SIAM J. CONTROL OPTIM.
Vol. 36, No. 6, pp. 1890–1907, November 1998

003

ADAPTIVE LINEAR QUADRATIC GAUSSIAN CONTROL: THE COST-BIASED APPROACH REVISITED*

M. C. CAMPI† AND P. R. KUMAR‡

Abstract. In adaptive control, a standard approach is to resort to the so-called certainty equivalence principle which consists of generating some standard parameter estimate and then using it in the control law as if it were the true parameter. As a consequence of this philosophy, the estimation problem is decoupled from the control problem and this substantially simplifies the corresponding adaptive control scheme. On the other hand, the complete absence of dual properties makes certainty equivalent controllers run into an identifiability problem which generally leads to a strictly suboptimal performance.

In this paper, we introduce a cost-biased parameter estimator to overcome this difficulty. This estimator is applied to a linear quadratic Gaussian controller. The corresponding adaptive scheme is proven to be stable and optimal when the unknown system parameter lies in an infinite, yet compact, parameter set.

Key words. adaptive control, linear quadratic Gaussian control, self-optimizing control, cost-biased approach, certainty equivalence

AMS subject classifications. 15A15, 15A09, 15A23

PII. S0363012997317499

1. Introduction.

Consider a linear time-invariant system

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

where $x_t \in \mathbf{R}^n$ is the state, $u_t \in \mathbf{R}^m$ the control variable, and w_t is a noise process of independent, Normal $N(0,1)$ random variables. The system matrices A° and B° are unknown.

Our control objective is to select the input u_t in such a way as to minimize the long-term average quadratic cost criterion

$$\limsup_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \left[x_s^T Q x_s + u_s^T R u_s\right], \qquad Q = Q^T \geq 0, \quad R = R^T > 0. \tag{2}$$

To this purpose, we observe the state x_t and, based on this, we first generate an estimate of the system matrices A° and B° and then exploit these estimates in a certainty equivalence fashion.

*Received by the editors February 26, 1997; accepted for publication (in revised form) January 8, 1998; published electronically July 17, 1998.

http://www.siam.org/journals/sicon/36-6/31749.html

†Department of Electrical Engineering and Automation, University of Brescia, Via Branze 38, 25123 Bresica, Italy. The research of this author was supported by MURST under the 60% project "Adaptive identification, prediction and control" and the 40% project "Model identification, system control and signal processing." This research was conducted while M. C. Campi was visiting the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign in the summer of 1995.

‡Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, 1308 West Main Street, Urbana, IL 61801 (prkumar@decision.csl.uiuc.edu). The research of this author was supported by U.S. Army Research Office contract DAAH-04-95-1-0090 and Joint Service Electronics Program contract N00014-96-1-0129.

A common way to generate an estimate of A° and B° is to resort to the least squares method which corresponds to minimizing the performance index

$$V_t(A, B) = \sum_{s=1}^{t} \|x_s - Ax_{s-1} - Bu_{s-1}\|^2. \tag{3}$$

It is well known, however, that the corresponding certainty equivalent adaptive control law can suffer from an identifiability problem and that this can result in a degradation of the control system performance; see [1, 2, 3, 4]. In particular, for the case where matrices A° and B° belong to a finite known set, it is shown in [2] that the least squares estimate can converge with positive probability to a false estimate, which then leads to a strictly suboptimal value of the long-term average cost criterion. For the case of controlled Markov chains, such a counterexample had earlier been exhibited in [1]. Parameter consistency is guaranteed under certain conditions which are satisfied only in specific adaptive control situations, as, e.g., studied in [5] and [6].

This inability to identify the open loop system from closed-loop measurements is one of the fundamental obstacles to self-optimizing adaptive control. To overcome this, one approach is to occasionally probe the system. This can be done by either adding dither to the control or by occasionally breaking the control loop. However, such perturbations should be of small enough magnitude or infrequent enough so that they do not in themselves add to the cost incurred. An account of this approach can be found in Chen and Guo [7, 8, 9, 10, 11, 12].

To overcome this general problem of identifiability in closed loop, a very different approach, which still preserves the certainty equivalent structure of the adaptive controller and holds out the promise of general self-optimizing controllers, was proposed in [13] for the class of controlled Markov chains. The novelty of this adaptive controller is the employment of a *cost-biased maximum likelihood* parameter estimator, rather than the usual maximum likelihood parameter estimator. This cost biasing modifies the log-likelihood criterion by incorporating an additional term which favors parameter estimates with smaller optimal costs. For controlled Markov chains with a finite parameter set, it was shown in [13] that such a cost biasing eliminates parameters with costs larger than the optimal cost from occurring as limit points of the estimator. As a consequence, the corresponding adaptive controller was proved to provide optimal performance. This result was extended in [14] to the case of general parameter sets, for controlled Markov chains with finite state spaces. Another extension to the case of a finite parameter set, but allowing for a general state space and nonlinear systems, was provided in [15]. In the reference most pertinent to this paper, [2], it was shown that the cost-biased maximum likelihood-based certainty equivalent controller yielded an optimal cost for linear systems with quadratic costs, as in (1) and (2), provided that the parameter set is finite.

The assumption that the parameter set is finite is crucial in the derivations of [2]. Indeed, it was shown in [2] that the log-likelihood ratio $V_t(A^\circ, B^\circ) - V_t(A, B)$ stays bounded for any *fixed* parameter (A, B), and, therefore, a wrong fixed parameter (A, B) can gain, at most, a finite advantage over the true parameter (A°, B°) in the standard least squares criterion. Thus, when the number of possible parameters is finite, the maximum of these finite advantages is still finite, and so a mild biasing is sufficient to prevent elements (A, B) with larger cost than the optimal cost from occurring as limit points of the parameter estimator. This mildness of the biasing is important in order not to destroy the ability of the least squares estimate to identify closed-loop dynamics. Unfortunately, this argument is no longer true when turning to a more general setting allowing for infinitely possible true parameterizations. Indeed, in such a case, $inf_{(A,B)}[V_t(A^\circ, B^\circ) - V_t(A, B)]$ is no longer bounded and the above argument valid for the finite case fails to apply. As a consequence of this and other difficulties, the infinite parameter set case has remained so far unsolved.

It is the purpose of this paper to establish the optimality of a certainty equivalent controller based on the cost-biased maximum likelihood parameter estimator for linear quadratic Gaussian systems, in the case of compact parameter uncertainty set. The aforementioned difficulty that the log-likelihood ratio $V_t(A^\circ, B^\circ) - V_t(A, B)$ is unbounded is circumvented by resorting to a Bayesian embedding approach. In this setting, one can show that the least squares estimate converges along the directions of diverging information to the true parameter value. As a consequence, a sequence of parameters (A'_t, B'_t) can be determined with the property that it converges to the true parameter (A°, B°) and for which $inf_{(A,B)}[V_t(A'_t, B'_t) - V_t(A, B)]$ remains bounded. Loosely speaking, (A'_t, B'_t) can be used in the analysis in place of (A°, B°) and, by a careful use of continuity arguments, the optimality of the adaptive controller can be established.

The paper is organized as follows. Our adaptive control scheme is described in section 2. In section 3, the properties of the cost-biased maximum likelihood parameter estimator are worked out. Section 4 is devoted to the study of the self-tuning properties of the adaptive scheme, and its stability and optimality are established in section 5.

