Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

MoC11.1

Stochastic linear–quadratic adaptive control: A conceptual scheme

David Levanony Peter E. Caines

Abstract A conceptual adaptive linear–quadratic (LQ) control scheme is proposed. Its derivation is based on a study of a family of *asymptotic maximum likelihood* (AML) estimators, and their associated limit sets. The geometric properties of such limit sets, lead to the formulation of a time–varying, constrained optimization problem, whose solution is an inherently consistent estimate of the system's unknown parameters. When incorporated within a certainty–equivalence adaptive control scheme, these estimates yield optimal long–run LQ closed–loop performance.

I. INTRODUCTION

The work of Kumar and Becker [10] and Kumar and Lin [11] introduced the technique of biasing standard parameter estimation schemes in adaptive control algorithms for controlled Markov chains. This biasing is a function of the system performance that would result if the true system were described by the current parameter estimate. The basic idea is to direct the process of parameter estimates in a way that takes account of both identification and steady state system performance. This line of research was continued for controlled diffusion processes by Borkar [1], and, more recently, by Campi and Kumar [4] and Prandini and Campi [21]. One notable advantage of this approach is that it permits a significant weakening of the persistent excitation (PE) requirement that is to be found in the work on consistency based stochastic adaptive control and which is a particular difficulty in the adaptive stabilization in both the continuous and discrete time cases (see Lai and Wei [14], Caines [2, 3], Chen and Guo [5]).

The adaptive control scheme proposed in this paper involves parameter estimates which are *implicitly* biased. Unlike the work cited above, these estimates are a product of the solution of a time–varying, constrained optimization problem, constructed so as to make its solution inherently consistent, thereby, incorporated within a certainty–equivalence scheme, yielding optimal long–run LQ performance. The adaptive scheme discussed in this paper is regarded as *conceptual*, since any application would require an instantaneous optimization procedure. It is therefore that such a scheme is not recursively realizable (as are all biased-cost schemes derived in the above cited publications, with the exception of [10, 11] in which *finite* parameter sets are considered). Nonetheless, based on the work presented here, a recursive algorithm, which generates *approximate* finite time solutions, while still maintaining consistency, is derived in [17, 19].

A particular feature of our approach is the use of the geometric analysis of Polderman [22, 23, 24] who studied the structure of the sets of parameters corresponding to (i) indistinguishable closed loop dynamical behaviour and (ii) optimal LQ closed loop performance. This analysis was specifically carried out with the analysis of LQ adaptive control schemes in view.

Consider a class of completely observed LTI systems whose states evolve according to the Îto equations

$$dx_t = Ax_t dt + Bu_t dt + Cdw_t,$$

where x, u, w take values in $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^r$ (respectively) and w is a standard Brownian process independent of x. For a given system, parameterized by unknown $(A^*, B^*) \stackrel{\Delta}{=} \theta^*$, our objective is to generate estimates $\{\theta_t; t \geq t_o\}$ to be used in an LQ certainty equivalence control to obtain the optimal LQ (long run) cost which would be obtained if θ^* were known. In the solution to this problem provided in this paper we consider the same control policy as described in Caines [4]. The basic difference lies in the parameter estimation algorithm. In Caines [4] a standard RLS algorithm has been employed under a (sample-wise) PE condition which requires independent verification. In the adaptive parameter estimation and control work of Duncan and Pasik-Duncan [7, 8], the role of this PE condition is effectively replaced by the condition that a certain determinant $det\widetilde{A}_t$ shall be almost surely bounded away from zero. This condition is verifiable in certain cases of interest. Alternatively, the required PE property can be *created* by the injection of a diminishing excitation signal, as has been shown by Duncan, Guo and Pasik–Duncan [9]. In this work, no external dither is utilized.

Let $\phi_t^T = (x_t^T, -x_t^T K_t^T)$ denote the regression vector where $K_t = K(\theta_t)$ is computed via the control law; then, roughly speaking, one class of PE conditions is equivalent to assuming that the matrix $\int_0^t \phi_s \phi_s^T ds$ (properly normalized) converges to a strictly positive definite limit. Another important class is that involving the comparative rates of growth of eigenvalues of this matrix (see Lai and Wei [14], Chen and Guo [5]). Note, however, that if $K_t \rightarrow K$, the existence of such a positive definite limit is highly questionable. PE conditions are almost invariably used to ensure consistency; an alternative formulation, which we pursue in this work, is that where one simply considers the natural convergence $\theta_t \rightarrow \mathcal{I}$, as $t \rightarrow \infty$, where $\mathcal{I}$ is some limit set, which occurs without further conditions for essentially all parameter estimation algorithms. (In our

Dept. of Electrical and Computer Engineering, Ben Gurion University, Beer Sheva 84105, Israel. levanony@ee.bgu.ac.il.

