Journal of Mathematical Systems, Estimation, and Control
Vol. 6, No. 3, 1996, pp. 1–31

On Fixed Gain Recursive Estimation Processes*

László Gerencsér†

Abstract

An important class of fixed gain recursive estimation processes can be approximated by random differential equations, the right hand sides of which are L-mixing random fields. It will be shown that the solution trajectories of these random differential equations follow the trajectories of the corresponding deterministic differential equation obtained by averaging so that the tracking error is majorated by an L-mixing process,the moments of which can be estimated. The result is applied to prove a theorem on the pathwise tracking ability of a time-varying recursive estimation scheme.

Key words: time-varying systems, recursive estimation, stochastic ordinary differential equations, stochastic approximation, L-mixing processes, averaging principle

AMS Subject Classifications: 60H10, 62L20, 93E12

1 Introduction

The purpose of this work is to present a rigorous mathematical analysis of a time-varying recursive parameter estimation scheme, that is suitable for the "identification" of linear stochastic systems. For the time invariant parameter estimation problem this scheme was proposed in [19] and [4] and its usefulness was demonstrated in [20], [2] and [3]. An estimation scheme that is suitable for the tracking of time varying parameters is obtained using fixed gain instead of decreasing gain, typically $1/n$. The first steps

*Received September 20, 1990; received in final form May 9, 1994. Summary appeared in Volume 6, Number 3, 1996.
†This research was supported by the National Foundation of Hungary under contract No. 102071; and by the National Sciences of Engineering Research Council under grant No. 01329.

towards the analysis of this algorithm were presented in [7], in which some classical averaging methods were applied (cf. [5]).

Fixed gain recursive estimators for adaptive filtering were used in [18] using a weak convergence framework. The main result of that paper is a characterization of a fixed gain recursive estimator process by a limit theorem for a sequence of problems with fixed gains which tend to zero. Another aspect of the same problem, namely the rate of convergence of the estimator process was recently considered in [13]. Structured problems of time-varying estimation, i.e. problems where the parameter process is assumed to be the output of a linear stochastic system of known structure, were considered in [21]. The nature of time varying estimators was explored in more detail for off-line identification using exponential forgetting in [6], and was further elaborated in [12] and [16].

The major advance of the present paper is that we characterize the actual estimator process instead of a somewhat artificial limiting process; we give a pathwise characterization of the estimation error rather than an upper bound for its moments; and finally, we do not impose any structure on parameter process.

Fixed gain estimators are of particular interest in at least two important problems of the statistical theory of linear stochastic systems: model selection (cf.[15]) and change point detection (cf.[14]). In both papers a key role is played by what is called fixed gain predictive stochastic complexity (cf.[23, 11]).

The first and major part of this paper is devoted to the analysis of what is called the frozen-parameter system. This is a random differential equation, the right hand side of which is a so-called L-mixing random field. We shall develop an averaging principle, which captures the properties of the tracking error in a fairly precise way. The application of these results to time-varying recursive estimation will be given in Section 3.

To start the technical discussion we introduce some notations and definitions which have been introduced partly in [8]. The set of real numbers will be denoted by $\mathbb{R}$, the p-dimensional Euclidean space will be denoted by $\mathbb{R}^p$ and we write $\mathbb{R}^+ = \{t : t \geq 0\}$.

Let a probability space $(\Omega, \mathcal{F}, P)$ be given, let $D \subset \mathbb{R}^p$ be an open domain and let $(x_t(\theta))$: $\Omega \times \mathbb{R}^+ \times D \to \mathbb{R}^n$ be a stochastic process. Here θ is considered as a parameter. We say that $(x_t(\theta))$ is M-bounded if for all $1 \leq q < \infty$

$$M_q(x) = \sup_{\substack{t\geq 0\\ \theta \in D}} \mathrm{E}^{1/q} |x_t(\theta)|^q < \infty.$$

Here $|\cdot|$ denotes the Euclidean norm. We shall use the same terminology if θ or t degenerate into a single point. Also we shall use the following notation: if $(x_t(\theta))$ is M-bounded, we write $(x_t(\theta)) = O_M(1)$. Moreover, if (c_t) is a positive real-valued function, we write $x_t(\theta) = O_M(c_t)$ if $x_t(\theta)/c_t =$

$O_M(1)$.

A key tool in the analysis of estimator processes is the concept of L-mixing which we are now going to introduce. Let $(\mathcal{F}_s), s \geq 0$ be a family of monotone increasing σ-algebras, and $(\mathcal{F}_s^+), s \geq 0$ be a family of monotone decreasing σ-algebras. We assume that $(\mathcal{F}_s^+)$ is continuous from the right, i.e. $\mathcal{F}_s^+ = \sigma\{\cup_{0<\varepsilon} \mathcal{F}_{s+\varepsilon}^+\}$. Furthermore assume that for all $s \geq 0, \mathcal{F}_s$ and $\mathcal{F}_s^+$ are independent. For $s < 0 \quad \mathcal{F}_s^+ = \mathcal{F}_0^+$.

A stochastic process $(x_t(\theta)), t \geq 0 \quad \theta \epsilon D$ is L-mixing with respect to $(\mathcal{F}_t, \mathcal{F}_t^+)$ uniformly in θ if it is $(\mathcal{F}_t)$-progressively measurable, M-bounded and if we set for $1 \leq q < \infty$

$$\gamma_q(\tau) = \gamma_q(\tau, x) = \sup_{\substack{t \geq \tau \\ \theta \epsilon D}} \mathrm{E}^{1/q} |x_t(\theta) - \mathrm{E}(x_t(\theta) | \mathcal{F}_{t-\tau}^+)|^q \qquad \tau \geq 0,$$

then we have

$$, _q = , _q(x) = \int_0^\infty \gamma_q(\tau) d\tau < \infty.$$

It can be shown that $\gamma_q(\tau)$ is measurable and thus the integral makes sense. If the process does not depend on a parameter, then we define L-mixing in an obvious way. We shall sometimes use the notations $x_{t,s}^+(\theta) = \mathrm{E}(x_t(\theta)|\mathcal{F}_s^+)$, for $s \leq t$ and write $x_t(\theta)$ as $x_t(\theta) = x_{t,s}^+(\theta) + x_{t,s}^0(\theta)$.

The above definitions extend to discrete-time processes in an obvious way. Let $D \subset \mathbb{R}^p$ be an open domain and let the stochastic process $(x_n(\theta)) : \Omega \times \mathbb{Z} \times D \to \mathbb{R}^n$ be a stochastic process. We say that $(x_n(\theta))$ is M-bounded if for all $1 \leq q < \infty$

$$M_q(x) = \sup_{\substack{n \geq 0 \\ \theta \epsilon D}} \mathrm{E}^{1/q} |x_n(\theta)|^q < \infty.$$

Let $(\mathcal{F}_n), n \geq 0$ be a family of monotone increasing σ-algebras, and $(\mathcal{F}_n^+), n \geq 0$ be a monotone decreasing family of σ-algebras. We assume that for all $n \geq 0, \mathcal{F}_n$ and $\mathcal{F}_n^+$ are independent. For $n \leq 0$ we set $\mathcal{F}_n^+ = \mathcal{F}_0^+$. A stochastic process $(x_n(\theta)), n \geq 0$ is L-mixing with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$ uniformly in θ if it is $(\mathcal{F}_n)$-progressively measurable, M-bounded and if we set for $1 \leq q < \infty$

$$\gamma_q(\tau, x) = \gamma_q(\tau) = \sup_{\substack{n \geq \tau \\ \theta \epsilon D}} \mathrm{E}^{1/q} |x_n(\theta) - \mathrm{E}(x_n(\theta) | \mathcal{F}_{n-\tau}^+)|^q$$

where τ is a positive integer then

$$, _q = , _q(x) = \sum_{\tau=1}^{\infty} \gamma_q(\tau) < \infty.$$

Some of the basic results of the theory of L- mixing processes were developed in [8], and a summary of them was given in [12].

To capture the smoothness of a stochastic process $(x_t(\theta))$ with respect to θ we define

$$\Delta x/\Delta\theta = (\Delta x/\Delta\theta)_t(\theta, \theta + h) = |x_t(\theta + h) - x_t(\theta)|/|h|$$

for $t \geq 0, \theta \neq \theta + h\epsilon D$. A stochastic process $(x_t(\theta))$ is M-Lipschitz-continuous in θ if the process $\Delta x/\Delta\theta$ is M-bounded; i.e. if for all $1 \leq q < \infty$, we have

$$M_q(\Delta x/\Delta\theta) = \sup_{\substack{t\geq 0 \\ \theta\neq\theta+h\epsilon D}} \mathrm{E}^{1/q}|x_t(\theta + h) - x_t(\theta)|^q/|h| < \infty.$$

We define , $_q(\Delta x/\Delta\theta)$ in an analogous way. Finally, we introduce the notations

$$M'_q(x) = M_q(x) + M_q(\Delta x/\Delta\theta) \quad \text{and} \quad ,'_q(x) = ,_q(x) + ,_q(\Delta x/\Delta\theta).$$

Now we start the discussion of the subject matter of this paper. Consider the random differential equation

$$\dot{\theta}_t = H(t, \theta_t, \omega) + \delta H(t, \omega) \qquad \theta_s = \xi \qquad s \geq 0 \tag{1.1}$$

where $(H(t, \theta, \omega))$ is a random field defined on the probability space $(\Omega, \mathcal{F}, P)$ for $t \geq 0$ and $\theta\epsilon D$, where D is an open domain in $\mathbb{R}^p$ and $\delta H(t, \omega)$ is a perturbation term. Define the random field $\Delta H/\Delta\theta$ as above by

$$\Delta H/\Delta\theta(t, \theta, \theta + h, \omega) = |H(t, \theta + h, \omega) - H(t, \theta, \omega)|/|h|$$

for $\theta, \theta + h\epsilon D, h \neq 0$. We assume

Condition 1.1 $(H(t, \theta, \omega))$ and $(\Delta H/\Delta\theta(t, \theta, \theta + h, \omega))$ are continuous in t and bounded in (t, θ, ω) and $(t, \theta, \theta + h, \omega)$ respectively, say

$$|H(t, \theta, \omega)| \leq K \quad \text{and} \quad |\Delta H/\Delta\theta(t, \theta, \theta + h, \omega)| \leq L.$$

Condition 1.2 The process $\delta H(t, \omega)$ is a measurable, bounded process, satisfying $|\delta H(t, \omega)| \leq \delta K \leq K$.