2. The adaptive control system. Throughout this paper, let $[A, B] \in \mathbf{R}^{n\times(n+m)}$ denote the matrix obtained by concatenating matrices $A \in \mathbf{R}^{n\times n}$ and $B \in \mathbf{R}^{n\times m}$.

In our adaptive control problem, matrices A° and B° of system (1) are unknown and belong to a known compact set Θ as precisely stated in the following assumptions.

(A.i) There is a known compact set $\Theta \subset R^{n\times(n+m)}$ such that

$$[A^\circ, B^\circ] \in \text{interior}(\Theta).$$

(A.ii) (A, B) is reachable and $(A, Q^{1/2})$ is observable, $\forall [A, B] \in \Theta$.

Given the system parameters $[A, B] \in \Theta$, the control law minimizing the cost (2) for the system $x_{t+1} = Ax_t + Bu_t + w_{t+1}$ is easily derived (see, e.g., Kumar and Varaiya [16] or Bertsekas [17] for a comprehensive presentation of linear quadratic control problems). First, one has to compute the positive semidefinite solution to the algebraic Riccati equation

$$P = A^T PA - A^T PB(B^T PB + R)^{-1}B^T PA + Q.$$

The existence and uniqueness of such a solution is a consequence of the reachability and observability assumption (A.ii). Denoting such a solution by $P(A, B)$, the control law is then given by

$$u_t = K(A, B)x_t, \tag{4}$$

where $K(A, B)$ is the linear quadratic Gaussian (LQG) optimal gain defined by

$$K(A, B) = -(B^T P(A, B)B + R)^{-1}B^T P(A, B)A. \tag{5}$$

The corresponding optimal cost is denoted by $J(A, B)$.

When one is facing an adaptive control problem, the system matrices (A°, B°) are not known and some estimates $\widehat{A}_t$ and $\widehat{B}_t$ of them are needed. Once these estimates have been generated, in the certainty equivalence approach they are simply used as if they were the true system matrices. Correspondingly, the adaptive control law is given by

$$u_t = K(\widehat{A}_t, \widehat{B}_t)x_t. \tag{6}$$

The heart of our adaptive control scheme lies in the cost-biased least squares estimator of the system matrices as described below.

Choose a deterministic sequence μ_t such that $\mu_t \to \infty$ and $\mu_t = o(\log t)$ as $t \to \infty$. The parameter estimate sequence $\{[\widehat{A}_t, \widehat{B}_t]\}$ is given by

$$[\widehat{A}_t, \widehat{B}_t] = \begin{cases} \arg\min_{[A,B]\in\Theta} \left\{ \sum_{s=1}^{t} \|x_s - Ax_{s-1} - Bu_{s-1}\|^2 + \mu_t J(A,B) \right\}, \text{ for } t \text{ even}, \\ \left[\widehat{A}_{t-1}, \widehat{B}_{t-1}\right], \text{ for } t \text{ odd} \end{cases} \tag{7}$$

(when there is more than one minimizer, any of them can be chosen).

The distinguishing feature of the criterion (7) is the term $\mu_t J(A,B)$, which introduces a mild bias in favor of parameters (A,B) with lower optimal costs. The biasing is "mild" because $\mu_t = o(\log t)$. On the other hand, it is nonnegligible because $\mu_t \to \infty$. Without this term one would simply have the usual least squares parameter estimator, with its attendant difficulty in identifying the system in closed loop.

The intuitive rationale for the cost biasing in the least squares criterion is as follows. Suppose that one simply employs a straightforward least squares parameter estimator. Then, generically, it can be shown that the least squares parameter estimates sequence $[\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]$ converges to a limiting random variable $[\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}]$ (see [18]). Such a limiting estimate results in a limiting controller $u_t = K(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS})x_t$. It is natural to expect that the least squares estimator will asymptotically identify, at a minimum, the closed-loop behavior of the system. Thus, one expects that the behavior of the true system with the loop closed by $u_t = K(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS})x_t$ will be the same as the closed-loop estimated system, i.e., their closed-loop gains are equal:

$$A^\circ + B^\circ K(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}) = \widehat{A}_\infty^{LS} + \widehat{B}_\infty^{LS} K(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}).$$

This implies that the cost of running the true system (A°, B°) with the feedback gain $K(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS})$ is the same as the cost of running the estimated system $(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS})$ with the feedback $K(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS})$. The latter is, however, the optimal configuration for the system $x_{t+1} = \widehat{A}_\infty^{LS} x_t + \widehat{B}_\infty^{LS} u_t + w_{t+1}$, while the former is not necessarily an optimal configuration for the true system. Thus one has

$$J(\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}) \geq J(A^\circ, B^\circ).$$

This means that the least squares estimator has a natural tendency to return estimates with larger optimal cost than the optimal cost associated with the true system. This motivates the idea of somehow introducing a bias into the parameter estimator so that it favors parameters (A,B) with smaller values of $J(A,B)$.

Thus, one conceives of adding a term such as $\mu_t J(A,B)$ to the squared error in (7). However, one needs to choose μ_t with care. One does not want to destroy the ability of the least squares estimator to identify the closed-loop dynamics. This is achieved by choosing μ_t small enough so that $\mu_t = o(logt)$. On the other hand, one definitely wants the $\mu_t J(A,B)$ term to assert itself, and this is achieved by choosing $\mu_t \to \infty$. Hence, we arrive at the cost-biased least squares parameter estimator (7).

Notation. For brevity, the following notation will be used throughout the paper: $P^\circ := P(A^\circ, B^\circ)$, $\widehat{P}_t := P(\widehat{A}_t, \widehat{B}_t)$, $K^\circ := K(A^\circ, B^\circ)$, $\widehat{K}_t := K(\widehat{A}_t, \widehat{B}_t)$, $J^\circ := J(A^\circ, B^\circ)$, and $\widehat{J}_t := J(\widehat{A}_t, \widehat{B}_t)$. □

3. The properties of the parameter estimates. In this section, we study the properties of the estimates $[\widehat{A}_t, \widehat{B}_t]$ returned by the estimator (7). Our main result is that the introduction of the cost-bias term $\mu_t J(A,B)$ in the identification criterion prevents parameters $[A,B]$ with cost $J(A,B)$ strictly larger than the optimal cost from occurring as limit points of $[\widehat{A}_t, \widehat{B}_t]$ (Theorem 2). In this way, our modification is proven successful in counteracting the natural tendency of least squares to return estimates with larger cost than the optimal one. In addition, we show that the estimator preserves the capability of the least squares method of identifying the control system closed-loop dynamics (Theorem 3).

We start by summarizing some known results on the least squares estimates relevant to the forthcoming developments.

Denote by $[\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]$ the least squares estimate of $[A^\circ, B^\circ]$:

$$\left[\widehat{A}_t^{LS}, \widehat{B}_t^{LS}\right] := \arg\min_{[A,B]\in\mathbf{R}^{n\times(n+m)}} \sum_{s=1}^{t} \|x_s - Ax_{s-1} - Bu_{s-1}\|^2 .$$

The partial ability of the least squares estimates $(\widehat{A}_t^{LS}, \widehat{B}_t^{LS})$ to estimate a portion of the open-loop system can be stated precisely using the notion of the *excited subspace*, originally introduced in [19].