Dept. of Electrical Engineering, McGill University, Montreal, Quebec H2A 2A7, Canada, and The Canadian Institute for Advanced Research. peterc@cim.mcgill.ca

case, $\mathcal{I}$ corresponds to parameters yielding indistinguishable system trajectories.) Our examination of the limit set $\mathcal{I}$ is inspired by the topological dynamics formulation of stability theory due to La Salle [16].

Let $J(\theta)$ be the long run optimal performance cost for a system with $(A, B) = \theta$. As has been established by Polderman [22, 23], the limit set of indistinguishable dynamics

$$\mathcal{I} = \{(A, B) | A - BK(A, B) = A^* - B^* K(A, B)\}$$

is a smooth manifold upon which J has a unique minimum θ^* (= the true parameter) which, in addition, we show has no local stationary points other than θ^* on $\mathcal{I}$. With those properties in mind, the intuition behind the performance biased adaptive control algorithm is as follows: A conventional certainty equivalence adaptive controller, using an MLE-type parameter estimates is utilized. The MLE-type estimates, being members of a properly defined family of AML estimates, are generated as solutions of a time–varying, constrained optimization problem. More specifically, minimization of the control cost J, subject to the constraint that the gradient of the log–likelihood is bounded by a vanishing function $\{\delta_t,\ t \geq 0\}$, is sought–after. Hence, as the time–varying constraint makes the resulting solutions (or estimates) to converge to the common AML limit set $\mathcal{I}$, it follows that minimization of J is *asymptotically* restricted to $\mathcal{I}$, over which θ^* is the unique minimizer of J. With the desired strong consistency thereby secured, optimal, closed-loop LQ performance follows.

Our results are organized as follows: The adaptive control problem is formulated in the next Section 2. In Section 3, a class of AML estimates is defined and further, is shown to have a common limit set $\mathcal{I}$. Section 4 is devoted to the investigation of the geometric properties of the limit set $\mathcal{I}$. Finally, in Section 5, the aforementioned time–varying, constrained optimization problem is formulated and its solutions are shown to be consistent estimates of θ^*. The resulting optimal LQ performance is then established.

Due to obvious constrains on the paper's length, many of the proofs are omitted. These appear in [17].

II. Problem Statement

We consider the system

$$dx_t = Ax_t dt + Bu_t dt + C dw_t, \qquad \text{(II.1)}$$

where $x_t \,\epsilon\, \mathbb{R}^n, u_t \,\epsilon\, \mathbb{R}^m$ and $w_t \,\epsilon\, \mathbb{R}^p$ for all $t \geq 0, x_0$ is a non-random initial condition and w is a standard Brownian motion. Let the process $w = \{w_t,\ t \geq 0\}$ be measurable with respect to the increasing family of σ-fields $\mathcal{F}_t$ for all $t \geq 0$. The solution process $\{x_t,\ t \geq 0\}$ for (II.1) is dependent upon the values taken by A, B, C when these are treated as non–random, time independent variables.

Let x generate the increasing family of σ-fields $\mathcal{F}^x_t, t \geq 0$. Then the process u is assumed to satisfy the (adaptive) non-anticipating control condition that u_t is only an $\mathcal{F}^x_t$ measurable function for $t \geq 0$, and hence is not an explicit function of A, B, C. We shall denote this condition by $u \,\epsilon\, \mathcal{U}$.

The objective in the application of adaptive control to the system (II.1) is to achieve, along almost all sample paths,

$$J^o_\infty \triangleq \inf_{u \epsilon \mathcal{U}} \lim_{t \to \infty} \frac{1}{t} \int_0^t (\|x_s\|^2 + \|u_s\|^2) ds \qquad \text{(II.2)}$$

where the control laws employed are such that the inner limit exists, along almost all paths.

Let us define the matrix parameter $\Theta = [A, B]^T$ and the corresponding vector parameter $\theta = [co\ell(A, B)] \,\epsilon\, \mathbb{R}^{n(n+m)}$, where $[co\ell(A, B)]^T = [(A_1^T, B_1^T) \cdots, (A_n^T, B_n^T)]$, with A_i (respectively B_i) denoting the i-th row of A (respectively B). Further define the $n \times (n + m)n$ matrix Ψ_t by

$$\Psi_t = \begin{bmatrix} \phi_t^T & 0 & \cdots & 0 \\ 0 & \phi_t^T & 0 & 0 \\ 0 & \cdots & \cdots & \phi_t^T \end{bmatrix},$$

where $\phi_t^T \triangleq (x_t^T, u_t^T)$. Assuming a full rank noise (i.e. $CC^T > 0$), we take for simplicity $C = I$. The system equation (II.1) may be conveniently re-expressed as

$$dx_t = \Psi_t \theta dt + dw_t, \quad t \geq 0. \qquad \text{(II.3)}$$

For convenience, when $u \,\epsilon\, \mathcal{U}$, we shall refer to both (II.1) and (II.3) as the system $\Xi(\theta)$.