This condition ensures that (1.1) has a unique solution which can be continued until θ_t escapes D. To get an averaging principle for random differential equation we need to assume some kind of mixing property of the random field $(H(t, \theta, \omega))$.

Condition 1.3 H and $\Delta H/\Delta\theta$ are L-mixing uniformly in θ for $\theta\epsilon D$ and in $\theta, \theta + h\epsilon D$ respectively, with respect to a pair of families of σ-algebras $(\mathcal{F}_t, \mathcal{F}_t^+)$.

An important step in the study of random differential equations is the development of an averaging method. The concrete form of this method may vary from problem to problem, cf. e.g. the paper of [5] and the references therein. A common feature of these methods is that the solution of the random differential equation is compared to the solution of a deterministic differential equation, the "averaged differential equation." However, instead of exact averaging we consider approximate averaging. Let us consider a decomposition of the expectation of $H(t, \theta_t, \omega)$ into a dominant and a residual term as follows:

$$\mathrm{E}H(t, \theta, \omega) = G(t, \theta) + \delta G(t). \tag{1.2}$$

Then the ordinary differential equation

$$\dot{y}_t = G(t, y_t) \qquad y_s = \xi, \qquad s \geq 0 \tag{1.3}$$

is the approximate averaged differential equation.

Condition 1.4 We assume that $G(t, y)$ is defined on $\mathbb{R}^+ \times D$; it is continuous and bounded in (t, y) together with its first and second partial derivatives as indicated below

$$|G(t, y)| \leq K, \quad \|\partial G(t, y)/\partial y\| \leq L \quad \|\partial^2 G(t, y)/\partial y^2\| \leq L.$$

Here $\|\cdot\|$ denotes the operator norm of a matrix.

This condition ensures the existence and uniqueness of the solution of (1.3), which we denote by $y(t, s, \xi)$ in some finite or infinite time interval. Further details will be provided in Condition 1.6.

Condition 1.5 The perturbation term $\delta G(t)$ satisfies $|\delta G(t)| \leq \delta K$.

Note, that the constants in the upper bounds in Condition 1.4 and 1.5 are identical with the constants in Conditions 1.1 and 1.2. This is no serious restriction and makes the calculations and the interpretation of the results easier.

We associate with the differential equation (1.3) a flow in the phase space in the usual way: for any ξ we define $\phi_{t,s}(\xi) = y(t, s, \xi)$. Let $D_0 \subset D$ be any subset of D such that for some $t, s \geq 0$ we have $\phi_{t,s}(\xi) \epsilon D$ for any $\xi \epsilon D_0$. Then the image of D_0 under $\phi_{t,s}$ will be denoted as $\phi_{t,s}(D_0)$, i.e., we set

$$\phi_{t,s}(D_0) = \{y : y = y(t, s, \xi), \xi \epsilon D_0\}.$$

Assume now that $\phi_{t,s}(D_0)$ can be defined for any $t, s \geq 0$.Then we denote the union of these sets by $\phi(D_0)$; i.e., we set

$$\phi(D_0) = \{y : y = y(t, s, \xi) \quad \text{for some} \quad t \geq s \geq 0\}.$$

The ε neighborhood of the set D_0 will be denoted by $S(D_0, \varepsilon)$ i.e. $S(D_0, \varepsilon) = \{\theta : |\theta - z| < \varepsilon \text{ for some } z \epsilon D_0\}$. The interior of a compact domain D is denoted by $\text{int}D$.

Condition 1.6 There exist compact domains $D_\xi \subset D_y \subset D_\theta \subset D_0 \subset D$ such that we have $\phi(D_\xi) \subset \text{int}D_y$, $S(D_y, d) \subset \text{int}D_\theta$ for some $d > 0$ and $\phi(D_\theta) \subset \text{int}D_0$. (The subscripts indicate the processes, which live in the corresponding domains.)

It is well-known (cf. [22], Ch. 24 Th. 17, or [17] Ch. V Th. 1.1), that $y(t, s, \xi)$ is a continuously differentiable function of (t, s, ξ) and even $(\partial^2/\partial\xi^2)y(t, s, \xi)$ exists and is continuous in (t, s, ξ). We can therefore express exponential asymptotical stability of (1.3) in the following way:

Condition 1.7 For some $c_0 > 0$ $\alpha > 0$ we have for all $0 < s < t, \theta \epsilon D_\theta$

$$\|\frac{\partial y}{\partial \theta}(t, s, \theta)\| \leq c_0 e^{-\alpha(t-s)}.$$

It is no loss of generality to assume that $c_0 > 1$.

Note, that $(\partial/\partial\theta)y(t, s, \theta) = Y(t, s, \theta)$ is the solution of the linear differential equation

$$\dot{Y}(t, s, \theta) = \frac{\partial}{\partial y} G(t, y(t, s, \theta))Y(t, s, \theta) \qquad Y(s, s, \theta) = I. \tag{1.4}$$

It is interesting to note here that exponential asymptotical stability implies under the conditions given above that the second order derivatives $(\partial^2/\partial\theta^2)y(t, s, \theta) = (\partial/\partial\theta)Y(t, s, \theta)$ also decay exponentially in $(t - s)$ (cf. Lemma 4.2 of Appendix).

In the formulation of the theorem below we also need the process $\overline{H}$ defined by

$$\overline{H}(t, \theta, \omega) = H(t, \theta, \omega) - \text{E}H(t, \theta, \omega).$$

We shall introduce constants which depend only on the constants appearing in the conditions above, i.e. K, δK, L, d, c_0, and α. These constants will be called system constants.

Theorem 1.1 *Assume that Conditions 1.1-1.7 are satisfied. Then for $\xi \epsilon D_\xi$, any initial time $s \geq 0$ and sufficiently large d the solution θ_t is defined for all $t \geq s$, $\theta_t \epsilon D_\theta$ and $|\theta_t - y_t| \leq c_0\alpha^{-1} \cdot 4K$. Here d is sufficiently large, if $d > c_0\alpha^{-1}\lambda 4K$. Moreover setting $T = \alpha^{-1}$ we have for small α*

$$\sup_{nT \leq t \leq (n+1)T} |\theta_t - y_t| \leq \delta_n$$

where (δ_n) *is an L-mixing process with respect to* $(\mathcal{F}_{nT}, \mathcal{F}^+_{nT})$ *and we have for any* $2 < q < \infty$ *and* $r > p$

$$M_q(\delta) \le C\alpha^{-1/2}(M'_{qr}(\overline{H}) \cdot ,'_{qr}(\overline{H}))^{1/2} + C\alpha^{-1}\delta K \qquad 1.5$$

$$,_q(\delta) \le C\alpha^{-1/2}(M'_{qr}(\overline{H}), '_{qr}(\overline{H}))^{1/2} + C\alpha^{-1}\delta K + C\,,'_{qr}(\overline{H}) \qquad 1.6$$

where $C = c_2 \cdot exp 30 c_0^4 (1 + L\alpha^{-1})(1 + K\alpha^{-1})$, *and here* c_2 *depends only on* p, q, r, D_θ *and* D.

An important feature of the theorem is that the moments of the tracking error are directly estimated through systems characteristics. On the other hand, the tracking error as a stochastic process has also been characterized to a degree which is quite useful for applications.

It is of interest to check the effect of scaling onto the upper bounding process (δ_n). Let us consider a family of problems parametrized by λ, where say $0 < \lambda \le 1$:

$$\dot\theta^\lambda_t = \lambda(H^\lambda(t, \theta^\lambda_t, \omega) + \delta H^\lambda(t, \omega)) \qquad \theta^\lambda_s = \xi. \qquad (1.7)$$

We shall assume the validity of Conditions 1.1, 1.3, 1.4 and 1.6 uniformly in λ. For the latter condition this means that the domains $D_\xi \subset D_y \subset D_\theta \subset D_0$ and d are independent of λ. The presence of the scaling parameter λ justifies some modifications in connection with Conditions 1.2, 1.5 and 1.7.

Condition 1.2$'$ The process $\delta H^\lambda(t, \omega)$ is a measurable, bounded process, satisfying $|\delta H^\lambda(t, \omega)| \le \lambda \cdot K$, where K is independent of λ.

Condition 1.5$'$ The perturbation term $\delta G^\lambda(t)$ satisfies $|\delta G^\lambda(t)| \le \lambda \cdot K$, where K is independent of λ.

Let us associate with (1.7) the family of ordinary differential equations

$$\dot y^\lambda_t = \lambda G^\lambda(t, y^\lambda_t) \qquad y^\lambda_s = \xi. \qquad (1.8)$$

and let its general solution be $y^\lambda(t, s, \xi)$.

Condition 1.7$'$ For some $c_0 > 0, \alpha > 0$ we have for all $0 < s < t$, $\theta \epsilon D_\theta$, $0 < \lambda \le 1$

$$\left\| \frac{\partial y^\lambda}{\partial \theta}(t, s, \theta) \right\| \le c_0 e^{-\lambda\alpha(t-s)}$$

and here c_0 and α do not depend on λ.

Remark If $G^\lambda(t, y)$ does not depend on t, then Condition 1.7$'$ is a direct consequence of Condition 1.7. On the other hand, for truly time-varying differential equations $\dot y^\lambda_t = \lambda G(t, y^\lambda_t)$ with fixed $G(t, y)$ on the right hand side, Condition 1.7$'$ may not be satisfied: for small λ's the above differential equation may lose stability. This is why there should be some dependence between λ and $G(t,y)$, which is expressed by the superscript λ.

Theorem 1.2 *Assume that the Conditions 1.1,1.3,1.4 and 1.6 are satisfied uniformly in λ,and Conditions 1.2',1.5' and 1.7' are also satisfied. Then for $\xi \epsilon D_\xi$, any initial time s, $0 < \lambda \le 1$ and for sufficiently large d, θ^λ_t is defined for all $t \ge s$, $\theta^\lambda_t \epsilon D_\theta$ and $|\theta^\lambda_t - y^\lambda_t| \le c_0 \alpha^{-1} 4K$ and we have*

$$|\theta^\lambda_t - y^\lambda_t| \le \delta^\lambda_{ct}$$

where (δ^λ_{ct}) is an L-mixing process with respect to $(\mathcal{F}_t, \mathcal{F}^+_t)$ and for any $q \ge 1$

$$M_q(\delta^\lambda_c) \le c\lambda^{1/2}, \qquad \textit{and} \qquad , _q(\delta^\lambda_c) \le c\lambda^{-1/2}$$

where the constant c is independent of λ.