DEFINITION 1. *Defining $v_s^T := [x_s^T \quad u_s^T]$, the subspace*

$$\mathcal{E}^\perp := \left\{ z \in \mathbf{R}^{n+m} : \qquad z^T \sum_{s=1}^{\infty} v_s v_s^T z < \infty \right\}$$

is called the unexcited *subspace. Its orthogonal complement $\mathcal{E}$ is the* excited *subspace.*

Given $[A,B]$, let $[A,B]_{\mathcal{E}}$ and $[A,B]_{\mathcal{E}^\perp}$ denote the matrices in $\mathbf{R}^{n\times(n+m)}$ formed by projecting the rows of $[A,B]$ onto $\mathcal{E}$ and $\mathcal{E}^\perp$, respectively.

The main properties of the least squares estimate are stated in Theorem 1 below (the proof of point (i) can be derived as a slight modification to that of Theorem 1 in [18], whereas point (ii) follows from Theorem 2 in [20] and Theorem 2 in [21]).

THEOREM 1. *There exists a set $N \in \mathbf{R}^{n+m}$ with zero Lebesgue measure such that, if $[A^\circ, B^\circ]$ does not belong to N, then*

(i)

$$\lim_{t\to\infty} [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] = [\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}] \qquad a.s.,$$

where $[\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}]$ is an almost surely (a.s.) bounded random variable.

(ii)

$$[\widehat{A}_\infty^{LS}, \widehat{B}_\infty^{LS}]_{\mathcal{E}} = [A^\circ, B^\circ]_{\mathcal{E}} \qquad a.s.$$

In particular, point (ii) asserts that the asymptotic estimation error is confined to the unexcited subspace. This is not surprising since the uncertainty in the excited directions is overcome by the information, which diverges with time. This turns out to be a crucial property in the derivation of several results concerning our adaptive scheme.

Throughout, we assume that $[A^\circ, B^\circ]$ does not belong to N.

Our first result on the cost-biased estimate $[\widehat{A}_t, \widehat{B}_t]$ proves that it abandons the region with costs larger than the optimal cost, for t large enough. A key role is played by the composite estimate

$$[A'_t, B'_t] := [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]_{\mathcal{E}} + [A^\circ, B^\circ]_{\mathcal{E}^\perp}.$$

THEOREM 2.

$$\limsup_{t\to\infty} \hat{J}_t \le J^\circ \qquad a.s.$$

Proof. Define

$$\begin{aligned} V_t(A,B) &:= \sum_{s=1}^{t} \|x_s - Ax_{s-1} - Bu_{s-1}\|^2, \\ D_t(A,B) &:= V_t(A,B) + \mu_t J(A,B). \end{aligned}$$

Note for future use that

$$V_t(A,B) - V_t(\widehat{A}_t^{LS}, \widehat{B}_t^{LS}) = \sum_{s=1}^{t} \left\| \left\{ [A,B] - [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] \right\} v_{s-1} \right\|^2.$$

Indeed, recalling that the minimizer of $V_t(A,B)$ is given by $[\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] = (\sum_{s=1}^{t} x_s v_{s-1}^T)(\sum_{s=1}^{t} v_{s-1} v_{s-1}^T)^{-1}$, one has

$$\begin{aligned} & V_t(A,B) - V_t(\widehat{A}_t^{LS}, \widehat{B}_t^{LS}) - \sum_{s=1}^{t} \left\| \left\{ [A,B] - [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] \right\} v_{s-1} \right\|^2 \\ &= \sum_{s=1}^{t} \|x_s\|^2 + \sum_{s=1}^{t} v_{s-1}^T [A,B]^T [A,B] v_{s-1} - 2\sum_{s=1}^{t} v_{s-1}^T [A,B]^T x_s \\ &- \sum_{s=1}^{t} \|x_s\|^2 - \sum_{s=1}^{t} v_{s-1}^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] v_{s-1} + 2\sum_{s=1}^{t} v_{s-1}^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T x_s \\ &- \sum_{s=1}^{t} v_{s-1}^T [A,B]^T [A,B] v_{s-1} - \sum_{s=1}^{t} v_{s-1}^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] v_{s-1} \\ &\qquad + 2\sum_{s=1}^{t} v_{s-1}^T [A,B]^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] v_{s-1} \\ &= -2\sum_{s=1}^{t} v_{s-1}^T [A,B]^T x_s - 2Trace \left\{ [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] \left(\sum_{s=1}^{t} v_{s-1}^T v_{s-1} \right) [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T \right\} \\ &+ 2\sum_{s=1}^{t} v_{s-1}^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T x_s + 2Trace \left\{ [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] \left(\sum_{s=1}^{t} v_{s-1}^T v_{s-1} \right) [A,B]^T \right\} \end{aligned}$$

$$= -2\sum_{s=1}^{t} v_{s-1}^T [A,B]^T x_s - 2Trace\left\{\sum_{s=1}^{t} x_s v_{s-1}^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T\right\}$$
$$+ 2\sum_{s=1}^{t} v_{s-1}^T [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]^T x_s + 2Trace\left\{\sum_{s=1}^{t} x_s v_{s-1}^T [A,B]^T\right\}$$
$$= 0.$$

For every $[A, B] \in S_\epsilon := \{[A, B] \in \Theta \ : \ J(A, B) \geq J^\circ + \epsilon\}, \epsilon > 0$, the following chain of inequalities holds true:

$$\begin{aligned} D_t(A,B) - D_t(A'_t, B'_t) &\geq V_t(\widehat{A}_t^{LS}, \widehat{B}_t^{LS}) + \mu_t J(A,B) \\ &\quad - V_t(A'_t, B'_t) - \mu_t J(A'_t, B'_t) \\ &\geq -\sum_{s=1}^{t} \left\| \left\{ [A'_t, B'_t] - [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] \right\} v_{s-1} \right\|^2 \\ &\quad + \mu_t \left\{ J^\circ + \epsilon - J(A'_t, B'_t) \right\}. \end{aligned} \tag{8}$$

Recalling that $J(A, B) = TraceP(A, B)$ (see [16] or [17]), and that $P(\cdot, \cdot)$ is a continuous function of the entries of matrices A and B for any $[A, B] \in \Theta$ (see [22]), we can conclude that $J(\cdot, \cdot)$ is continuous in $[A^\circ, B^\circ]$. Since $[A'_t, B'_t] \to [A^\circ, B^\circ]$ (which follows from (ii) of Theorem 1), we therefore have

$$J^\circ + \epsilon - J(A'_t, B'_t) \to \epsilon \qquad \text{a.s.}$$

Thus, the second term on the right-hand side of (8) tends to infinity as $t \to \infty$. On the other hand, by the very definition of unexcited subspace and $[A'_t, B'_t]$, the first term stays bounded. Therefore, the right-hand side of (8) is diverging, uniformly in $[A, B] \in S_\epsilon$. That is, $D_t(A, B)$ is strictly larger than $D_t(A'_t, B'_t)$ for any $[A, B] \in S_\epsilon$ when t is large enough. Finally, by noting that $[A'_t, B'_t] \in \Theta$ for t large enough, the conclusion is drawn that $[\widehat{A}_t, \widehat{B}_t]$ leaves set S_ϵ in finite time. In view of the arbitrariness of $\epsilon > 0$, the proof is complete. □

We now introduce C_δ as the set of parameters $[A, B]$ such that the gain of the corresponding optimal closed-loop system differs from the gain of the true system with the loop closed by $K(A, B)$ by at least δ in norm, i.e.,

$$C_\delta := \{[A, B] \in \Theta \ : \ \|[A^\circ + B^\circ K(A, B)] - [A + BK(A, B)]\| \geq \delta\}.$$

We now prove that the estimate $[\widehat{A}_t, \widehat{B}_t]$ can visit C_δ only rarely, and so our cost-biased estimator (7) still possesses good closed-loop identification properties.