We use the notation θ^* to denote the value of the deterministic parameter of the system (II.3) generating a given set of observations for a given control law $u \,\epsilon\, \mathcal{U}$. This parameter θ^* will be referred to as the true system parameter and we assume that

$$\theta^* \,\epsilon\, \mathcal{S} \triangleq \{\theta = co\ell(A, B) : (A, B) \text{ stabilizable}\}.$$

The adaptive control algorithms we study in this paper are based upon the class of certainty equivalence (CE) algorithms which have the following form for the adaptive LQ problem: for each $t \geq 0$,

(i) Compute an estimate $\theta_t \,\epsilon\, \mathcal{S}$ of $\theta^* \,\epsilon\, \mathcal{S}$.

(ii) Use the feedback control law

$$u_t = -K(\theta_t) x_t,$$

where

$$K(\theta) = K(A, B) = B^T V(A, B) \qquad \text{(II.4)}$$

where $V(A, B)$ is the (unique) positive definite symmetric solution to the algebraic Riccati equation (ARE):

$$A^T V + VA - VBB^T V + I = 0. \qquad \text{(II.5)}$$

It is well known that for any given system, parameterized by a stabilizable pair (A, B), the optimal achievable performance (II.2) equals to

$$J(A, B) \triangleq \mathrm{Tr} V(A, B). \qquad \text{(II.6)}$$

Therefore, (with a slight abuse of notation) we shall refer to $J(\theta) \triangleq \mathrm{Tr} V(\theta)$ as the (synthetic) cost function. As will be shown below, a minimization of (II.6), which takes place during the adaptation procedure, leads to the desired optimal performance, that is, along almost all sample paths, one obtains,

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t (\|x_s\|^2 + \|u_s\|^2)ds = J(\theta^*) = J^o_\infty. \quad \text{(II.7)}$$

One of the motives for the use of the gradient search maximum likelihood (ML) class of parameter estimation schemes in the adaptive control algorithm introduced in this paper is that such schemes are sufficiently flexible to permit various modifications to the algorithm while retaining consistency; another motive is that they are comparatively efficient numerically. To facilitate the analysis in this paper, we reduce the proof of the consistency of the ML scheme (for systems subject to feedback control which is dependent upon current parameter estimates) to a convergence analysis of recursive least squares schemes (for feedback systems) and this, in turn, is reduced to the study of minimum variance estimates (i.e. conditional expectations) within a Bayesian framework. As a consequence, after the presentation of an introductory result (Lemma 3.1) in Section 3 concerning AML estimates, we proceed to establish a Bayesian convergence result (Theorem 3.2) and then an RLS convergence result (Corollary 3.3).

III. Asymptotic Maximum Likelihood Estimation

Let $\{(u_t, x_t),\ t \geq 0\}$ denote an observed input-state sample path of the system $\Xi(\theta^*)$, and let $L_t(\theta) \triangleq L(\theta, (u_0^t, x_0^t)) = \int_0^t \theta^T \Psi_s^T dx_s - \frac{1}{2}\int_0^t |\Psi_s \theta|^2 ds$ denote a log-likelihood function of (u_0^t, x_0^t) at any $\theta \,\epsilon\, \mathcal{S}$ whose gradient $\nabla L_t(\theta)$ is given by,

$$\begin{aligned}\nabla L_t(\theta) &= \int_0^t \Psi_s^T dx_s - \int_0^t \Psi_s^T \Psi_s ds\theta \\ &= \int_0^t \Psi_s^T dw_s - \int_0^t \Psi_s^T \Psi_s ds(\theta - \theta^*) \\ &\triangleq m_t - \Phi_t \widetilde{\theta}, \end{aligned} \quad \text{(III.1)}$$

where $m_t, t \geq 0$ is an $n(n+m)$ dimensional martingale, $\widetilde{\theta} \triangleq \theta - \theta^*$ and

$$\Phi_t \triangleq \int_0^t \Psi_s^T \Psi_s ds.$$

In terms of (III.1) we may present the following preliminary maximum likelihood estimation (MLE) result, which we note does not constitute a consistency result without the addition of further hypotheses.