Note that the essential supremum for the absolute value of the deviation does not decrease with λ. Using Theorem 1.2 we can derive the following result, which gives a pathwise characterization of the process (θ^λ_t):

Corollary 1.3 *Let $(F(\theta))$ be a continuously differentiable function of θ defined in D. Then we have*

$$\limsup_{T\to\infty} \frac{1}{T}\int_0^T |F(\theta^\lambda_t) - F(y^\lambda_t)| dt \le c\lambda^{1/2}$$

with probability 1, where c is a deterministic constant, which is independent of λ.

In the application we present in Section 3 we shall need a discrete time version of Theorem 1.2. Let us consider a discrete-time random field $H^\lambda(n,\theta,\omega)$ and a discrete-time random process $\delta H^\lambda(n,\omega)$. We will consider processes defined by

$$\theta^\lambda(n+1) = \theta^\lambda(n) + \lambda(H^\lambda((n+1),\theta^\lambda(n),\omega) + \delta H^\lambda((n+1),\omega)). \quad (1.10)$$

We define a continuous time extension of the correction terms as follows: set for $n \le t < n+1$ $H^\lambda(t,\theta,\omega) = H^\lambda(n+1,\theta,\omega)$ and $\delta H^\lambda(t,\omega) = \delta H^\lambda(n+1,\omega)$. We define $\mathcal{F}_t = \mathcal{F}_{n+1}$, $\mathcal{F}^+_t = \mathcal{F}^+_{n+1}$. Let the functions $G^\lambda(n,\theta), \delta G^\lambda(n)$ satisfy

$$\mathrm{E} H^\lambda(n,\theta,\omega) = G^\lambda(n,\theta) + \delta G^\lambda(n),$$

and let $G^\lambda(t,\theta)$ be the the piecewise constant extension of $G^\lambda(n,\theta)$, defined by $G^\lambda(t,\theta) = G^\lambda(n,\theta)$ for $n \le t < n+1$. We define $\delta G^\lambda(t,\theta)$ analogously. Then the associated differential equation is defined as in (1.8).

Theorem 1.4 *Assume that H^λ, δH^λ, G^λ, δG^λ satisfy the conditions of Theorem 1.2, and θ_n^λ is generated by (1.10). Then replacing θ_t^λ by θ_n^λ the conclusion of Theorem 1.2 hold.*

2 The Proofs

First we prove Theorem 1.1 in several steps.

Step 1 First we show that $\theta_t \epsilon D_\theta$ for all t.

Assume $s = 0$ and apply Lemma 4.1 of the Appendix to express $\theta_t - y_t$. Note that along the trajectory θ_t we have $H + \delta H - G = H - \mathrm{E}H + H - G = \overline{H} + \delta H + \delta G$. Thus by Lemma 4.1 we get that as long as θ_t does not leave D_θ we have

$$\theta_t - y_t = \int_0^t \frac{\partial}{\partial \xi} y(t, s, \theta_s)(\overline{H}(s, \theta_s, \omega) + \delta H(s, \omega) + \delta G(s))ds. \qquad (2.1)$$

Taking into account the stability condition (Condition 1.7) and the inequalities $|\overline{H}| \leq 2K, |\delta H| \leq K, |\delta G| \leq K$, we get

$$|\theta_t - y_t| \leq \int_0^t c_0 e^{-\alpha(t-s)} \cdot 4K ds \leq c_0 \cdot 4K\alpha^{-1}.$$

Thus if K is sufficiently small, then $|\theta_t - y_t|$ will always be smaller than the distance between D_y and D_θ^c, where D_θ^c denotes the complement of the set D_θ; hence θ_t will stay in D_θ for all t. Thus Step 1 is complete and the first proposition of the theorem is proved.

To prepare Step 2 we need to introduce some notations. Let us subdivide the positive real line into intervals of length T. In the interval $(n, (n+1)T)$ we consider the solution trajectory of (1.3) starting from $\theta(nT)$ at time nT, say $\overline{y}_t$; i.e., $\overline{y}_t$ is defined by

$$\dot{\overline{y}}_t = G(t, \overline{y}_t), \qquad \overline{y}(nT) = \theta(nT).$$

We shall give an estimate of $|\theta_t - \overline{y}_t|$ in $nT \leq t < (n+1)T$, which is much sharper than what we obtained at the beginning of Step 1.

Step 2 We shall prove that for $T = \alpha^{-1}$ we have

$$\sup_{nT \leq t \leq (n+1)T} |\theta_t - \overline{y}_t| \leq c^*(\eta_n^* + \delta\eta^*),$$

where (η_n^*) is defined in terms of the random field $H = (H(s, y, \omega))$ along deterministic trajectories as follows:

$$\eta_n^* = \sup_{\substack{nT \leq t \leq (n+1)T \\ \theta \varepsilon D_\theta}} \left| \int_{nT}^t (\partial/\partial\xi) y((n+1)T, s, y(s, nT, \theta)) \cdot \overline{H}(s, y(s, nT, \theta), \omega) ds \right| \qquad (2.2)$$

and $\delta\eta^*$ is a constant given by

$$\delta\eta^* = 2c_0\alpha^{-1} \cdot \delta K. \tag{2.3}$$

Furthermore,

$$c^* = \exp 4c_0^3(1 + L\alpha^{-1})(1 + K\alpha^{-1}). \tag{2.4}$$

c is a system constant.

For the proof we first note that Condition 1.6 ensures that $\overline{y}_t \epsilon D_0$ for all $t \geq 0$. There is no loss of generality to assume that $n = 0$, and then we get from (2.1)

$$|\theta_t - \overline{y}_t| \leq |\int_0^t \left(\frac{\partial}{\partial\xi} y(t, s, \overline{y}_s)\right)\left(\overline{H}(s, \overline{y}_s, \omega) + \delta H(s, \omega) + \delta G(s)\right) ds| + \\ + r_1 + r_2 + r_3 \tag{2.5}$$

where

$$r_1 = |\int_0^t (\frac{\partial}{\partial\xi} y(t, s, \overline{y}_s)(\overline{H}(s, \theta_s, \omega) - \overline{H}(s, \overline{y}_s, \omega))ds|$$

$$r_2 = |\int_0^t \left(\frac{\partial}{\partial\xi} y(t, s, \theta_s) - \frac{\partial}{\partial\xi} y(t, s, \overline{y}_s)\right)\overline{H}(s, \theta_s, \omega)ds|$$

and

$$r_3 = |\int_0^t \left(\frac{\partial}{\partial\xi} y(t, s, \theta_s) - \frac{\partial}{\partial\xi} y(t, s, \overline{y}_s)\right)(\delta H(s, \omega) + \delta G(s))ds|.$$

Since $\|(\partial/\partial\xi)y(t, s, \xi)\| \leq c_0$, and $\overline{H}$ is Lipschitz-continuous in θ with Lipschitz-constant, L we get that

$$r_1 \leq \int_0^t c_0 L|\theta_s - \overline{y}_s| ds.$$

On the other hand, it can be shown (cf. Lemma 4.2 the of Appendix) that Conditions 1.4 - 1.7 imply

$$\|\frac{\partial^2}{\partial\xi^2} y(t, s, \xi)\| \leq c_0^3 L\alpha^{-1};$$

hence $(\partial/\partial\xi)y(t, s, \xi)$ is Lipshitz-continuous in ξ with Lipschitz-constant $c_0^3 L\alpha^{-1}$ and we thus get

$$r_2 \leq \int_0^t c_0^3 L\alpha^{-1}|\theta_s - \overline{y}_s| 2K ds.$$

Similarly,

$$r_3 \leq \int_0^t c_0^3 L\alpha^{-1}|\theta_s - \overline{y}_s| 2K ds.$$

Thus

$$r_1 + r_2 + r_3 \leq \int_0^T (c_0 L + c_0^3 L\alpha^{-1} 4K) \cdot |\theta_s - \overline{y}_s| ds = c_1 \int_0^T |\theta_s - \overline{y}_s| ds \quad (2.6)$$

with

$$c_1 = c_0 L + c_0^3 L\alpha^{-1} \cdot 4K. \quad (2.7)$$

Thus we can write (2.5) as

$$|\theta_t - \overline{y}_t| \leq | \int_0^t (\frac{\partial}{\partial \xi} y(t,s,\overline{y}_s)\Big(\overline{H}(s,\overline{y}_s,\omega) + \delta H(s,\omega) + \delta G(s)\Big) ds| +$$

$$+ \int_0^t c_1 |\theta_s - \overline{y}_s| ds. \quad (2.8)$$

To estimate the first term on the right hand side note that $(\partial/\partial\xi)y(t,s,\xi)$ is the solution of the variational equation (1.4). Write $Y_t = Y(t,0,\xi)$. Thus the first term on the right hand side of (2.8) can be written as

$$|Y_t Y_T^{-1} \int_0^t Y_T Y_s^{-1} \Big(\overline{H}(s,\overline{y}_s,\omega) + \delta H(s,\omega) + \delta G(s)\Big) ds|. \quad (2.9)$$

Here $\|Y_t\| \leq c_0$ by Condition 1.7, and $\|Y_T^{-1}\| \leq \exp LT$ since Y_t^{-1} satisfies the linear differential equation

$$\dot{Y}_t^{-1} = -\frac{\partial}{\partial y} G(t,y_t) \cdot Y_t^{-1} \qquad Y^{-1}(0) = I,$$

and here the norm of the coefficient matrix is bounded by L. (We get the estimate $\|Y_T^{-1}\| \leq \exp LT$ writing the above differential equation as an integral equation, using the inequality $\|(\partial/\partial y)G(t,y)\| \leq L$ and applying the Bellman-Gronwall-lemma.) Thus

$$\|Y_t Y_T^{-1}\| \leq c_0 \exp LT. \quad (2.10)$$

Write the integral term in (2.9) as the sum of three integral and take supremum over $0 \leq t \leq T$ and over the initial condition $\theta = \theta(0)$ which enters implicitly through $\overline{y}_s$. For the first integral we get the random variable defined in (2.2) with $n = 0$.