THEOREM 3.

$$\sum_{s=1}^{t} 1([\widehat{A}_s, \widehat{B}_s] \in C_\delta) = O(\mu_t) \qquad a.s., \qquad \forall \delta > 0.$$

Proof. We first prove that

$$\sum_{s=1}^{t} \left\| \left\{ [A'_t, B'_t] - [\widehat{A}_t, \widehat{B}_t] \right\} v_{s-1} \right\|^2 = O(\mu_t), \qquad t \text{ even} \qquad \text{a.s.} \tag{9}$$

Indeed,

$$\sum_{s=1}^{t} \left\| \left\{ [A'_t, B'_t] - [\widehat{A}_t, \widehat{B}_t] \right\} v_{s-1} \right\|^2$$

$$\leq 2 \sum_{s=1}^{t} \left\| \left\{ [A^\circ, B^\circ]_{\mathcal{E}^\perp} - [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]_{\mathcal{E}^\perp} \right\} v_{s-1} \right\|^2 + 2 \sum_{s=1}^{t} \left\| \left\{ [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] - [\widehat{A}_t, \widehat{B}_t] \right\} v_{s-1} \right\|^2 .$$

The first term is bounded because of the definition of unexcited subspace. As for the second term, it can be handled as follows:

$$\begin{aligned} \sum_{s=1}^{t} \left\| \left\{ [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}] - [\widehat{A}_t, \widehat{B}_t] \right\} v_{s-1} \right\|^2 &= V_t(\widehat{A}_t, \widehat{B}_t) - V_t(\widehat{A}_t^{LS}, \widehat{B}_t^{LS}) \\ &= D_t(\widehat{A}_t, \widehat{B}_t) - D_t(A'_t, B'_t) + \mu_t \left\{ J(A'_t, B'_t) - \widehat{J}_t \right\} \\ &\quad + \left\{ V_t(A'_t, B'_t) - V_t(\widehat{A}_t^{LS}, \widehat{B}_t^{LS}) \right\} . \end{aligned}$$

The last term equals $\sum_{s=1}^{t} ||\{[A'_t, B'_t] - [\widehat{A}_t^{LS}, \widehat{B}_t^{LS}]\} v_{s-1}||^2$ and is bounded, whereas, by noting that $[A'_t, B'_t] \in \Theta$ for t large enough, the first term is less than or equal to zero in the limit. Result (9) then follows from the fact that $J(A'_t, B'_t) - \widehat{J}_t$ is bounded (remember that $J(\cdot, \cdot)$ is a continuous function on Θ and Θ is a compact set).

Note now that the matrix

$$[A^\circ + B^\circ K(A, B)] - [\bar{A} + \bar{B} K(A, B)]$$

is continuous as a function of $[A, B] \in \Theta$ and $[\overline{A}, \overline{B}] \in \Theta$ (this follows from the expression (5) of the gain $K(A, B)$ and the continuity of $P(A, B)$ in Θ (see [22])). Therefore, $\forall [\widetilde{A}, \widetilde{B}] \in C_\delta$, there exists a neighborhood $N(\widetilde{A}, \widetilde{B})$ of $[\widetilde{A}, \widetilde{B}]$ and a nonzero matrix H such that

$$\begin{aligned} &\left([A^\circ + B^\circ K(A, B)] - [\overline{A} + \overline{B} K(A, B)] \right)^T \left([A^\circ + B^\circ K(A, B)] - [\overline{A} + \overline{B} K(A, B)] \right) \\ &\geq H^T H, \qquad \forall [A, B], [\overline{A}, \overline{B}] \in N(\widetilde{A}, \widetilde{B}). \end{aligned} \tag{10}$$

The set of all these neighborhoods constitutes a cover of C_δ, from which a finite subcover $\{N_j\}_{j=1}^q$ can be extracted. The thesis of the theorem can then be recast as

$$\sum_{s=1}^{t} 1([\widehat{A}_s, \widehat{B}_s] \in N_j) = O(\mu_t) \qquad \text{a.s.}, \qquad \forall j \in [1, q]. \tag{11}$$

Equation (11) will be proven by contradiction. To this purpose, set

$$\#_{j,t} := \sum_{\substack{s=1 \\ s \ even}}^{t} 1([\widehat{A}_s, \widehat{B}_s] \in N_j)$$

and assume that there exist $\overline{j} \in [1, q]$ and a sequence of even time points $\{t_k\}$ such that $[\widehat{A}_{t_k}, \widehat{B}_{t_k}] \in N_{\overline{j}} \forall k$, and

$$\lim_{k\to\infty} \frac{1}{\mu_{t_k}} \#_{\overline{j},t_k} = \infty. \tag{12}$$

We prove that (12) implies

$$\liminf_{k\to\infty} \frac{1}{\#_{\overline{j},t_k}} \sum_{\substack{s=1 \\ s\ even}}^{t_k} \alpha_{\overline{j},s+1} > 0, \tag{13}$$

where

$$\alpha_{\overline{j},s+1} := (\|Hx_{s+1}\|^2 \wedge 1) 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}}) \tag{14}$$

(H is the matrix introduced in (10) associated with $N_{\overline{j}}$) and, in turn, this contradicts (9).

For the proof of (13), define $\mathcal{F}_s := \sigma(w_1, \dots, w_s)$ and note first that

$$\begin{aligned} E[\|Hx_{s+1}\|^2 \wedge 1 \mid \mathcal{F}_s] &= E[\|H(A^\circ x_s + B^\circ u_s) + Hw_{s+1}\|^2 \wedge 1 \mid \mathcal{F}_s] \\ &\geq \text{Prob}(\|H(A^\circ x_s + B^\circ u_s) + Hw_{s+1}\| \geq 1 \mid \mathcal{F}_s) \\ &\geq 1 - \text{Prob}\,(\|H(A^\circ x_s + B^\circ u_s)\| - 1 < \|Hw_{s+1}\| \\ &\qquad\qquad < \|H(A^\circ x_s + B^\circ u_s\| + 1 \mid \mathcal{F}_s)) \\ &\geq 1 - \sup_\alpha \text{Prob}(\alpha - 1 < \|Hw_{s+1}\| < \alpha + 1) \\ &\geq c, \end{aligned}$$

for a suitable constant $c > 0$, the last inequality following from the fact that $H \neq 0$. We therefore have

$$\frac{1}{\#_{\overline{j},t_k}} \sum_{\substack{s=1 \\ s\ even}}^{t_k} E[\alpha_{\overline{j},s+1} \mid \mathcal{F}_s] \geq \frac{1}{\#_{\overline{j},t_k}} \sum_{\substack{s=1 \\ s\ even}}^{t_k} 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}}) \cdot c = c. \tag{15}$$

On the other hand,

$$\begin{aligned} &\sum_{\substack{s=1 \\ s\ even}}^{t_k} \left\{ \alpha_{\overline{j},s+1} - E[\alpha_{\overline{j},s+1} \mid \mathcal{F}_s] \right\} \\ &= \sum_{\substack{s=1 \\ s\ even}}^{t_k} 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}}) \left\{ (\|Hx_{s+1}\|^2 \wedge 1) - E[\|Hx_{s+1}\|^2 \wedge 1 \mid \mathcal{F}_s] \right\} \\ &= o\left(\sum_{\substack{s=1 \\ s\ even}}^{t_k} 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}}) \right), \end{aligned} \tag{16}$$

on the set where

$$\sum_{\substack{s=1 \\ s\ even}}^{t_k} 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}}) = \infty \tag{17}$$

(see [23]). Since (17) is satisfied if (12) holds, equations (15) and (16) prove that (13) follows from (12).