Lemma 3.1

For the system $\Xi(\theta^*)$, and the observed process $\{(u_t, x_t),\ t \geq 0\}$, let $\{\theta_t = \theta_t(\theta^*, \omega),\ t \geq 0\}$, be a process which is progressively measurable with respect to the σ-fields $\mathcal{F}_t^x, t \geq 0$. Assume that $\{\theta_t,\ t \geq 0\}$ belongs to the class of *Asymptotic ML estimates* ($\mathcal{AML}$) in the sense that

$$\nabla L_t(\theta_t) \to 0 \qquad \text{a.s. as } t \to \infty, \quad \text{(III.2)}$$

which is denoted by $\{\theta_t,\ t \geq 0\} \,\epsilon\, \mathcal{AML}$. Further assume that almost surely Φ_t is non-singular for all t sufficiently large. Then there exists an a.s. finite random variable $\theta_\infty = \theta_\infty(\theta^*, \omega)$ for which

$$\theta_t \to \theta_\infty \qquad \text{a.s. as } t \to \infty \quad \text{(III.3)}$$

for all $\theta^* \notin \mathcal{N}$, where $\mathcal{N}$ is a Lebesgue null set in $\mathbb{R}^{n(n+m)}$ independent of ω.

□

The characterization of the limit set of $\{\theta_t\} \,\epsilon\, \mathcal{AML}$ is the purpose of the final phase of this section. Following the Bayesian embedding approach (Kumar [13]), we begin with a consistency proof in a Gaussian setting:

Theorem 3.2

Consider the system (II.1) where it is assumed that $C = I$, (A^*, B^*, x_0) have a joint Gaussian distribution and $w = \{w_t, t \geq 0\}$ is a standard Brownian vector process, independent of (A^*, B^*, x_0).

Suppose that the system is controlled by $u_t = -K_t x_t,\ t \geq 0$, where K_t is a causal, $\mathcal{F}_t^x$-measurable, continuous and bounded semi-martingale (matrix) which converges a.s. to a finite, possibly random limit K_∞. Let

$$\widehat{\Theta}_t \triangleq \mathrm{E}[\Theta^* | \mathcal{F}_t^x] = \mathrm{E}([A^*, B^*]^T | \mathcal{F}_t^x) \quad \text{(III.4)}$$

be the MV (minimum variance) estimate of $\Theta^* \triangleq [A^*, B^*]^T$. Then

$$\widehat{\Theta}_t \to \{[A, B]^T : A - BK_\infty = A^* - B^* K_\infty\}, \text{ a.s., as } t \to \infty \quad \text{(III.5)}$$

Corollary 3.3

Let Θ_t^{RLS} be the matrix RLS estimate corresponding to θ_t^{RLS} of Lemma 3.1 (i.e. $\theta_t^{RLS} = co\ell\{\Theta_t^{RLS}\}$). Then, under the conditions of Theorem 3.2 but with $[A^*, B^*]$ deterministic and stabilizable, Θ_t^{RLS} is a.s. convergent to some Θ_∞^{RLS} with

$$\Theta_\infty^{RLS} \,\epsilon\, \{[A, B]^T : A - BK_\infty = A^* - B^* K_\infty\}, \text{ a.s.}$$

for all $[A^*, B^*]^T = \Theta^* \notin \mathcal{N}$ where $\mathcal{N}$ is a Lebesgue null set in $\mathbb{R}^{(n+m)\times n}$ independent of ω.

□

Proof: The RLS algorithm is, in the Bayesian setup of Theorem 3.2, the Kalman filter for $\mathrm{E}[\Theta^* | \mathcal{F}_t^x)] = \widehat{\Theta}_t$. Hence, the conclusion follows by the convergence result of Theorem 3.2 and the mutual absolute continuity of the Lebesgue and the Gaussian measures on $\mathbb{R}^{(n+m)\times n}$.

□

To conclude this section, the next theorem shows that AML estimates posses the same limit set as RLS estimates do:

Theorem 3.4

Consider the system (II.1) with $C = I$ and $[A^*, B^*]$ a stabilizable (deterministic) pair. Suppose that $\Phi_t > 0$ a.s. for all $t \geq t_0$, for some $t_0 < \infty$, and let $\{\theta_t, t \geq t_0\} \,\epsilon\, \mathcal{AML}$ in the sense of (III.2) with $\theta_t \,\epsilon\, \mathcal{S}$ for all $t \,\epsilon\, [t_0, \infty]$, a.s. Then

$\theta_\infty = \lim_{t\to\infty}\theta_t$ exists and is finite, w.p.1, and $\theta_\infty \,\epsilon\, \mathcal{I}$ a.s., where

$$\theta_\infty \,\epsilon\, \mathcal{I} \triangleq \{\theta = co\ell(A,B) : A - BK(\theta) = A^* - B^*K(\theta)\}. \tag{III.6}$$

This holds for all stabilizable pairs $[A^*, B^*]$ outside a Lebesgue null set $\mathcal{N}$ in $\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times m}$. □

The following Theorem, provides an alternative, data-based interpretation to the limit set $\mathcal{I}$, by relating the limit *errors* to the data:

Theorem 3.5

Under the conditions of Theorem 3.4, the limits θ_∞ of $\mathcal{AML}$ estimators $\{\theta_t,\ t \geq t_0\}$, are characterized by the property,

$$\theta_\infty \,\epsilon\, \left\{\theta \,\epsilon\, \mathcal{S} | \theta - \theta^* \,\epsilon\, Ker(\overline{\Phi}(\theta))\right\} = \mathcal{I}(\theta^*) = \mathcal{I}, \tag{III.7}$$

where,

$$\overline{\Phi}(\theta_\infty) = \lim_{t\to\infty}\frac{1}{t}\Phi_t =$$
$$blockdiag(n)\{\begin{bmatrix} I \\ -K(\theta_\infty)\end{bmatrix}\overline{P}(\theta_\infty)[I,\ -K(\theta_\infty)^T]\},$$

and,

$$\begin{aligned}\overline{P}(\theta_\infty) &= \lim_{t\to\infty}\frac{1}{t}\int_0^t x_s x_s^T ds \\ &= \int_0^\infty [expF(\theta_\infty)t]CC^T[expF^T(\theta_\infty)t]dt,\end{aligned}$$

with $blockdiag(n)\{G\}$ being a block-diagonal matrix whose n blocks are the matrices G, and $F(\theta) = A - BK(\theta)$.

IV. Geometric Properties

In this section we examine the geometric characteristics of the limit set $\mathcal{I}$, the set of systems with indistinguishable closed-loop dynamics. As is apparent from Theorem 3.2, standard CE LQ schemes may only lead to parameter estimate convergence into $\mathcal{I}$ and producing suboptimal performance. To deal with this situation, the geometric study in this section provides information which facilitates the construction of an adaptive control scheme which will result in the desired optimal performance.

Let

$$\mathcal{C} = \{\theta \,\epsilon\, \mathcal{S}; K(\theta) = K(\theta^*)\}$$

Since $\theta_t \to \mathcal{I},\ as\ t\to\infty$, one would achieve the optimal performance (II.7) if, further, $\mathcal{I}\subset\mathcal{C}$. However, Polderman [22, 23] showed that

(i) $\mathcal{I}\cap\mathcal{C} = \theta^*$

(ii) $V(\theta) \geq V(\theta^*) \qquad \forall\theta \,\epsilon\, \mathcal{I}.$

(Recall that V is the solution of ARE (II.5).)

We show below that without consistent identification (i.e. $\theta_t \to \theta^*$) only suboptimal performance can be achieved, as, due to (i), by using standard $\mathcal{AML}$ estimates one may encounter

$$\theta_t \not\to \theta^* \tag{IV.1}$$

thus getting $J(\theta_\infty) > J(\theta^*)$ (where, as define in (II.6), $J(\theta)$ is the optimal achievable performance for a system parameterized by θ). We now examine the first order derivatives of V and J on $\mathcal{I}$. We show that, in addition of being the unique minimum of J over $\mathcal{I}, \theta^*$ is also a unique stationary value point of the gradient of J with respect to B (Lemma 4.1 below). Such a result is important as a key tool in establishing the consistency of various gradient and Newton type algorithms, in particular, the consistency of the *recursive* algorithm derived in [17, 19].

First note that for any $\theta \,\epsilon\, \mathcal{I}$, the calculation of $V(\theta)$ can be made either by the ARE (II.5) or by

$$[A^* - B^*K(\theta)]^T V + V[A^* - B^*K(\theta)] + K(\theta)K^T(\theta) + I = 0 \tag{IV.2}$$

where $K(\theta) = B(\theta)^T V(\theta) = B^T V(\theta)$. Note that by (IV.2) $V(\theta) = V(A,B)$ is actually a function of B only (where Definition (III.6) determines the corresponding A for all $\theta \,\epsilon\, \mathcal{I}$).

Let $dJ(\theta)/dB$ be an $n\times m$ matrix whose typical (i,j) entry is $dJ(\theta)/dB_{ij}$. Note that since for all $\theta \,\epsilon\, \mathcal{I},\ J$ is a function of B only, one has,

$$\frac{dJ(\theta)}{dB_{ij}} = \frac{\partial J(\theta)}{\partial B_{ij}} + \sum_{p,q=1}^{n}\frac{\partial A_{pq}}{\partial B_{ij}}\frac{\partial J(\theta)}{\partial A_{pq}}.$$

Lemma 4.1

θ^* is the unique stationary value point of dJ/dB over $\mathcal{I}$, that is,

$$\frac{dJ(\theta)}{dB} = 0 \Leftrightarrow \theta = \theta^*, \quad \theta \,\epsilon\, \mathcal{I}. \tag{IV.3}$$

□

While we skip the lengthy proof, the Lemma is demonstrated below to hold in the scalar case.