The contribution of the second and third integrals will be majorated by

$$\int_0^T c_0 e^{-\alpha(T-s)}(2\delta K) \leq 2c_0\alpha^{-1} \cdot \delta K = \delta\eta^*.$$

Thus the first term in (2.8) is majorated by $c_0 \exp LT \cdot (\eta^*(0) + \delta\eta^*)$ and thus applying the Bellman-Gronwall lemma (cf.(2.6)), we get that

$$\sup_{0 \le t \le T} |\theta_t - \overline{y}_t| \le c'(\eta^*(0) + \delta\eta^*)$$

with $c' = \exp c_1 T \cdot c_0 \exp(LT)$. We will find a simple upper bound for c' to see its dependence on the system constants in a more transparent manner. Using the definition of c_1 (cf. (2.7)) and the inequality $c_0 < e^{c_0}$, we get

$$c' \le \exp(c_0 LT + 4L\alpha^{-1} {c_0}^3 KT + c_0 + LT).$$

Now since $c_0 > 1$ we can majorate the expression in the bracket by $c_0^3 LT + 4Lc_0^3 \alpha^{-1} KT + c_0^3 + c_0^3 LT$. Taking out c_0^3 and substituting $T = \alpha^{-1}$ it is easy to see that the last expression is majorated by $4c_0^3(1 + L\alpha^{-1})(1 + K\alpha^{-1})$ which is exactly the exponent in c^* as given in (2.4). Thus we get

$$\sup_{0 \le t \le T} |\theta_t - \overline{y}_t| \le c^*(\eta^*(0) + \delta\eta^*). \tag{2.11}$$

The general case, i.e. $n > 0$ is handled the same way; thus Step 2 is completed.

Step 3 We shall prove that the process (η_n^*) is $L-$mixing with respect to $(\mathcal{F}_{nT}, \mathcal{F}_{nT}^+)$ and for any $2 < q < \infty$, and $r > p$ we get

$$M_q(\eta^*) \le c_3 c_2 T^{1/2} (M'_{qr}(\overline{H}), \Gamma'_{qr}(\overline{H}))^{1/2} \tag{2.12}$$

and

$$\Gamma_q(\eta^*) \le c_3 c_2 \Gamma'_{qr}(\overline{H}) \tag{2.13}$$

where c_2 depends only on q, p, r and the domains D_θ and D and c_3 is defined by

$$c_3 = c_0^4(1 + L\alpha^{-1}). \tag{2.14}$$

First we show that the process $u_s(\theta) = \overline{H}(s, y(s, nT, \theta), \omega)$ for $nT \le s < (n+1)T$ is L-mixing with respect to $(\mathcal{F}_s, \mathcal{F}_s^+)$, uniformly in θ for $\theta \epsilon D_\theta$ and we have for any $1 \le q < \infty$ $M_q(u) \le M_q(\overline{H})$, $\Gamma_q(u) \le \Gamma_q(\overline{H})$. To see this we prove the following more general statement:

Lemma 2.1 *Let $(v(t, y))$ $y \epsilon D$ be a stochastic process which is an L-mixing process uniformly in y with respect to a family of σ-algebras $(\mathcal{F}_t, \mathcal{F}_t^+), t \ge 0$. Here $D \subset \mathbb{R}^p$ is an open domain. Assume that $(y_t), t \ge 0$ is a measurable deterministic function taking values in D. Then the process $u_t = v(t, y_t)$ is L-mixing with respect to $(\mathcal{F}_t, \mathcal{F}_t^+)$ and we have for any $m \ge 1$*

$$M_m(u) \le M_m(v) \qquad \text{and} \qquad \Gamma_m(u) \le \Gamma_m(v).$$

Proof: The first inequality is trivial. To prove the second inequality note that for $t \geq s > 0$ we have

$$u_t - E(u_t|\mathcal{F}_s^+) = (v_t(y) - E(v_t(y)|\mathcal{F}_s^+))_{|y=y_t},$$

since y_t is deterministic. Hence for any $q \geq 1$

$$E^{1/q}|u_t - E(u_t|\mathcal{F}_s^+)|^{1/q} = E^{1/q}|v_t(y) - E(v_t(y)|\mathcal{F}_s^+)|^q_{y=(y_t)} \leq \gamma_q(t-s, v)$$

and thus the second inequality follows.

The next lemma shows that L-mixing is preserved under the operations by which (η_n^*) is generated from H.

Lemma 2.2 *Let $u_t(\theta), t \geq 0, \theta \epsilon D$ be a separable vector-valued stochastic process such that both u and $\Delta u/\Delta\theta = (u_t(\theta+h) - u_t(\theta))/h$ are uniformly L mixing in $\theta, \theta + h\epsilon D$, with respect to $(\mathcal{F}_t, \mathcal{F}_t^+)$. Assume that $\mathrm{E}u_t(\theta) = 0$ for all $t \geq 0, \theta\epsilon D$ and let $D_\theta \subset D$ be a compact domain. For fixed $T > 0$ we define a discrete time process (u_n^{**}) by*

$$u_n^{**} = \sup_{\substack{nT \leq t < (n+1)T \\ \theta\epsilon D_\theta}} \int_{nT}^t h_s(\theta) u_s(\theta) ds,$$

where $h_s(\theta)$ is a deterministic measurable and bounded matrix-valued function of (s, θ). Let us assume, e.g., that $\|h(s,\theta)\| \leq k$. Moreover, $h(s,\theta)$ is Lipschitz-continuous in θ, say $\|h(s,\theta+d) - h(s,\theta)\| \leq l|d|$. Then (u_n^{**}) is L-mixing with respect to $(\mathcal{F}_{nT}, \mathcal{F}_{nT}^+)$ and for any $q > 2$ and any $r > p$, we have

$$M_q(u^{**}) \leq c_2(k+l)T^{1/2}(M'_{q_r}(u) \cdot ,'_{q_r}(u))^{1/2}, \quad ,_q(u^{**}) \leq c_2(k+l),'_{q_r}(u)$$

where c_2 depends only on q, p, r and the domains D_θ and D.

Proof: We assume $h_s(\theta) = 1$. The general case is dealt with similarly. The inequalities given as Theorems 1.1 and 5.1 imply that

$$u_n^*(\theta) = \sup_{nT \leq t < (n+1)T} \int_{nT}^t u_s(\theta) ds$$

is M-bounded and for $q > 2$

$$M_q(u^*(\theta)) \leq c_q T^{1/2} M_q^{1/2}(u(\theta)),\ ^{1/2}_q(u(\theta)).$$

Let $m < n$ and approximate $u_n^*(\theta)$ by

$$u_{n,m}^{*+}(\theta) = \sup_{nT \leq t < (n+1)T} \int_{nT}^t u_{s,mT}^+(\theta) ds$$

where $u^+_{s,mT}(\theta) = \mathrm{E}(u_s(\theta)|\mathcal{F}^+_{mT})$. We have for any fixed $\theta \epsilon D$

$$|u^*_n(\theta) - u^{*+}_{n,m}(\theta)| \leq \int_{nT}^{(n+1)T} |u_s(\theta) - u^{*+}_{s,mT}(\theta)| ds$$

and taking the $L_q(\Omega, F, P)$ norm of both sides for some $q \geq 1$ we get

$$\begin{aligned} E^{1/q}|u^*_n(\theta) - u^{*+}_{s,m}(\theta)|^q &\leq \int_{nT}^{(n+1)T} \gamma_q(s - mT, u(\theta)) ds \\ &= \int_{\tau T}^{(\tau+1)T} \gamma_q(s, u(\theta)) ds, \end{aligned}$$

where $u(\theta)$ denotes the stochastic process $(u_t(\theta))$ for fixed θ. Fix $\tau = n - m$ and take supremum over n for $n \geq \tau$ to get

$$\gamma_q(\tau, u^*(\theta) \leq \int_{\tau T}^{(\tau+1)T} \gamma_q(s, u(\theta)) ds.$$

Summation over τ from 1 to ∞ gives $,_q(u^*(\theta)) \leq 2,_q(u(\theta))$.

Now the same argument can be applied for the process $(\Delta u^*_n / \Delta\theta)$ defined as follows: $(\Delta u^*_n/\Delta\theta)_n(\theta, \theta + h) = (u^*_n(\theta + h) - u^*_n(\theta))/|h|$. The application of the maximal inequality given in [8] as Theorem 3.4 completes the proof.

Let us now apply Lemma 2.2 with $h_s(\theta) = (\partial/\partial\xi) y((n+1)T, s, y(s, nT, \theta))$ and $u_s(\theta) = \overline{H}(s, y(s, nT, \theta), \omega)$. We have $\|h(s, \theta)\| \leq c_0$ and

$$\begin{aligned} \|\partial h^i(s,\theta)/\partial\theta\| &= \|(\partial^2/\partial\xi^2) y^i((n+1)T, s, y(s, nT, \theta)).\partial y(s, nT, \theta)/\partial\theta\| \\ &\leq Lc_0^3 \alpha^{-1} \cdot c_0. \end{aligned}$$

Thus the constant $k + l$ in Lemma 2.2 now becomes $c_0 + Lc_0^4\alpha^{-1}$ which is majorated by c_3 and the proposition of this step follows.

In the following step we shall estimate the distance between the piecewise continuous trajectory $(\bar{y}_t)$ constructed above and the solution trajectory of (1.3), which we denote by (y_t). For this we consider the difference $(\theta_{rT} - \overline{y}_{rT_-})$ and compute how it is propagated by the flow corresponding to (1.3).

Step 4 We have

$$\sup_{nT \leq t < (n+1)T} |\theta_t - y_t| \leq \delta_n \tag{2.15}$$

where

$$\delta_n = \sup_{0 \leq m \leq n} \left(3c_0 \sum_{r=m}^{n} e^{-(n-r)} c^* \eta^*_r \right) + 3c_0 \sum_{r=0}^{n} e^{-(n-r)} c^* \delta\eta^*. \tag{2.16}$$

Note that here δ_n is independent of the initial time $s \leq t$ and the initial value $\theta_s = y_s = \xi \epsilon D_\xi$.