We now prove that (13) contradicts (9).

The convergence result $[A'_t, B'_t] \to [A^\circ, B^\circ]$ (see Theorem 1) implies that

$$\begin{aligned}\left([A'_t + B'_t K(A,B)] - [\overline{A} + \overline{B}K(A,B)]\right)^T \left([A'_t + B'_t K(A,B)] - [\overline{A} + \overline{B}K(A,B)]\right)\\ \geq \left(\frac{1}{2}H\right)^T \left(\frac{1}{2}H\right) \qquad\qquad \forall [A,B], [\overline{A},\overline{B}] \in N_{\overline{j}},\end{aligned}$$

for t sufficiently high (see (10)). In view of this, the following chain of inequalities can be derived when (12), and, consequently, inequality (13) hold true:

$$\begin{aligned}\infty &= \lim_{k\to\infty} \frac{1}{\mu_{t_k}} \#_{\overline{j},t_k} \cdot \liminf_{k\to\infty} \frac{1}{\#_{\overline{j},t_k}} \sum_{\substack{s=1\\ s\ even}}^{t_k} \alpha_{\overline{j},s+1}\\ &\leq \lim_{k\to\infty} \frac{1}{\mu_{t_k}} \sum_{\substack{s=1\\ s\ even}}^{t_k-2} \alpha_{\overline{j},s+1}\\ &\leq \lim_{k\to\infty} \frac{1}{\mu_{t_k}} \sum_{\substack{s=1\\ s\ even}}^{t_k-2} \|Hx_{s+1}\|^2 \cdot 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}})\\ &\leq \lim_{k\to\infty} \frac{1}{\mu_{t_k}} \sum_{\substack{s=1\\ s\ even}}^{t_k-2} 4\left\| \left\{[A'_{t_k} + B'_{t_k}\widehat{K}_s] - [\widehat{A}_{t_k} + \widehat{B}_{t_k}\widehat{K}_s]\right\} x_{s+1}\right\|^2 \cdot 1([\widehat{A}_s, \widehat{B}_s] \in N_{\overline{j}})\\ &\leq \lim_{k\to\infty} \frac{1}{\mu_{t_k}} \sum_{\substack{s=1\\ s\ even}}^{t_k-2} 4\left\| \left([A'_{t_k}, B'_{t_k}] - [\widehat{A}_{t_k}, \widehat{B}_{t_k}]\right) v_{s+1}\right\|^2.\end{aligned}$$

This contradicts (9). Thus, (12) is false with probability 1, and so (11) is proven. □

4. The self-tuning property. A key issue in the analysis of any adaptive control method consists of determining whether it is able to generate, at least asymptotically, control laws close to the optimal control law for the true system. The objective of the present section is to prove that this is indeed the case for our adaptive scheme, except for very rare time instants. This result will play a crucial role in the next section where we address stability and optimality issues.

THEOREM 4.

$$\sum_{s=1}^{t} 1(\|\widehat{K}_s - K^\circ\| > \rho) = O(\mu_t) \qquad a.s., \qquad \forall \rho > 0.$$

Proof. Since Θ is compact,

$$\sup_{[A,B]\in\Theta} \lambda_{\max}[A + BK(A,B)] < 1.$$

This implies that $A^\circ + B^\circ K(A,B)$ is stable for $[A,B]$ belonging to the closed set $\overline{C^c_\delta}$ (where the overbar indicates closure and the superscript "c" indicates the complement of the set), for δ small enough.

Denote by $J(A, B; K)$ the cost for the system $x_{t+1} = Ax_t + Bx_t + w_{t+1}$ controlled by $u_t = Kx_t$, whenever the corresponding closed-loop system is stable. It is known that (see [16] or [17])

$$J(A, B; K) = Trace P(A, B; K), \tag{18}$$

where $P(A, B; K)$ is the unique positive semidefinite solution of the Lyapunov equation

$$P = K^T RK + [A + BK]^T P[A + BK] + Q. \tag{19}$$

From this, it is easy to verify that $J(A^\circ, B^\circ; K(A, B))$ is a continuous function of $[A, B] \in \overline{C^c_\delta}$. On the other hand, the optimal gain K° for the true system (1) is unique within the class of stabilizing gains:

$$J(A^\circ, B^\circ; K) > J^\circ, \qquad \forall K \neq K^\circ, \qquad K \text{ stabilizing}.$$

Therefore, there exists $\nu(\rho) > 0$ such that every gain $K = K(A, B)$, $[A, B] \in \overline{C^c_\delta}$, for which

$$J(A^\circ, B^\circ; K) \leq J^\circ + \nu(\rho)$$

also satisfies the bound

$$\|K - K^\circ\| \leq \rho. \tag{20}$$

Note now that since $A + BK(A, B)$ is close to $A^\circ + B^\circ K(A, B)$ when $[A, B] \in \overline{C^c_\delta}$, δ small, from (19), we have

$$\sup_{[A,B] \in C^c_\delta} \|P(A^\circ, B^\circ; K(A, B)) - P(A, B; K(A, B))\| \to 0, \qquad \delta \to 0,$$

and, in view of (18),

$$\sup_{[A,B] \in C^c_s} |J(A^\circ, B^\circ; K(A, B)) - J(A, B; K(A, B))| \to 0, \qquad \delta \to 0.$$

Fix $\delta(\rho)$ such that

$$\sup_{[A,B] \in C^c_{\delta(\rho)}} \|J(A^\circ, B^\circ; K(A, B)) - J(A, B; K(A, B))\| \leq \frac{1}{2}\nu(\rho). \tag{21}$$

Finally,

$$\begin{aligned}
&\limsup_{t\to\infty} \frac{1}{\mu_t} \sum_{s=1}^{t} 1(\|\widehat{K}_s - K^\circ\| > \rho) \\
&\quad \leq \limsup_{t\to\infty} \frac{1}{\mu_t} \sum_{s=1}^{t} 1(J(A^\circ, B^\circ; \widehat{K}_s) - J^\circ > \nu(\rho)) \qquad \text{(using (20))} \\
&\quad \leq \limsup_{t\to\infty} \frac{1}{\mu_t} \sum_{s=1}^{t} 1\left(|J(A^\circ, B^\circ; \widehat{K}_s) - \widehat{J}_s| > \frac{1}{2}\nu(\rho)\right) \\
&\qquad \text{(using Theorem 2)} \\
&\quad \leq \limsup_{t\to\infty} \frac{1}{\mu_t} \sum_{s=1}^{t} 1\left([\widehat{A}_s, \widehat{B}_s] \in C_{\delta(\rho)}\right) \qquad \text{(using (21))} \\
&\quad < \infty \qquad \text{(using Theorem 3)}. \qquad \square
\end{aligned}$$

5. Stability and optimality. According to Theorem 4, the adaptive gain $\widehat{K}_s$ is close to the optimal gain K° except at very rare time instants, the number of which grows at most as μ_t. At these exceptional time points, the closed-loop system may be unstable. However, due to their rare occurrence, we establish that they cannot endanger the stability of the adaptive closed-loop control system. The corresponding stability result is given in Theorem 5. The proof of Theorem 5 relies heavily on the results of [24] concerning stability of rarely destabilized time-varying systems. It is very similar to that of Theorem 12 in [2] and is provided here only for the sake of completeness.