For a precise definition of the *full* derivative of J w.r.t. B in (IV.3) let,

$$\nabla J(\theta) = [(\frac{\partial J(\theta)}{\partial co\ell A})^T, (\frac{\partial J(\theta)}{\partial co\ell B})^T]^T. \tag{IV.4}$$

Recall that for any $\theta \,\epsilon\, \mathcal{I}, A = A(B)$ is determined by (III.6). Let $D = D(\theta)$ be an $n^2\times nm$ matrix whose entries are

$$D_{ij}(\theta) = \frac{\partial(co\ell A)_i}{\partial(co\ell B)_j}$$

Hence, the *full* derivative of J w.r.t. B is written as,

$$co\ell\frac{dJ(\theta)}{dB} = D^T(\theta)\frac{\partial J(\theta)}{\partial co\ell A} + \frac{\partial J(\theta)}{\partial co\ell B} \tag{IV.5}$$

Corollary 4.2

$$J(\theta) > J(\theta^*),\ \forall\theta \,\epsilon\, \mathcal{I},\ \theta\neq\theta^*. \tag{IV.6}$$

Proof: Polderman [22, 23] already established that

$$TrV(\theta) = J(\theta) \geq TrV(\theta^*) = J(\theta^*),\ \forall\theta \,\epsilon\, \mathcal{I},\ \theta\neq\theta^*. \tag{IV.7}$$

and further, that $\mathcal{I}$ is a smooth manifold. Hence, together with (IV.3), uniqueness follows. □

The scalar example: With $\theta = (a, b)^T$, the ARE (II.5), rewritten as,

$$2(a - bk(\theta))v(\theta) + k^2(\theta) + 1 = 0,\ k(\theta) = bv(\theta), \quad \text{(IV.8)}$$

leads through differentiation (with $v = J$), to that

$$\frac{\partial J}{\partial b} = -k\frac{\partial J}{\partial a},\ \frac{\partial J}{\partial a} = \frac{-v}{a - bk} > 0,\ \ \forall\theta \,\epsilon\, \mathcal{S}. \quad \text{(IV.9)}$$

Differentiating (with respect to b) the $\mathcal{I}$–defining relation,

$$a - bk(\theta) = a^* - b^*k(\theta), \quad \text{(IV.10)}$$

taking into account (IV.9), results in the following implicit equation for $d = \partial a/\partial b$ ($\theta \,\epsilon\, \mathcal{I}$):

$$d = (b - b^*)b(d - k)\frac{\partial v}{\partial a} + 2k - b^*v. \quad \text{(IV.11)}$$

Equation (IV.5) takes the form,

$$\frac{dJ(\theta)}{db} = \frac{\partial J}{\partial a}(d - k),\ \theta \,\epsilon\, \mathcal{S}. \quad \text{(IV.12)}$$

From here, it is easily verifiable that, on $\mathcal{I}$,

$$\theta = \theta^* \Leftrightarrow d = k \Leftrightarrow \frac{dJ(\theta)}{db} = 0. \quad \text{(IV.13)}$$

As an alternative characterization of $\mathcal{I}$, we look at the direction *orthogonal* to $\mathcal{I}$, obtained through the differentiation of the relation (IV.10) with respect to θ: Rewrite (IV.10),

$$g(\theta) = g(a, b) = a - a^* - k(a, b)(b - b^*) = 0. \quad \text{(IV.14)}$$

Then,

$$\nabla g(a, b) = (\frac{\partial g}{\partial a}, \frac{\partial g}{\partial b})^T, \quad \text{(IV.15)}$$

is given by

$$\frac{\partial g}{\partial a} = 1 + \frac{k}{a - bk}(b - b^*), \quad \text{(IV.16)}$$

$$\frac{\partial g}{\partial b} = -v(b - b^*) - k\big(1 + \frac{k}{a - bk}(b - b^*)\big). \quad \text{(IV.17)}$$

On the other hand, the tangent space to $\mathcal{I}$ (at any $\theta \,\epsilon\, \mathcal{I}$), is defined by $q = (d, 1)^T$ (where $d = d(\theta)$). We now show that,

$$(\nabla g(\theta), q(\theta)) = 0,\ \ \forall\theta \,\epsilon\, \mathcal{I}. \quad \text{(IV.18)}$$

The computation of this inner product is straightforward using (IV.11, IV.15, IV.16, IV.17):

$$\begin{aligned}(\nabla g, q) &= d + \frac{kd}{a - bk}(b - b^*) - v(b - b^*)\\ &\quad -k\big(1 + \frac{k}{a - bk}(b - b^*)\big)\\ &= (d - k)\big(1 + \frac{k}{a - bk}(b - b^*)\big) - v(b - b^*)\\ &= (d - k)\frac{a - b^*k}{a - bk} - (k - vb^*). \quad \text{(IV.19)}\end{aligned}$$

Now, take (IV.11) to write,

$$\begin{aligned}d - k &= (b - b^*)b\frac{\partial v}{\partial a}(d - k) + k - b^*v\\ &= -(b - b^*)\frac{k}{a - bk}(d - k) + k - b^*v. \quad \text{(IV.20)}\end{aligned}$$

From here one has,

$$(d - k)\big(1 + \frac{k}{a - bk}(b - b^*)\big) = k - b^*v, \quad \text{(IV.21)}$$

rewritten as,

$$(d - k)\frac{a - b^*k}{a - bk} = k - b^*v. \quad \text{(IV.22)}$$

Substituting this relation in (IV.19) results in (IV.18).