To prove the proposition note that we have for $nT \leq t < (n+1)T$,

$$\overline{y}_t - y_t = \sum_{r=1}^{n} \int_0^1 \frac{\partial}{\partial\theta} y(t, rT, z(r,\lambda)) \cdot (\theta_{rT} - \overline{y}_{rT_-}) d\lambda$$

where $\overline{y}_{rT_-}$ denotes the left hand-side limit of $\overline{y}_t$ at $t = rT$, and $z(r,\lambda) = \lambda\theta_{rT} + (1-\lambda)\overline{y}_{rT_-}$. Using the stability condition imposed onto $\dot{y} = G(t,y)$ and (2.6), we get with $T = \alpha^{-1}$

$$|\overline{y}_t - y_t| \leq \sum_{r=1}^{n} c_0 e^{-(n-r)} |\theta_{rT} - \overline{y}_{rT_-}| \leq c_0 \sum_{r=1}^{n} e^{-(n-r)} c^* (\eta_r^* + \delta\eta^*). \quad (2.17)$$

Finally, using (2.6) and the inequality $|\theta_t - y_t| \leq |\theta_t - \overline{y}_t| + |\overline{y}_t - y_t|$ in $nT \leq t < (n+1)T$, we get

$$\sup_{nT \leq t < (n+1)T} |\theta_t - y_t| \leq c_0 \sum_{r=1}^{n} e^{-(n-r)} c^* (\eta_{r-1}^* + \delta\eta^*) + (\eta_n^* + \delta\eta^*)$$

$$\leq c_0 e \sum_{r=0}^{n} e^{-(n-r)} c^* (\eta_r^* + \delta\eta^*). \quad 2.18$$

It is easy to see that the same argument applies if the initial time s is not equal to 0 say $mT \leq s < (m+1)T$. Then the upper bound on the right hand side of (2.18) has to be modified so that the summation starts at $r = m$. Also we majorate $c_0 e$ by $3c_0$ and thus the desired result is obtained.

Step 5 It will now be shown that the process (δ_n) is L-mixing with respect to $(\mathcal{F}_{nT}, \mathcal{F}_{nT}^+)$ and we have

$$M_q(\delta) \leq c_4(M_q(\eta^*) + \delta\eta^*), \quad , _q(\delta) \leq c_4(M_q(\eta^*) + \delta\eta^*) + c_4, \ _q(\eta^*) \quad (2.19)$$

where

$$c_4 = 24 c_0 c^*. \quad (2.20)$$

Lemma 2.3 *Let* (u_n), $n = 0, 1 \cdots$ *be an* L*-mixing process with respect to a pair of families of* σ*-algebras* $(\mathcal{F}_n, \mathcal{F}_n^+)$. *Define the process* (v_n) *by*

$$v_n = \sup_{0 \leq m \leq n} \sum_{r=m}^{n} \rho^{n-r} u_r$$

with $|\rho| < 1$. *Then* (v_n) *is L-mixing with respect to* $(\mathcal{F}_n, \mathcal{F}_n^+)$ *and we have for* $q \geq 1$

$$M_q(v) \leq (1-\rho)^{-1} M_q(u), \quad , {}_q(v) \leq 2(1-\rho)^{-2} M_q(u) + 2(1-\rho)^{-1}, {}_q(u). \tag{2.21}$$

Proof: We have

$$|v_n| \leq \sum_{r=0}^{n} \rho^{n-r} |u_r|$$

from which the first part of (2.21) follows by the triangle inequality for the $L_q(\Omega, \mathcal{F}, P)$ norm. Let now τ be a fixed positive integer and approximate v_n by

$$v^*_{n,n-\tau} = \sup_{n-\tau \leq m \leq n} \sum_{r=m}^{n} \rho^{n-r} u^+_{r,n-\tau}$$

where $u^+_{r,n-\tau} = \mathrm{E}(u_r | \mathcal{F}^+_{n-\tau})$. Then obviously $v^*_{n,n-\tau}$ is $\mathcal{F}^+_{n-\tau}$ measurable. Furthermore, it is easy to see that

$$|v_n - v^*_{n,n-\tau}| \leq \sum_{r=0}^{n-\tau-1} \rho^{n-r} |u_r| + \sum_{r=n-\tau}^{n} \rho^{n-r} |u^0_{r,n-\tau}|$$

where $u^0_{r,n-\tau} = u_r - u^+_{r,n-\tau}$. Taking the $L_q(\Omega, \mathcal{F}, P)$ norm of the right hand side, we get

$$\begin{aligned} E^{1/q} |v_n - v^*_{n,n-\tau}|^q &\leq \rho^{n-r} M_q(u) + (r * \gamma_q)(\tau) \\ &\leq (1-\rho)^{-1} \rho^{\tau+1} M_q(u) + (r * \gamma_q)(\tau) \end{aligned}$$

where $*$ denotes discrete time convolution which now is applied to the series $r = (\rho^{\tau-1})_{\tau=1}^{\infty}$ and $(\gamma_q(\tau, u))_{\tau=1}^{\infty}$. Applying Lemma 2.1 of [8] gives

$$\gamma_q(\tau, v) \leq 2(1-\rho)^{-1} \rho^{\tau+1} M_q(u) + 2(r * \gamma)(\tau).$$

Let us perform summation over τ form 1 to ∞. The contribution of the first term on the right hand side is $2(1-\rho)^{-2} \rho M_q(u) \leq 2(1-\rho)^{-2} M_q(u)$. For the second term we apply the inequality

$$\sum_{\tau=1}^{\infty} (r * \gamma_q)(\tau) \leq \sum_{\tau=1}^{\infty} (\rho^{\tau-1}) \cdot \sum_{\tau=1}^{\infty} \gamma_q(\tau, u) = (1-\rho)^{-1}, {}_q(u).$$

Thus the proof of the lemma is complete.

Apply the lemma above with the process $u_{1r} = 3c_0 c^* \eta^*_r$ and the deterministic process $u_{2r} = 3c_0 c^* \delta \eta^*_r$ which appear on the definition of δ_n in (2.16) and with $\rho = e^{-1}$. Then $(1-\rho)^{-1} \leq (1-1/2)^{-1} = 2$, hence the largest constant multiplier in (2.21), which is $2(1-\rho)^{-2}$ is majorated by

8. Since (u_{2r}) is deterministic, we have $, (u_2) = 0$. Thus we immediately get the proposition of Step 5.

To complete the proof of Theorem 1.1 we have to put the estimates together. For the sake of convenience we summarize the relevant inequalities (2.15), (2.19), (2.12), (2.13) and (2.3), which connect the various processes that we considered and (2.20), (2.4), and (2.14), which define the various constants:

$$\sup_{nT \leq t < (n+1)T} |\theta_t - y_t| \leq \delta_n$$

$$\begin{array}{rcll} M_q(\delta) & \leq & c_4(M_q(\eta^*) + \delta\eta^*), & , _q(\delta) \leq c_4(M_q(\eta^*) + \delta\eta^*) + c_4 , _q(\eta^*), \\ M_q(\eta^*) & \leq & c_3 c_2 T^{1/2}(M'_{qr}(\overline{H}) , '_{qr}(\overline{H}))^{1/2}, & , _q(\eta^*) \leq c_3 c_2 , '_{qr}(\overline{H}). \end{array}$$

$$\delta\eta^* = 2c_0\alpha^{-1} \cdot \delta K.$$

For the constants we have

$$c_4 = 24c_0c^*, \quad c^* = \exp 4c_0^3(1 + L\alpha^{-1})(1 + K\alpha^{-1}) \quad c_3 = c_0^4(1 + L\alpha^{-1})$$

and c_2 depends only on q, p, r and the domains D_θ and D. Combining the above inequalities we get for $q > 2$,

$$\begin{array}{rcl} M_q(\delta) & \leq & c_4 c_3 c_2 T^{1/2}(M'_{qr}(\overline{H}) \cdot , '_{qr}(\overline{H}))^{1/2} + c_4 \cdot (2c_0\alpha^{-1} \cdot \delta K); \\ , _q(\delta) & \leq & c_4 c_3 c_2 T^{1/2}(M'_{qr}(\overline{H}) \cdot , '_{qr}(\overline{H}))^{1/2} \\ & & + c_4 \cdot (2c_0\alpha^{-1} \cdot \delta K) + c_4 c_3 c_2 , '_{qr}(\overline{H}). \end{array}$$

We give a simple upper bound for $c_4c_3c_2$. Using the inequality $c < e^c$ with $c = 24c_0$ and the definition of c^*, we get

$$c_4 \leq \exp 24c_0 \cdot \exp 4c_0^3(1+L\alpha^{-1})(1+K\alpha^{-1}) = \exp 28c_0^4(1+L\alpha^{-1})(1+K\alpha^{-1}).$$

Taking into account the definition of c_3, we find that $c_3 \leq \exp c_0^4(1 + L\alpha^{-1})(1 + K\alpha^{-1})$ and thus $c_4c_3 \leq \exp 30c_0^4(1 + L\alpha^{-1})(1 + K\alpha^{-1})$, which if multiplied by c_2 is exactly the constant appearing in the theorem. Finally, to majorate the expression $c_4 \cdot (2c_0\alpha^{-1} \cdot \delta K)$, we use the inequality

$$2c_0 \leq \exp 2c_0 \leq \exp 2c_0^4(1 + L\alpha^{-1})(1 + K\alpha^{-1}),$$

which if multiplied by the upper bound given for c_4 above and assuming $c_2 \geq 1$ gives that $c_4 \cdot 2c_0 \leq C$, where C is the constant given in the theorem. Thus the proof of Theorem 1.1 is complete.

