathcal{J}^\star_{(a_t,b_t)} - \mathcal{J}^\star_{(a_{t+1},b_{t+1})} \leq \epsilon_t.$$

Finally, letting N^+ denote the set of instant points $t \in [1, N]$ such that $p(a_t, b_t) - p(a_{t+1}, b_{t+1}) \geq 0$,

$$\begin{aligned}
&\sum_{t=1}^{N} |p(a_t, b_t) - p(a_{t+1}, b_{t+1})| \\
&\leq \sup_{[a\ b] \in \Theta} p(a, b) + 2 \sum_{t \in N^+} (p(a_t, b_t) - p(a_{t+1}, b_{t+1})) \\
&\leq \sup_{[a\ b] \in \Theta} p(a, b) + 2 \sum_{t \in N^+} \epsilon_t \\
&= o(N) \quad a.s.,
\end{aligned}$$

so concluding the proof. □

REFERENCES

[1] K. Aström and B. Wittenmark, 1973, *On self-tuning regulators*, Automatica, 9, pp. 185-199.

[2] Y. Bar-Shalom and E. Tse, 1974, *Dual effort, certainty equivalence and separation in stochastic control*, IEEE Trans. on Automatic Control, AC-19, pp. 494-500.

[3] Y. Bar-Shalom and K. Wall, 1974, *Dual adaptive control and uncertainty effects in macroeconomic modelling*, Automatica, 16, pp. 147-156.

[4] A. Becker, P. R. Kumar and C. Z. Wei, 1985, *Adaptive control with the stochastic approximation algorithm: Geometry and convergence*, IEEE Trans. on Automatic Control, AC-30, pp. 330-338.

[5] D. Bertsekas, 1987, *Dynamic Programming*, Prentice-Hall, NJ.

[6] S. Bittanti, P. Bolzern and M. C. Campi, 1990, *Recursive least squares identification algorithms with incomplete excitation: convergence analysis and application to adaptive control*, IEEE Trans. on Automatic Control, AC-35, pp. 1371-1373.

[7] S. Bittanti and M. C. Campi, 1996, *Least squares based self-tuning control systems*, In: Identification, Adaptation, Learning. The science of learning models from data (S.Bittanti and G.Picci eds.). Springer-Verlag NATO ASI series - Computer and systems sciences, 508-514.

[8] V. S. Borkar, 1993, *On the Milito Cruz adaptive control scheme for Markov chains*, J. of Optimization Theory and Applications, 77, pp. 387-397.

[9] P. E. Caines, 1992, *Continuous-time stochastic adaptive control: non-explosion, ϵ-consistency and stability*, Syst. Contr. Lett., 19, pp. 169-176.

[10] P. E. Caines and S. Lafortune, 1984, *Adaptive control with recursive identification for stochastic linear systems*, IEEE Trans. on Automatic Control, AC-29, pp. 312-321.

[11] M. C. Campi, 1991, *On the convergence of minimum-variance directional-forgetting adaptive control schemes*, Automatica, 28, pp. 221-225.

[12] M. C. Campi, 1996, *The problem of pole-zero cancellation in transfer function identification and application to adaptive stabilization*, Automatica, 32, pp. 849-857.

[13] M. C. Campi, 1997, *Achieving optimality in adaptive control: the Bet On the Best approach*, In: Proc. 36th Conf. on Decision and Control.

[14] M. C. Campi and P. R. Kumar, 1998, *Adaptive linear quadratic Gaussian control: the cost-biased approach revisited*, SIAM Journal on Control and Optimization, 36, pp. 1890-1907.

[15] H. F. Chen and L. Guo, 1986, *Convergence rate of least squares identification and adaptive control for stochastic systems*, International Journal of Control, 44, pp. 1459-1476.

[16] H. F. Chen and L. Guo, 1987a, *Asymptotically optimal adaptive control with consistent parameter estimates*, SIAM Journal on Control and Optimization, 25, pp. 558-575.

[17] H. F. Chen and L. Guo, 1987b, *Optimal adaptive control and consistent parameter estimates for ARMAX model with quadratic cost*, SIAM Journal on Control and Optimization, 25, pp. 845-867.

[18] H. F. Chen and L. Guo, 1988, *A robust stochastic adaptive controller*, IEEE Trans. on Automatic Control, AC-33, pp. 1035-1043.