Remark: Note that by (IV.16, IV.17), $\nabla g(\theta) = (1, -k(\theta))^T$ if and only if $\theta = \theta^*$. It follows that $\nabla g(\theta)$ and $\nabla J(\theta)$ become parallel only at $\theta = \theta^*$. This fact (easily verifiable in the vector–valued parameter case), serves in the construction of a *recursive* scheme, which utilizes a projection of the gradient ∇J on $\mathcal{I}$, a projection which, according to the above, vanishes only at θ^* [17, 19].

V. A Conceptual Adaptive Scheme

The results quoted in the previous section show that for optimal adaptive LQ performance it is necessary to generate consistent parameter estimates. Sufficiency follows from,

Theorem 5.1: Let $\{\widehat{\theta}_t,\ t \geq 0\}$ be a consistent estimate in the sense that

$$\widehat{\theta}_t \to \theta^*,\ a.s.\ as\ t \to \infty,$$

where in addition, $\widehat{\theta}_t \,\epsilon\, \mathcal{S}\ \forall t \geq 0$. Then, with $\{\widehat{\theta}_t,\ t \geq 0\}$ incorporated within an adaptive feedback law of the form,

$$u_t = -K(\widehat{\theta}_t)x_t,$$

the resulting long–run LQ performance is optimal in the sense of (II.7), a.s.

Proof: See Duncan and Pasik-Duncan [6]. □

This leads us to adopt a methodology related to the biased ML approach of Kumar [12] and Borkar [1], (see also Kumar and Becker [10], Kumar and Lin [11], Campi and Kumar [4], Prandini and Campi [21]). The techniques of the aforementioned authors invoke the minimization of a weighted sum of the log-likelihood function and the computed performance J of the controlled system.

In light of Theorem 3.4, the point of view adopted in this paper is that, in the limiting case of an infinite observation sample, the control task is to minimize J over the parameterized system descriptions and parameterized controllers that satisfy the *constraint* given by the vanishing of the gradient of the log-likelihood function.

The conceptual adaptive algorithm is as follows. Suppose that a positive scalar stochastic process $\{\delta_t; t \geq 0\}$ is given with δ_t monotonically decreasing to zero. Then we formulate the adaptive optimization problem as,

$$\begin{aligned}&\text{minimize} \quad J(\theta, \theta),\\ &\text{subject to} \quad \begin{cases}\theta \,\epsilon\, \mathcal{S}\\ \|\nabla L_t(\theta)\| \leq \delta_t,\end{cases}\end{aligned} \quad \text{(V.1)}$$

where, with a slight change of notation, the first parameter, θ, appearing in the function $J(\cdot, \cdot)$ denotes the system

$\Xi(\cdot)$, while the second refers to the parameterization of the feedback control law $K(\theta)$. Hence $J(\theta, \theta')$ denotes the LQ performance of the system $\Xi(\theta)$ subject to the control law $u_t = -K(\theta')x_t$, and we adopt the convention that $\nabla J(\theta, \theta)$ denotes the gradient of $J(\theta, \theta)$ with respect to the θ parameter in both entries.

Lemma 5.2

Suppose that $\Phi_{t_0} > 0$, a.s., for some finite t_0. Assume that a strong solution $\{\theta_t,\ t \geq 0\}$ to (V.1) exists. Then,

$$\theta_t \to \theta_\infty(\omega) \,\epsilon\, \mathcal{I} \quad \text{as } t \to \infty, \text{ a.s.}$$

Proof: As a strong solution exists, it is necessary (by construction) that,

$$\|\nabla L_t(\theta_t)\| \to 0 \quad \text{a.s. } t \to \infty \text{ a.s.}$$

(where $\delta_t \to 0$). Hence, $\{\theta_t\}$ is an $\mathcal{AML}$ estimate and therefore, by Theorem 3.4, converges to a finite limit in $\mathcal{I}$. □

Corollary 5.3

Let the conditions of Lemma 5.2 hold. Then, with probability 1, $\theta_\infty(\omega) = \theta^*$, that is, the estimates $\{\theta_t\}$ are strongly consistent and the associated CE based adaptive scheme is a.s. optimal in the sense of (II.7).