THEOREM 5.

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t}[\|x_s\|^p + \|u_s\|^p] < \infty \qquad a.s., \qquad \forall p > 0.$$

Proof. We start by noting that, since $A^\circ + B^\circ K^\circ$ is a stable matrix, there exists a suitable norm on $\mathbf{R}^n$ such that, under the corresponding induced matrix norm, $\|A^\circ + B^\circ K^\circ\| < 1$ (see, e.g., [25]). Throughout this proof all the norm symbols refer to this particular norm.

It is easy to verify that the following inequality holds true for any integer n and real numbers a, b, and $\epsilon > 0$,

$$(a+b)^{2^n} \le (1+\epsilon^2)^{2^n-1}a^{2^n} + (1+\epsilon^{-2})^{2^n-1}b^{2^n}. \tag{22}$$

Taking into account the relation $x_{t+1} = A^\circ x_t + B^\circ \widehat{K}_t x_t + w_{t+1}$, from (22) we obtain

$$\|x_{t+1}\|^{2^n} \le (1+\epsilon^2)^{2^n-1}\|A^\circ + B^\circ\widehat{K}_t\|^{2^n}\|x_t\|^{2^n} + (1+\epsilon^{-2})^{2^n-1}\|w_{t+1}\|^{2^n},$$

for any integer n and positive real ϵ.

Now fix $\bar{n}$ such that $2^{\bar{n}} \ge p$ and choose $\bar{\epsilon} > 0$ such that $(1+\bar{\epsilon}^2)^{2^{\bar{n}}-1}\|A^\circ + B^\circ K^\circ\|^{2^{\bar{n}}} < 1$. Further, select ρ in such a way that

$$a := \sup_{K\,:\,\|K-K^\circ\|\le\rho} (1+\bar{\epsilon}^2)^{2^{\bar{n}}-1}\|A^\circ + B^\circ K\|^{2^{\bar{n}}} < 1$$

and also let

$$b := \sup_{[A,B]\in\Theta} (1+\bar{\epsilon}^2)^{2^{\bar{n}}-1}\|A^\circ + B^\circ K(A,B)\|^{2^{\bar{n}}}.$$

Then

$$\|x_{t+1}\|^{2^{\bar{n}}} \le \gamma_t\|x_t\|^{2^{\bar{n}}} + (1+\bar{\epsilon}^{-2})^{2^{\bar{n}}-1}\|w_{t+1}\|^{2^{\bar{n}}}, \tag{23}$$

where

$$\gamma_t = \begin{cases} a, & \text{if } \|\widehat{K}_t - K^\circ\| \le \rho, \\ b, & \text{otherwise.} \end{cases}$$

We now apply Theorem 2 in [24] to (23) (see also Remark 1 in the same paper). By noting that $\sum_{s=1}^{t} 1(\|\widehat{K}_s - K^\circ\| > \rho) = O(\mu_t)$ (Theorem 4) and that $\mu_t = o(\log t)$, from that theorem we can conclude that

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t}\|x_s\|^{2^{\bar{n}}} < \infty \qquad \text{a.s.}$$

This implies that (recall that $2^{\bar{n}} \geq p$)

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} \|x_s\|^p < \infty \qquad \text{a.s.}$$

Since $\|u_s\| \leq \sup_{[A,B]\in\Theta} \|K(A,B)\|\|x_s\|$, we also have

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} \|u_s\|^p < \infty \qquad \text{a.s.}$$

This proves the stability result. □

We are now in a position to prove the optimality of the adaptive scheme, namely, that the incurred cost equals the optimal cost that could be obtained if the true system parameter were known at the start.

THEOREM 6.

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} \left[x_s^T Q x_s + u_s^T R u_s\right] = J^\circ \qquad \textit{a.s.}$$

Proof. The dynamic programming equation for model $x_{s+1} = \widehat{A}_s x_s + \widehat{B}_s u_s + w_{s+1}$ is (see [16])

$$\begin{aligned}
&\widehat{J}_s + x_s^T \widehat{P}_s x_s \\
&\quad = x_s^T Q x_s + u_s^T R u_s + E[(\widehat{A}_s x_s + \widehat{B}_s u_s + w_{s+1})^T \widehat{P}_s (\widehat{A}_s x_s + \widehat{B}_s u_s + w_{s+1}) \mid \mathcal{F}_s] \\
&\quad = x_s^T Q x_s + u_s^T R u_s + E[x_{s+1}^T \widehat{P}_s x_{s+1} \mid \mathcal{F}_s] \\
&\quad + \left\{ (\widehat{A}_s x_s + \widehat{B}_s u_s)^T \widehat{P}_s (\widehat{A}_s x_s + \widehat{B}_s u_s) - (A^\circ x_s + B^\circ u_s)^T \widehat{P}_s (A^\circ x_s + B^\circ u_s) \right\}.
\end{aligned}$$

From this,

$$\begin{aligned}
&\underbrace{\frac{1}{t}\sum_{s=1}^{t} \widehat{J}_s}_{A} + \underbrace{\frac{1}{t}\sum_{s=1}^{t} \left\{ x_s^T \widehat{P}_s x_s - E[x_{s+1}^T \widehat{P}_{s+1} x_{s+1} \mid \mathcal{F}_s] \right\}}_{B} \\
&\qquad = \frac{1}{t}\sum_{s=1}^{t} [x_s^T Q x_s + u_s^T R u_s] + \underbrace{\frac{1}{t}\sum_{s=1}^{t} E[x_{s+1}^T (\widehat{P}_s - \widehat{P}_{s+1}) x_{s+1} \mid \mathcal{F}_s]}_{C} \\
&\qquad + \underbrace{\frac{1}{t}\sum_{s=1}^{t} \left\{ (\widehat{A}_s x_s + \widehat{B}_s u_s)^T \widehat{P}_s (\widehat{A}_s x_s + \widehat{B}_s u_s) - (A^\circ x_s + B^\circ u_s)^T \widehat{P}_s (A^\circ x_s + B^\circ u_s) \right\}}_{D}.
\end{aligned} \tag{24}$$

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

(A) From Theorem 2 we have

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} \widehat{J}_s \leq J^\circ.$$

(B)

$$\frac{1}{t}\sum_{s=1}^{t}\left\{x_s^T\widehat{P}_s x_s - E[x_{s+1}^T\widehat{P}_{s+1}x_{s+1} \mid \mathcal{F}_s]\right\}$$

$$= \frac{1}{t}x_1^T\widehat{P}_1 x_1 - \frac{1}{t}x_{t+1}^T\widehat{P}_{t+1}x_{t+1} + \frac{1}{t}\sum_{s=1}^{t}\left\{x_{s+1}^T\widehat{P}_{s+1}x_{s+1} - E[x_{s+1}^T\widehat{P}_{s+1}x_{s+1} \mid \mathcal{F}_s]\right\}.$$