[19] H. F. Chen and L. Guo, 1991, *Identification and stochastic adaptive control*, Birkhauser.

[20] T. E. Duncan, L. Guo and B. Pasik-Duncan, 1999, *Adaptive continuous-time linear quadratic Gaussian control*, IEEE Trans. on Automatic Control, AC-44, pp. 1653-1662.

[21] T. E. Duncan and B. Pasik-Duncan, 1986, *A parameter estimate associated with the adaptive control of stochastic systems*, In: Analysis and optimization of systems, Control and information sc., Springer., 83, pp. 508-514.

[22] T. E. Duncan and B. Pasik-Duncan, 1991, *Some methods for the adaptive control of continuous time linear stochastic systems*, In: Topics in stochastic systems: modelling, estimation and adaptive control, Control and information sc., Springer, 161, pp. 242-268.

[23] G. C. Goodwin, P. J. Ramadge and P. E. Caines, 1981, *Discrete-time stochastic adaptive control*, SIAM Journal on Control and Optimization, 19, pp. 829-853.

[24] L. Guo, 1996, *Self-convergence of weghted least-squares with applications to stochastic adaptive control*, IEEE Trans. on Automatic Control, AC-41, pp. 79-89.

[25] L. Guo and H. F. Chen, 1991, *The Astrom-Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers*, IEEE Trans. on Automatic Control, AC-36, pp. 802-812.

[26] P. Hall and C. C. Heyde, 1980, *Martingale limit theory and its applications*, Academic Press, N.Y.

[27] P. R. Kumar, 1983a, *Simultaneous identification and adaptive control of unknown systems over finite parameter sets*, IEEE Trans. on Automatic Control, AC-28, pp. 68-76.

[28] P. R. Kumar, 1983b, *Optimal adaptive control of linear-quadratic-Gaussian systems*, SIAM Journal on Control and Optimization, 21, pp. 163-178.

[29] P. R. Kumar and A. Becker, 1982, *A new family of optimal adaptive controllers for Markov chains*, IEEE Trans. on Automatic Control, AC-27, pp. 137-146.

[30] T. L. Lai and C. Z. Wei, 1982, *Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems*, Annals of Statistics, 10, pp. 154-166.

[31] D. Levanony and P. E. Caines, 2001a, *Stochastic Lagrangian adaptation*, Internal Report,

Mc Gill University.

[32] D. Levanony and P. E. Caines, 2001b, *On persistent excitation for linear systems with stochastic coefficients*, SIAM Journal on Control and Optimization, 40, pp. 882-897.

[33] D. Levanony and P. E. Caines, 2005, *Stochastic linear-quadratic adaptive control: a conceptual scheme*, In: Proc. 44th Conf. on Decision and Control, pp. 1747-1752.

[34] W. Lin, W, P. R. Kumar and T. I. Seidman, 1985, *Will the self-tuning approach work for general cost criteria?* Syst. Contr. Lett., 6, pp. 77-85.

[35] R. A. Milito and J. B. Cruz, 1987, *An optimization-oriented approach to the adaptive control of Markov chains*, IEEE Trans. on Automatic Control, AC-32, pp. 754-762.

[36] J. W. Polderman, 1986a, *On the necessity of identifying the true parameter in adaptive LQ control*, Syst. Contr. Lett., 8, pp. 87-91.

[37] J. W. Polderman, 1986b, *A note on the structure of two subsets of the parameter space in adaptive control problems*, Syst. Contr. Lett., 8, pp. 25-34.

[38] M. Prandini and M. C. Campi, 2001, *Adaptive LQG control of input-output systems: a cost-biased approach*, SIAM Journal on Control and Optimization, 39, pp. 1499-1519.

[39] K. S. Sin and G. C. Goodwin, 1982, *Stochastic adaptive control using a modified least squares algorithm*, Automatica, 18, pp. 815-321.

[40] C. T. Striebel, 1965, *Sufficient statistics in the optimal control of stochastic systems*, J. Math. Anal. Appl., 12, pp. 576-592.

[41] J. H. van Schuppen, 1994, *Tuning of Gaussian stochastic control systems*, IEEE Trans. on Automatic Control, AC-39, pp. 2178-2190.