Proof of Theorem 1.2: Theorem 1.2 follows from Theorem 1.1. Conditions 1.2',1.5' and 1.7' imply that the right hand sides of (1.5) and (1.6) are majorated by $C\lambda^{1/2}$. Indeed, in the present case, $\overline{H}$ becomes $\lambda\overline{H}^\lambda$, α

becomes $\lambda\alpha$ and δK becomes $\lambda^2\delta K$. From the above scaling property it follows that C is independent of the parameter λ. Applying Theorem 1.1 with $T=(\lambda\alpha)^{-1}$, we get that

$$\sup_{nT\le t\le (n+1)T} |\theta_t^\lambda - y_t^\lambda| \le \delta_n^\lambda.$$

Now if the piecewise constant extension of (δ_n^λ) is denoted by (δ_{ct}^λ), i.e. $\delta_{ct}^\lambda=\delta_n$ for $n\le t<n+1$, then (δ_{ct}^λ) is L-mixing with respect to $(\mathcal{F}_t,\mathcal{F}_t^+)$. Obviously, we have $M_q(\delta_c^\lambda)=M_q(\delta^\lambda)\le c\lambda^{1/2}$. Furthermore,

$$,\ {}_q(\delta_c^\lambda)\le 2T,\ {}_q(\delta^\lambda)=2(\lambda\alpha)^{-1},\ {}_q(\delta^\lambda) \tag{2.23}$$

by Lemma 2.2 in [8], and the right hand side is majorated by $2\alpha^{-1}c\lambda^{-1/2}$; thus Theorem 1.2 follows.

Proof of Corollary 1.3: First note that the process $|F(\theta_t^\lambda)-F(y_t^\lambda)|$ is $L-$mixing. Furthermore, since F is continuously differentiable in D and $\theta_t^\lambda, y_t^\lambda \epsilon D_\theta$, we have $|F(\theta_t^\lambda)-F(y_t^\lambda)|dt \le c|\theta_t^\lambda - y_t^\lambda| \le c\lambda^{1/2}$ with some $c>0$, and hence $\mathrm{E}|F(\theta_t^\lambda)-F(y_t^\lambda)|\le c\lambda^{1/2}$. Applying a strong law of large numbers given as Corollary 1.3 in [8], we get the proposition.

Proof of Theorem 1.4: Let θ_t^λ denote the piecewise linear extension of θ_n, defined for $n\le t<n+1$ by $\theta_t^\lambda=(1-(t-n))\theta_n^\lambda+(t-n)\theta_{n+1}^\lambda$ for $n\le t<n+1$. Then we can write

$$\dot\theta_t^\lambda=\lambda(H^\lambda(t,\theta_t^\lambda,\omega)+\delta H^\lambda(t,\omega))+\lambda(H^\lambda(n+1,\theta_n^\lambda,\omega)-H^\lambda(t,\theta_t^\lambda,\omega)). \tag{2.24}$$

Introduce the notation

$$\delta H_0^\lambda(t,\omega)=(H^\lambda(n+1,\theta^\lambda(n),\omega)-H^\lambda(n+1,\theta^\lambda(t),\omega)$$

and

$$\delta H_1^\lambda(t,\omega)=\delta H^\lambda(t,\omega)+\delta H_0^\lambda(t,\omega).$$

Then (2.24) can be written as

$$\dot\theta_t^\lambda=\lambda(H^\lambda(t,\theta_t^\lambda,\omega)+\delta H_1^\lambda(t,\omega)). \tag{2.25}$$

By the conditions of Theorem 1.4, all conditions of Theorem 1.2 except Condition 1.2′ are satisfied for (2.25). To verify this condition, note that by Condition 1.1, $|\theta_t^\lambda-\theta_n^\lambda|\le\lambda K$, and thus $|\delta H_1^\lambda(t,\omega)|\le\lambda KL$, and the validity of Condition 1.2′ follows.

3 A Time Varying Estimation Scheme

In this section we consider the general estimation scheme proposed in [19] and [4]. It has been shown in [20] that this general scheme is suitable for the description of a large class of recursive identification methods and also some special adaptive control methods, such as the Åström–Wittenmark self-tuning regulator (cf. [1]). Because of its wide applicability and its mathematical elegance, this scheme deserves some attention.

Here we consider only one of the possible variants of this general estimation scheme. For this we consider the following parameter-dependent state space equation in which the parameters are a true system-parameter θ^* and a probe parameter θ:

$$\overline{\overline{x}}_{n+1}(\theta,\theta^*) = A(\theta,\theta^*)\overline{\overline{x}}_n(\theta,\theta^*) + B(\theta,\theta^*)e_n \tag{3.1}$$

with $\theta, \theta^* \epsilon D \subset \mathbb{R}^p$, $\overline{\overline{x}}_n(\theta) \epsilon \mathbb{R}^m$, where D is a bounded open domain. The initial condition for this equation is 0, i.e., we set $\overline{\overline{x}}_0(\theta) = 0$.

Condition 3.1 It is assumed that $A(\theta,\theta^*)$ is uniformly stable for $(\theta,\theta^*)\epsilon D \times D$, i.e. $\|A^n(\theta,\theta^*)\| \leq ca^n$, with some $c > 0$ and $0 < a < 1$, for all $(\theta,\theta^*)\epsilon D$. Moreover the matrix-valued functions $(A(\theta,\theta^*))$ and $(B(\theta,\theta^*))$ and their derivatives are bounded for $(\theta,\theta^*)\epsilon D \times D$.

In many important applications the true system parameter and the probe value may have different values, different dimensions and different interpretations. For instance, in system identification the model class we choose may be simpler; i.e., we may misspecify our model. Then the true system parameter is different from the true parameter of the model. Or the probe values may be directly related to a controller, so again, the value for the first parameter is related to the true system parameter in an indirect manner. However, such extensions pose no technical difficulties to modify the derivations below.

The driving noise process may consist of components of the system noise, the observation noise and in the case of adaptive control of a dither, which we inject into the system. There are various standard assumptions of the noise process. Here we adopt one, which has been found particularly convenient for deriving strong results.

Condition 3.2 The driving noise process $(e_n), n \geq 0$ is a second order stationary process which is L-mixing with respect to a pair of families of σ-algebras $(\mathcal{F}_n, \mathcal{F}_n^+)$. Moreover, assume that the input noise process is bounded, say $|e_n| \leq \kappa$.

To determine θ^* we consider certain cross-covariances of the components of $\overline{\overline{x}}_n(\theta,\theta^*)$. For this purpose let Q be a p-dimensional, vector-valued quadratic function of the state vector $\overline{\overline{x}}_n$, and define

$$G(\theta,\theta^*) = \lim_{n\to\infty} \mathrm{E}Q(\overline{\overline{x}}_n(\theta,\theta^*)).$$

The function $G(\theta,\theta^*)$ is a well-defined function of (θ,θ^*), and it is bounded together with all its derivatives. The covariances are selected so that the nonlinear algebraic equation

$$G(\theta, \theta^*) = 0 \tag{3.2}$$

is solved by θ^*.

Also it is assumed that each component of $Q(\overline{\overline{x}}_n(\theta, \theta^*))$ is empirically computable. Thus in principle each component of $G(\theta, \theta^*)$ can be computed for any fixed θ, although for the exact value to obtain an infinite number of observations is needed. A major problem of the theory of stochastic systems is the determination of θ^* from empirical data. This is achieved in the time invariant case by the general recursive estimation scheme given in [19] and [4]. The time-varying estimation problem is the subject of this section and the general scheme will be given below in (3.7) and (3.8).

A simple method of describing a time-varying problem is to replace the system parameter θ^* in (3.1) by a time-varying parameter θ_n^*. Thus we get the following state space equation:

$$\overline{x}_{n+1}(\theta) = A(\theta, \theta_n^*)\overline{x}_n(\theta, \theta_n^*) + B(\theta, \theta_n^*)e_n \tag{3.3}$$

with initial condition $\overline{x}_0(\theta) = 0$. An important restriction is, though, that the system is slowly changing, which will be expressed by the following condition.

Condition 3.3 We assume that $\theta_n^* \epsilon D_\theta^*$, where D_θ^* is a compact subset of D and

$$\dot{S} \stackrel{\Delta}{=} \sup_{n \geq 0} |\theta_{n+1}^* - \theta_n^*| < \infty.$$

The quantity $\dot{S}$ is called the rate of change.

The above description of slowly time-varying systems is a special case of a more general definition in [25]. Time varying systems with rapid changes and structural constraints are considered in [24]. The purpose of this section is to present the design and analysis of a time-varying estimation scheme and establish a relationship between the tracking error and the rate of change.

First we consider a time-invariant ordinary differential equation associated with our estimation problem. We assume the validity of the following condition:

Condition 3.4 We assume that (3.2) has a unique solution $\theta = \theta^*$ with respect to θ in the domain D, and the Jacobian matrix $G_\theta(\theta^*, \theta^*)$ is nonsingular and stable. The solution of the ordinary differential equation

$$\dot{y}_t = G(y_t, \theta^*) \qquad y(0) = \theta \tag{3.4}$$

belongs to a compact domain D_0 whenever $\theta \epsilon D_\theta$ where D_θ is a compact domain such that $D_\theta \subset D$. Moreover, the differential equation (3.4) is globally asymptotically stable in θ^* whenever $\theta^* \epsilon D_\theta^*$ and $y_0 \epsilon D_\theta$, where D_θ^* is the compact domain introduced in Condition 3.3.

It is easy to see using standard compactness arguments that Condition 3.4 implies the validity of Condition 1.7 uniformly in θ^* for $\theta^* \epsilon D_\theta^*$. More exactly let the general solution of (3.4) be denoted by $y(t,s,\theta,\theta^*)$. Then

$$\left\| \frac{\partial y}{\partial \theta}(t,s,\theta,\theta^*) \right\| \leq c_0 e^{-\alpha(t-s)}$$

and here c_0 and α do not depend on θ^*.