The first term obviously tends to zero. As for the second one, note that, $\widehat{P}_{t+1} \leq \sup_{[A,B]\in\Theta} P(A,B)$ being bounded, it tends to zero provided that $\|x_t\|^2/t \to 0$. The fact that this is the case can be proven by contradiction. Suppose that there exists a time sequence $\{t_k\}$ and a real number $\alpha > 0$ such that $\|x_{t_k}\|^2 > \alpha t_k, \forall k$. Then $\limsup_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t}\|x_s\|^4 \geq \limsup_{k\to\infty}\frac{1}{t_k}\|x_{t_k}\|^4 \geq \limsup_{k\to\infty}\frac{1}{t_k}\alpha^2 t_k^2 = \infty$. This contradicts Theorem 5. In the third term,

$$\{\alpha_{s+1}\} := \{x_{s+1}^T\widehat{P}_{s+1}x_{s+1} - E[x_{s+1}^T\widehat{P}_{s+1}x_{s+1} \mid \mathcal{F}_s]\}$$

is a martingale difference. Therefore, $\frac{1}{t}\sum_{s=1}^{t}\alpha_{s+1} \to 0$, provided that

$$\sum_{s=1}^{\infty} s^{-2}E[\alpha_{s+1}^2 \mid \mathcal{F}_s] < \infty$$

(see [26]). Since $\widehat{P}_{s+1}$ is bounded, it is easily seen that this last condition is implied by $\sum_{s=1}^{\infty} s^{-2}[\|x_s\|^4 + \|u_s\|^4] < \infty$. Again, this conclusion can be drawn by contradiction from Theorem 5. In fact, if this conclusion were false, sequence $s^{-1/2}[\|x_s\|^4 + \|u_s\|^4]$ would be unbounded and, therefore, there would exist a sequence of times $\{t_k\}$ such that $[\|x_{t_k}\|^4 + \|u_{t_k}\|^4] \geq t_k^{1/2}\forall k$. From this, $\limsup_{k\to\infty}\frac{1}{t}\sum_{s=1}^{t}[\|x_s\|^4 + \|u_s\|^4]^4 \geq \limsup_{k\to\infty}\frac{1}{t_k}[\|x_{t_k}\|^4 + \|u_{t_k}\|^4]^4 \geq \limsup_{k\to\infty}\frac{1}{t_k}t_k^2 = \infty$, and this is in contradiction with Theorem 5. In conclusion,

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t}\left\{x_s^T\widehat{P}_s x_s - E[x_{s+1}^T\widehat{P}_{s+1}x_{s+1} \mid \mathcal{F}_s]\right\} = 0 \qquad \text{a.s.}$$

(C) We start by proving that

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t}\|\widehat{P}_s - \widehat{P}_{s+1}\|^2 = 0 \qquad \text{a.s.} \tag{25}$$

Since P° satisfies the equation

$$P^\circ = K^{\circ T}RK^\circ + [A^\circ + B^\circ K^\circ]^T P^\circ [A^\circ + B^\circ K^\circ] + Q,$$

and $\widehat{P}_s$ satisfies the equation

$$\widehat{P}_s = \widehat{K}_s^T R\widehat{K}_s + [\widehat{A}_s + \widehat{B}_s\widehat{K}_s]^T\widehat{P}_s[\widehat{A}_s + \widehat{B}_s\widehat{K}_s] + Q,$$

P° is close to $\widehat{P}_s$ when K° is close to $\widehat{K}_s$ and $A^\circ + B^\circ K^\circ$ is close to $\widehat{A}_s + \widehat{B}_s \widehat{K}_s$. In view of Theorems 3 and 4, the total of the numbers of time points in which this does not happen is $O(\mu_t)$. Therefore,

$$\sum_{s=1}^{t} 1(\|\widehat{P}_s - P^\circ\| > \rho) = O(\mu_t) \qquad \text{a.s.,} \qquad \forall \rho > 0.$$

Equation (25) then easily follows from

$$\begin{aligned}\frac{1}{t}\sum_{s=1}^{t} \|\widehat{P}_s - \widehat{P}_{s+1}\|^2 &\leq \frac{2}{t}\sum_{s=1}^{t}\left[\|\widehat{P}_s - P^\circ\|^2 + \|\widehat{P}_{s+1} - P^\circ\|^2\right] \\ &\leq \frac{4}{t}\sum_{s=1}^{t+1} \|\widehat{P}_s - P^\circ\|^2 1(\|\widehat{P}_s - P^\circ\| > \rho) + \frac{4(t+1)}{t}\rho^2 \\ &\to 4\rho^2,\end{aligned}$$

since ρ is an arbitrary positive real number.

Notice now that, by the Schwarz inequality,

$$\frac{1}{t}\sum_{s=1}^{t} |x_{s+1}^T(\widehat{P}_s - \widehat{P}_{s+1})x_{s+1}| \leq \left(\frac{1}{t}\sum_{s=1}^{t}\|\widehat{P}_s - \widehat{P}_{s+1}\|^2\right)^{1/2}\left(\frac{1}{t}\sum_{s=1}^{t}\|x_{s+1}\|^4\right)^{1/2}.$$

Therefore, $t^{-1}\sum_{s=1}^{t}\|x_{s+1}\|^4$ being bounded (Theorem 5), (25) implies

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} x_{s+1}^T(\widehat{P}_s - \widehat{P}_{s+1})x_{s+1} = 0 \qquad \text{a.s.} \tag{26}$$

Finally, the conclusion

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} E[x_{s+1}^T(\widehat{P}_s - \widehat{P}_{s+1})x_{s+1} \mid \mathcal{F}_s] = 0 \qquad \text{a.s.}$$

is drawn from (26) by observing that

$$\{\beta_{s+1}\} := \{x_{s+1}^T(\widehat{P}_s - \widehat{P}_{s+1})x_{s+1} - E[x_{s+1}^T(\widehat{P}_s - \widehat{P}_{s+1})x_{s+1} \mid \mathcal{F}_s]\}$$

is a martingale difference for which, by calculations resembling those developed in point (B), $\sum_{s=1}^{\infty} s^{-2}E[\beta_{s+1}^2 \mid \mathcal{F}_s] < \infty$.

(D) Since

$$\|P^TP - R^TR\| \leq \|P - R\|(\|P\| + \|R\|), \qquad \forall P, R \in \mathbb{R}^{n\times n},$$

we have

$$\begin{aligned}&|(\widehat{A}_s x_s + \widehat{B}_s u_s)^T\widehat{P}_s(\widehat{A}_s x_s + \widehat{B}_s u_s) - (A^\circ x_s + B^\circ u_s)^T\widehat{P}_s(A^\circ x_s + B^\circ u_s)| \\ &\quad = |x_s^T(\widehat{A}_s + \widehat{B}_s\widehat{K}_s)^T\widehat{P}_s(\widehat{A}_s + \widehat{B}_s\widehat{K}_s)x_s - x_s^T(A^\circ + B^\circ\widehat{K}_s)^T\widehat{P}_s(A^\circ + B^\circ\widehat{K}_s)x_s| \\ &\quad \leq \|x_s\|^2\|\widehat{P}_s\|\ \|(\widehat{A}_s + \widehat{B}_s\widehat{K}_s) - (A^\circ + B^\circ\widehat{K}_s)\|(\|\widehat{A}_s + \widehat{B}_s\widehat{K}_s\| + \|A^\circ + B^\circ\widehat{K}_s\|).\end{aligned}$$