Proof: As a solution to (V.1) is, by the constraint it is subjected to, made to converge to $\mathcal{I}$, it follows that minimization of J is *asymptotically* restricted to $\mathcal{I}$, on which θ^* was shown to be the unique minimizer of J, thereby making θ^* the only possible limit. Optimal performance follows from Theorem 5.1. □

REFERENCES

[1] Borkar, V.S., Self-Tuning Control of Diffusions without the Identifiability Condition. *J. Opt. Theory & Appl*, Vol. 68 (1991) 117-138.

[2] Caines, P.E. *Linear Stochastic Systems* (John Wiley, 1988).

[3] Caines, P.E., Continuous-Time Stochastic Adaptive Control: Non-explosion, ϵ-Consistency and Stability. *Systems & Control Letters* 19 (1992) 169-176.

[4] Campi, M.C. and Kumar, P.R., Adaptive Linear Quadratic Gaussian Control: The Cost–Biased Approach Revisited, *SIAM J. Control and Optimization*, Vol. 36 (1998) 1890-1907.

[5] Chen, H.F. and Guo, L., *Identification and Stochastic 'Adaptive Control* (Birkhöuser, 1991).

[6] Duncan, T.E. and Pasik-Duncan, B., Adaptive Control of Continuous–Time Linear Stochastic Systems, *Mathematics of Control, Signals and Systems* 3 (1990) 45-60.

[7] Duncan, T.E. and Pasik-Duncan, B., A Parameter Estimate Associated with the Adaptive Control of Stochastic Systems. In: *Analysis and Optimization of Systems*, L.N. Control & Inf. Sc. 83 (Springer, 1986) 508-514.

[8] Duncan, T.E. and Pasik-Duncan, B., Some Methods for the Adaptive Control of Continuous Time Linear Stochastic Systems. In: *Topics in Stochastic Systems: Modelling, Estimation and Adaptive Control*, Gerencsér, L. and Caines, P.E., Eds. L.N. Control & Info. Sc. 161 (Springer, 1991).

[9] Duncan, T.E., Guo, L. and Pasik-Duncan, B., Adaptive Continuous–Time Linear Quadratic Gaussian Control. *IEEE Trans. Automatic Control* 44 (1999) 1653-1662.

[10] Kumar, P.R. and Becker, A., A New Family of Optimal Adaptive Controllers for Markov Chains. *IEEE Trans. Automatic Control*, Vol. AC-27 (1982) 137-146.

[11] Kumar, P.R. and Lin, W., Optimal Adaptive Controllers for Unknown Markov Chains. *IEEE Trans. Automatic Control*, Vol. AC-27 (1982) 765-774.

[12] Kumar, P.R., Optimal Adaptive Control of Linear-Quadratic-Gaussian Systems, *SIAM J. Control and Optimization*, Vol. 21 (1983) 163-178.

[13] Kumar, P.R., Convergence of Adaptive Control Schemes Using Least Squares Estimates. *Proceedings of the 28th Conference on Decision and Control* (1989), 727-731.

[14] Lai, T.L. and Wei, C.Z., Least Squares Estimates in Stochastic Regression Models with Application to Identification and Control of Dynamic Systems, *Ann. Stat.* 10 (1982) 154-166.

[15] Lai, T.L. and Wei, C.Z., Asymptotically Efficient Self-Tuning Regulators. *SIAM J. Control Opt.*, Vol. 25 (1987) 466-481.

[16] La Salle, J.P., *The Stability of Dynamical Systems*. CBMS–NSF Regional Conference Series in Applied Mathematics 25 (SIAM, 1976).

[17] Levanony, D. and Caines, P.E., Stochastic Lagrangian Adaptation, McGill University Research Report (2001).

[18] Levanony, D. and Caines, P.E., On Persistent Excitation for Linear Systems with Stochastic Coefficients. *SIAM J. Control and Optimization* Vol. 40 (2001) 882-897.

[19] Levanony, D. and Caines, P.E., Stochastic Lagrangian Adaptive LQG Control. In: *Stochastic Theory and Control*, Pasik–Duncan, B., Ed., L.N. Control & Inf. Sc. 280 (Springer, 2002) 283-300.

[20] Levanony, D., On Persistent Excitation Conditions for the Consistent Filtering of Convergent Semimartingales. *Proceedings of the 43rd Conference on Decision and Control* (2004) 394-399.

[21] Prandini, M. and Campi, M.C., Adaptive LQG Control of Input–Output Systems: A Cost–Biased Approach, *SIAM J. Control and Optimization*, Vol. 39 (2001) 1499-1519.

[22] Polderman, J.W., A Note on the Structure of Two Subsets of the Parameter Space in Adaptive Control Problems. *Systems & Control Letters* 7 (1986) 25-34.

[23] Polderman, J.W., On the Necessity of Identifying the True Parameter in Adaptive LQ Control. *Systems & Control Letters* 8 (1986) 87-91.

[24] Polderman, J.W., Adaptive LQ Control: Conflict Between Identification and Control. *Linear Algebra & Applications* (1989) 219-244.