Obviously, if we multiply the right hand side of (3.4) by a fixed positive gain, say λ, i.e., and if we consider the differential equation

$$\dot{y}_t = \lambda G(y_t, \theta^*) \qquad y(0) = \theta,$$

then the solution trajectories get scaled but remain unchanged otherwise. Denoting the solutions by $y^\lambda(t,s,\theta,\theta^*)$, we get

$$\left\| \frac{\partial y^\lambda}{\partial \theta}(t,s,\theta,\theta^*) \right\| \leq c_0 e^{-\lambda\alpha(t-s)}.$$

A conceptual algorithm for estimating θ_n^* is given by the recursion

$$\theta_{n+1}^\lambda = \theta_n^\lambda + \lambda Q(\overline{x}_{n+1}(\theta_n^\lambda)). \tag{3.5}$$

The practical problem with this conceptual algorithm is that the correction term $Q(\overline{x}_{n+1}(\theta_n^\lambda))$ is not given by an explicit expression and is not computable recursively. In the final step of the construction this deficiency is removed as follows: let an initial value $\theta_0^\lambda \epsilon D_\xi$ be given, where D_ξ is a compact domain such that $D_\xi \subset D$. Furthermore let an initial auxiliary state vector x_0 be given,which is assumed to be an $\mathcal{F}_0$-measurable, bounded random variable. Generate a sequence of state vectors x_n for $1 \leq n \leq n_0$ by the equation (cf. (3.3))

$$x_{n+1} = A(\theta_0^\lambda, \theta_n^*) x_n(\theta_0^\lambda, \theta_n^*) + B(\theta_0^\lambda, \theta_n^*) e_n. \tag{3.6}$$

The purpose of this initial phase is to get a good approximation of $\overline{x}_n(\theta_0^\lambda)$ for the initial time $n = n_0$. Then for $n \geq n_0$ generate the auxiliary state vector x_n and the estimator θ_n^λ of θ_n^* recursively by the following equations:

$$x_{n+1} = A(\theta_n^\lambda, \theta_n^*) x_n + B(\theta_n^\lambda, \theta_n^*) e_n \tag{3.7}$$

$$\theta_{n+1}^\lambda = \theta_n^\lambda + \lambda Q(x_{n+1}). \tag{3.8}$$

In the theorem and below we will say that c is a system constant, if it depends only on the apriori constants and the domains given above, but it is independent of the sequence (θ_n^*). Thus particular system constants are independent of $\dot{S}$ or λ.

Theorem 3.1 *Assume that Conditions 3.1-3.4 are satisfied and let the compact domains $D_\xi \subset D_y \subset D_\theta \subset D_0 \subset D$ be such that they satisfy Condition 1.6 with respect to (3.4) for all $\theta^* \epsilon D_{\theta^*}$. Let $\theta_0^\lambda \epsilon D_\xi$. If $\dot{S}/\lambda$ is sufficiently small, d is sufficiently large, and n_0 is sufficiently large, then θ_n^λ will not leave D_θ and we have*

$$|\theta_n^\lambda - y_n^\lambda| \leq \delta_n^\lambda$$

where (δ_n^λ) is an L-mixing process such that for any $q \geq 1$ we have

$$M_q(\delta^\lambda) \leq c\lambda^{1/2} \qquad \text{and} \qquad ,\ _q(\delta^\lambda) \leq c\lambda^{-1/2},$$

where c is a system constant.

This theorem combined with the first part of Theorem A below gives a decomposition of an upper bound of the tracking error into a random and a deterministic component plus a negligible term.

Corollary 3.2 *Under the conditions of Theorem 3.1 we have*

$$|\theta_n^\lambda - \theta_n^*| \leq \delta_n^\lambda + c\dot{S}/\lambda + c'e^{-\lambda\alpha n}$$

where the process (δ_n^λ) is identical with the one given in the quoted theorem and c, c' are system constants.

The right hand side will be temporally denoted by δ_n^*. Obviously, the process (δ_n^*) is L-mixing and $,\ _q(\delta^*) = ,\ _q(\delta)$ for every $q \geq 1$. On the other hand, we have $M_q(\delta^*) \leq c(\lambda^{1/2} + \dot{S}/\lambda)$ with some system constant c. Now it is easy to see that this upper bound will be minimized for $\lambda = (2\dot{S})^{2/3}$. Substituting this optimizing value into the formulas of Theorem 3.1, we arrive at the following corollary:

Corollary 3.3 *Under the conditions of Theorem 3.1 choose λ so that we have $c_1\dot{S}^{2/3} \leq \lambda \leq c_2\dot{S}^{2/3}$ where c_1, c_2 are positive system constants. Then*

$$|\theta_n^\lambda - \theta_n^*| \leq \delta_n^* + c'e^{-\lambda\alpha_n}$$

where (δ_n^*) is an L-mixing process such that for any $q \geq 1$ we have

$$M_q(\delta^*) \leq c\dot{S}^{1/3} \qquad \text{and} \qquad ,\ _q(\delta^*) \leq c\dot{S}^{-1/3},$$

where c is a system constant.

Finally we get an analogy with Corollary 1.3 the following proposition:

Corollary 3.4 *Let $(F(\theta))$ be a continuously differentiable function of θ defined in D. Then under the conditions of Corollary 3.3 we have*

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{0}^{N} |F(\theta_n^\lambda) - F(\theta_n^*)| \le c\dot{S}^{1/3}$$

with probability 1, where c is a system constant.

Proof of Theorem 3.1: Let us introduce the notations:

$$H(n,\theta,\omega) = Q(\overline{x}_n(\theta))$$

$$\delta H(n+1,\omega) = Q(x_{n+1}) - Q(\overline{x}_{n+1}(\theta_n^\lambda)). \tag{3.9}$$

Then we can write (3.8) as

$$\theta_{n+1}^\lambda = \theta_n^\lambda + \lambda Q(\overline{x}_{n+1}(\theta_n^\lambda)) + \lambda\delta H(n+1).$$

We shall verify that H and δH satisfy the conditions of Section 1.

Lemma 3.3 *If $\dot{S}$ is sufficiently small, then the random fields $H = (H(n,\theta,\omega))$ and $\Delta H/\Delta\theta = ((H(n,\theta+h,\omega) - H(n,\theta,\omega))/|h|)$ are bounded and*

$$|H(n,\theta)| \le c\kappa^2, \quad |\Delta H/\Delta\theta(n,\theta+h,\theta)| \le c\kappa^2,$$

where c is a system constant. Thus the random field H satisfies Condition 1.1 uniformly in (θ_n^*).

Proof: It is well-known that if $\dot{S}$ is sufficiently small, say $\dot{S} \le \dot{S}_0$, then the time-varying linear system (3.3) is exponentially stable in the following sense: we have for $0 \le n \le m$ and

$$\overline{\Psi}(n,m) = \prod_{i=m}^{n} A(\theta,\theta_n^*)$$

the inequality

$$\|\overline{\Psi}(n,m)\| \le ca^{(n-m)} \tag{3.10}$$

with some $0 < a < 1$ and $c > 0$.

Since the input process (e_n) is bounded in absolute value by κ, it follows that $|\overline{x}_n(\theta)| \le c\kappa$, where c is a system's constant. Since Q is quadratic, we get the first claim of the lemma.

On the other hand, the process $\partial\overline{x}/\partial\theta = \overline{x}_\theta$ satisfies the difference equation

$$\overline{x}_{\theta,n+1}(\theta) = A(\theta,\theta_n^*)\overline{x}_{\theta,n}(\theta) + A_\theta(\theta,\theta_n^*)\overline{x}_n(\theta) + B_\theta(\theta,\theta_n^*)e_n \tag{3.11}$$

in which the second and third terms on the right-hand side are to be considered as input processes. Since they are bounded in absolute value by $C\kappa$, it follows that $|\overline{x}_{\theta n}(\theta)| \leq c\kappa$ where c is a system's constant. Now

$$\Delta H/\Delta\theta(n, \theta+h, \theta) = \int_0^1 Q_x(\overline{x}_n(\theta(\mu))d\mu \cdot \overline{x}(\theta(\mu))$$

where $\theta(\mu) = \theta + \mu \cdot h$. Since $\overline{x}_{\theta n}(\theta)$ is bounded in absolute value by $C\kappa$ uniformly in θ and $Q_x(x)$ grows linearly in x, the second estimation of the lemma follows.

The verification of Condition 1.2′, which is the least obvious task, is left to the end of the section.

Lemma 3.4 *If $\dot{S}$ is sufficiently small, then the random fields $H = (H(n,\theta,\omega))$ and $\Delta H/\Delta\theta = ((H(n,\theta+h,\omega) - H(n,\theta,\omega))/|h|)$ are L-mixing with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$, uniformly in $\theta \epsilon D_\theta$ and $\theta, \theta+h \epsilon D_\theta$, respectively. Thus the random field H satisfies Condition 1.3.*

Proof: We have seen in the proof of the previous lemma that the time varying system (3.3) is exponentially stable, if $\dot{S}$ is sufficiently small. Therefore Lemma 2.4 in [8] implies that the random field $(\overline{x}_n(\theta))$ is L-mixing and it is easy to see that $(\overline{x}_n(\theta))$ is L-mixing uniformly in θ for $\theta \epsilon D_\theta$. The argument for the random field $\Delta\overline{x}/\Delta\theta$ is analogous.

Let us now consider the random field $H(n,\theta) = Q(\overline{x}_n(\theta))$. The properties of $\overline{x}_n(\theta)$ that were derived above are inherited by $H(n,\theta)$ since Q is quadratic (cf. the remark after Definition 1.3); i.e., H and $\Delta H/\Delta\theta$ are L-mixing, and thus the lemma has been proved.

Let us define

$$G(n+1, \theta) = G(\theta, \theta_n^*).$$

Then the piecewise constant extension of $G(n+1,\theta)$ can be written as $G_c(t,\theta) = G(n+1,\theta) = G(\theta,\theta_n^*)$ for $n \leq t < n+1$.

It is easy to see that Condition 1.4 is satisfied with $K = c\kappa^2$ and $L = c\kappa^2$, where c is a system's constant. We proceed to verify Condition 1.5′. We note that

$$\delta G(n) = EH(n,\theta,\omega) - G(n,\theta) = EH(n,\theta,\omega) - G(\theta, \theta_{n-1}^*).$$

Lemma 3.5 *If $\dot{S}$ is sufficiently small, then we have for sufficiently large n_0*

$$|\delta G(n)| = |EH(n,\theta,\omega) - G(\theta,\theta_{n-1}^*)| \leq c\dot{S}\kappa^2$$

where c is a system constant. Thus, if $\dot{S}/\lambda$ is bounded, then Condition 1.5' is satisfied.

Proof: Note that the process $\overline{\overline{x}}_{n+1}(\theta,\theta_n^*)$ satisfies the following difference equation :

$$\overline{\overline{x}}_{n+1}(\theta,\theta_n^*) = A(\theta,\theta_n^*)\overline{\overline{x}}_n(\theta,\theta_{n-1}^*)+$$
$$+B(\theta,\theta_n^*)e_n + A(\theta,\theta_n^*)\cdot(\overline{\overline{x}}_n(\theta,\theta_n^*) - \overline{\overline{x}}_n(\theta,\theta_{n-1}^*)). \qquad 3.12$$

Let us denote the second term on the right hand side of (3.12) by r_n; i.e. set $r_n = A(\theta,\theta_n^*)\cdot(\overline{\overline{x}}_n(\theta,\theta_n^*) - \overline{\overline{x}}_n(\theta,\theta_{n-1}^*))$. Then we have $|r_n| \leq c\dot{S}\kappa$ where c is a system constant.