Also, $\|\widehat{P}_s\|$ is uniformly bounded over time. The same holds for $(\|\widehat{A}_s+\widehat{B}_s\widehat{K}_s\|+\|A^\circ+B^\circ\widehat{K}_s\|)$. Furthermore, using the Schwarz inequality,

$$\frac{1}{t}\sum_{s=1}^{t}\|x_s\|^2\|(\widehat{A}_s+\widehat{B}_s\widehat{K}_s)-(A^\circ+B^\circ\widehat{K}_s)\|$$

$$\leq\left(\frac{1}{t}\sum_{s=1}^{t}\|x_s\|^4\right)^{1/2}\left(\frac{1}{t}\sum_{s=1}^{t}\|(\widehat{A}_s+\widehat{B}_s\widehat{K}_s)-(A^\circ+B^\circ\widehat{K}_s)\|^2\right)^{1/2}.$$

By Theorem 5 the first term is bounded. In light of Theorem 3, the second term can be handled analogously to the calculations for $t^{-1}\sum_{s=1}^{t}\|\widehat{P}_s-\widehat{P}_{s+1}\|^2$ in point (C), thus yielding

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t}\|(\widehat{A}_s+\widehat{B}_s\widehat{K}_s)-(A^\circ+B^\circ\widehat{K}_s)\|^2=0 \qquad \text{a.s.}$$

This suffices to prove that $D\to 0$, a.s.

By inserting all the partial results in (24) we finally obtain

$$\limsup_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t}\left[x_s^TQx_s+u_s^TTu_s\right]\leq J^\circ \qquad \text{a.s.}$$

Since J° is the optimal cost for the true system, this proves the theorem. ∎

6. Concluding remarks. In an adaptive control context, the minimization of a given cost function is made difficult by the general identifiability problem stemming from the natural tendency of classical identification methods to return estimates with the corresponding optimal cost larger than the optimal cost for the true system. A way out of this problem is to employ a more fine-grained estimation scheme which exploits the properties of the set to which the estimates converge. Such a scheme has been presented and analyzed in this paper for the linear quadratic Gaussian control problem.

The results of this paper need to be extended in several directions to provide a fuller theory of optimal adaptation:

- *The presented scheme is nonrecursive.* However, one can conceive of somehow recursively minimizing our identification performance index so as to retain its asymptotic identification properties. This must be further investigated.
- *We assume full state observations.* This limitation needs to be removed.
- *Our adaptive scheme is, to some extent, tailored to linear quadratic Gaussian control.* In particular, a central role in the analysis is played by the uniqueness of the optimal gain in linear quadratic Gaussian control problems. It would be of interest to investigate how the biasing idea applies to other control strategies. An additional point is concerning the Gaussianity of the noise. This assumption is exploited in proving that the least squares estimate converges and that it tends to the true value in the excited subspace. In an attempt to remove the Gaussianity assumption one can use a weighted least squares algorithm, as suggested in [12], guaranteeing estimate convergence. In doing so, however, consistency in the excited subspace is lost and this may pose a difficulty in the derivation of many results.

• *Assumption $\mu_t = o(\log t)$ may be very conservative.* It is mainly motivated by the stability analysis and it is possible that our results still hold with μ_t growing at a faster rate. This and other choices made in the definition of our algorithm may be further investigated.

All the above problems suggest interesting research opportunities and a promise of self-optimizing adaptive control for nonlinear stochastic systems.

REFERENCES

[1] V. Borkar and P. P. Varaiya, *Adaptive control of Markov chains,* I: *Finite parameter set*, IEEE Trans. Automat. Control, 24 (1979), pp. 953–958.
[2] P. R. Kumar, *Optimal adaptive control of linear-quadratic-Gaussian systems*, SIAM J. Control Optim., 21 (1983), pp. 163–178.
[3] A. Becker, P. R. Kumar, and C. Z. Wei, *Adaptive control with the stochastic approximation algorithm: Geometry and convergence*, IEEE Trans. Automat. Control, 30 (1985), pp. 330–338.
[4] W. Lin, P. R. Kumar, and T. I. Seidman, *Will the self-tuning approach work for general cost criteria?*, Systems Control Lett., 6 (1985), pp. 77–85.
[5] T. E. Duncan and B. Pasik-Duncan, *Adaptive control of continuous-time linear stochastic systems*, Math. Control Signals Systems, 3 (1990), pp. 45–60.
[6] T. E. Duncan, P. Mandl, and B. Pasik-Duncan, *Control theory methods for consistency in some least squares identification problems,* IEEE Trans. Automat. Control, 38 (1993), pp. 1289–1292.
[7] H. F. Chen and L. Guo, *Convergence rate of least squares identification and adaptive control for stochastic systems*, Internat. J. Control, 44 (1986), pp. 1459–1476.
[8] H. F. Chen and L. Guo, *Asymptotically optimal adaptive control with consistent parameter estimates*, SIAM J. Control Optim., 25 (1987), pp. 558–575.
[9] H. F. Chen and L. Guo, *Optimal adaptive control and consistent parameter estimates for ARMAX model with quadratic cost*, SIAM J. Control Optim., 25 (1987), pp. 845–867.
[10] H. F. Chen and L. Guo, *A robust stochastic adaptive controller*, IEEE Trans. Automat. Control, 33 (1988), pp. 1035–1043.
[11] H. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Birkhäuser, Basel, 1991.
[12] L. Guo, *Self-convergence of weighted least-squares with applications to stochastic adaptive control*, IEEE Trans. Automat. Control, 41 (1996), pp. 79–89.
[13] P. R. Kumar and A. Becker, *A new family of optimal adaptive controllers for Markov chains*, IEEE Trans. Automat. Control, 27 (1982), pp. 137–146.
[14] W. Lin and P. R. Kumar, *Stochastic control of a queue with two servers of different rates*, in Analysis and Optimization of Systems, Lecture Notes in Control and Inform. Sci., A. Bensoussan and J. L. Lions, eds., Springer-Verlag, 1982, Chap. 44.
[15] P. R. Kumar, *Simultaneous identification and adaptive control of unknown systems over finite parameter sets*, IEEE Trans. Automat. Control, 28 (1983), pp. 68–76.
[16] P. R. Kumar and P. P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice–Hall, Englewood Cliffs, NJ, 1986.
[17] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice–Hall, Englewood Cliffs, NJ, 1987.
[18] P. R. Kumar, *Convergence of least-squares parameter estimate based adaptive control schemes*, IEEE Trans. Automat. Control, 35 (1990), pp. 416–424.
[19] S. Bittanti, P. Bolzern, and M.C. Campi, *Recursive least squares identification algorithms with incomplete excitation: Convergence analysis and application to adaptive control*, IEEE Trans. Automat. Control, 35 (1990), pp. 1371–1373.
[20] H. Rootzen and J. Sternby, *Consistency in least squares estimation: A Bayesian approach*, Automatica, 20 (1984), pp. 471–475.
[21] M. C. Campi, *The problem of pole-zero cancellation in transfer function identification and application to adaptive stabilization*, Automatica, 32 (1996), pp. 849–857.
[22] D. Delchamps, *Analytic feedback control and the algebraic Riccati equation*, IEEE Trans. Automat. Control, 29 (1984), pp. 1031–1033.
[23] Y. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, 2nd ed., Springer-Verlag, New York, 1988.

[24] P. R. Kumar and T. I. Seidman, *Stability in the sense of bounded average power*, IMA J. Math. Control Inform., 2 (1985), pp. 109–122.

[25] J. M. Ortega, *Numerical Analysis: A Second Course*, Academic Press, New York, 1972.

[26] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, Toronto, 1980.