Subtract (3.12) from (3.3). Then the term $B(\theta,\theta_n^*)e_n$ falls out and introduces the process $\delta\overline{x}_n \stackrel{\Delta}{=} \overline{x}_n(\theta) - \overline{\overline{x}}_n(\theta,\theta_{n-1}^*)$. We see that this process satisfies the simple difference equation

$$\delta\overline{x}_{n+1} = A(\theta,\theta_n^*)\delta\overline{x}_n + r_n. \qquad (3.13)$$

Now if $\dot{S} \leq \dot{S}_0$ then the time-varying linear system (3.13) is exponentially stable, hence a trivial input-output estimate gives $|\delta\overline{x}_{n+1}| \leq c\dot{S}\kappa$ for sufficiently large n_0. Now

$$|Q(\overline{x}_n(\theta)) - Q(\overline{\overline{x}}_n(\theta,\theta_{n-1}^*))| = |\int_0^1 Q_x(x(\mu))\cdot\delta\overline{x}_n d\mu| \qquad (3.14)$$

where $x(\mu) = \overline{x}_n(\theta) + \mu(\overline{\overline{x}}_n(\theta,\theta_{n-1}^*) - \overline{x}_n(\theta))$. Since Q is quadratic and $|\overline{x}_n(\theta)|$ and $|\overline{\overline{x}}_n(\theta,\theta_{n-1}^*)|$ are bounded by $C\kappa$, we get

$$\int_0^1 Q_x(x(\mu))\cdot\delta\overline{x}_n d\mu \leq C\dot{S}\kappa^2.$$

Using the trivial inequality $|\mathrm{E}\xi - \mathrm{E}\eta| \leq \mathrm{E}|\xi-\eta|$ for the random variables inside the absolute value sign of the left hand side of (3.14) and noting that $\mathrm{E}Q(\overline{\overline{x}}_n(\theta,\theta_{n-1}^*)) = G(\theta,\theta_{n-1}^*) + ca^n\kappa$, where c, a are system constants, the proposition of the lemma follows.

The associated ordinary differential equation is defined as follows:

$$\dot{y}_t^\lambda = \lambda G(y_t^\lambda,\theta_t^*) \qquad (3.15)$$

where $\theta_t^* = \theta_n^*$ for $n \leq t < n+1$ and λ is a fixed, small positive gain. Define the initial conditions for (3.15) as $y_s^\lambda = \theta\epsilon D_\theta$, and let the general solution of (3.15) be denoted by $y^\lambda(t,s,\theta)$.

The following theorem has been proved in [9]:

Theorem A *Assume that the differential equation (3.4) satisfies Conditions 3.4. For any parameter process (θ_n^*) satisfying Condition 3.3 choose λ so that $\dot{S}/\lambda$ is sufficiently small. Then the solution of (3.15) is defined for all $t \geq s$, and*

$$|y^\lambda(t,s,\theta) - y^\lambda(t,s,\theta,\theta_t^*)| \leq c\dot{S}/\lambda.$$

Moreover, (3.15) is exponentially stable: for any $0 < \alpha' < \alpha$ *there exist system constants* c, c'*, such that if* $\dot{S}/\lambda \le c$ *then we have*

$$\|\frac{\partial}{\partial \theta} y^{\lambda}(t,s,\theta)\| \le c' e^{-\lambda\alpha'(t-s)} \tag{3.16}$$

uniformly in θ *for* $\theta \epsilon D_{\theta}$*.*

Since $|y^{\lambda}(t,s,\theta,\theta^*_t) - \theta^*_t| \le ce^{-\lambda\alpha(t-s)}$ *where* c *is independent of* λ*, it follows that* $y^{\lambda}(t,s,\theta)$ *tracks* θ^*_t *with an error not exceeding in absolute value* $c\dot{S}/\lambda + ce^{-\lambda\alpha(t-s)}$*.*

The above theorem also implies the validity of Condition 1.7$'$. Furthermore, using this theorem it is easy to see that if we are given a set of domains $D_{\xi} \subset D_y \subset D_{\theta} \subset D_0 \subset D$ which satisfy Condition 1.6 in relation of the differential equation (3.4) for all $\theta^* \epsilon D_{\theta^*}$, then Condition 1.6 is satisfied with the same domains in relation to (3.15) whenever $\dot{S}/\lambda$ is sufficiently small.

Finally, we turn to the verification of Condition 1.2$'$. Note that we had by (3.9) $\delta H(n+1) = Q(x_{n+1}) - Q(\overline{x}_{n+1}(\theta^{\lambda}_n))$. We have to prove

$$|\delta H(n)| \le \lambda \cdot K \tag{3.17}$$

where K is independent of (θ^*_n) and λ. For this purpose we prove a stronger statement in the proof of which the interaction of the dynamics of x_n and θ^{λ}_n is more explicitly analyzed.

Let the largest system constant above be c. Choose a system constant c'_3 and n_0 so that

$$c(1-a)^{-1}\kappa < \frac{1}{2}c'_3\kappa \qquad \text{and} \qquad ca^{N+1}|x_0| < \frac{1}{2}c'_3\kappa$$

hold. Define $c_3 = c(c'_3)^2$. Furthermore, choose a system constant c_4 and n_0 so that

$$c(1-a)^{-1}c\kappa \cdot c_3\kappa^2\lambda < \frac{1}{2}c_4\kappa\lambda \qquad \text{and} \qquad ca^{n_0}|x_0| < \frac{1}{2}c_4\kappa\lambda$$

hold.

Lemma 3.6 *Under the conditions of Theorem 3.1 we have for sufficiently small* $\dot{S}\lambda$ *and* $\dot{S}/\lambda$ *and the system constants above* c_3 *and* c_4*:*

i. *The sequence of estimators* θ^{λ}_n *defined by (3.8) will be in* D_{θ} *for* $n \ge n_0$*.*

ii. *We have* $|\theta^{\lambda}_n - \theta^{\lambda}_{n-1}| \le c_3\kappa^2\lambda \le \dot{S}_0$ *for all* $n \ge n_0$*.*

iii. $|x_n - \overline{x}_n(\theta^\lambda_{n-1})| \le c_4 \kappa \lambda$ *for* $n \ge n_0$.

Obviously iii. implies the validity of Condition 1.2′ since by (3.9)

$$|\delta H(n)| \le c\kappa |x_n - \overline{x}_n(\theta^\lambda_{n-1})| \le c' \kappa c_4 \kappa \lambda.$$

Proof: We use induction on n. For $n = n_0$, i. and ii. are obviously satisfied. To verify iii. note that in the first phase of the algorithm, x_n and $\overline{x}_n(\theta^\lambda_0)$ are generated by the same linear system except that the initial condition at $n = 0$ is different. Therefore, taking into account (3.10) (with $\theta = \theta^\lambda_0$) we get that $|x_n - \overline{x}_n(\theta^\lambda_0)| \le ca^n |x_0|$. Thus since $ca^{n_0}|x_0| \le c_4\kappa\lambda$ large, then iii. is satisfied.

Assume that the propositions are true for $n \le N$. To prove the validity of iii. for $n = N+1$ we first note that if both θ^λ_n and θ^*_n belong to D_θ and they are slowly time varying, then the time varying linear system (3.7) is exponentially stable. To quantify this we assume that $\dot{S}_0$ is chosen so that if $|\theta^*_n - \theta^*_{n-1}| \le \dot{S}_0$ and $|\theta^\lambda_n - \theta^\lambda_{n-1}| \le \dot{S}_0$, then we have for $0 \le n \le m$ and

$$\Psi(n,m) = \prod_{i=m}^{n} A(\theta^\lambda_n, \theta^*_n)$$

the inequality

$$\|\Psi(n,m)\| \le ca^{(n-m)} \tag{3.18}$$

with some system constants $0 < a < 1$ and $c > 0$. Under the inductive hypothesis (3.18) is valid for $n_0 \le n \le N$.

Consider the state vectors x_{n+1} generated by (3.7) and the state vectors $\overline{x}^\lambda_{n+1}(\theta)$ with $\theta = \theta_n$ generated by the "frozen parameter" system (3.3). Subtracting (3.3) from (3.7) we get

$$\begin{aligned} x_{n+1} - \overline{x}_{n+1}(\theta^\lambda_n) = A(\theta^\lambda_n, \theta^*_n)(x_n - \overline{x}_n(\theta^\lambda_{n-1})) - \\ - A(\theta^\lambda_n, \theta^*_n)(\overline{x}_n(\theta^\lambda_n) - \overline{x}_n(\theta^\lambda_{n-1})). \end{aligned} \tag{3.19}$$

This is a difference equation for the variable $(x_n - \overline{x}_n(\theta^\lambda_{n-1}))$ with initial condition x_0 at $n = 0$. Let us interpret (3.19) as a linear system the input process of which is $A(\theta^\lambda_n, \theta^*_n)(\overline{x}_n(\theta^\lambda_n) - \overline{x}_n(\theta^\lambda_{n-1}))$. Since by the proof of Lemma 3.3 $|\overline{x}_{\theta n}(\theta)| \le c\kappa$, uniformly in θ for $\theta \epsilon D_\theta$, we have for $n \le N$

$$|\overline{x}_n(\theta^\lambda_n) - \overline{x}_n(\theta^\lambda_{n-1})| \le c\kappa |\theta^\lambda_n - \theta^\lambda_{n-1}| \le c\kappa \cdot c_3 \kappa^2 \lambda$$

where c is a system constant. Solving (3.19) by a discrete time Cauchy-formula for $(x_{n+1} - \overline{x}_{n+1}(\theta^\lambda_n))$ and using part i. and ii. of the inductive hypothesis we get for $n = N+1$

$$|x_{N+1} - \overline{x}_{N+1}(\theta^\lambda_N)| \le c(1-a)^{-1} c\kappa \cdot c_3 \kappa^2 \lambda + ca^{N+1}|x_0| \le c_4 \kappa \lambda.$$

Thus the validity of iii. for $n = N+1$ has been established.

To prove the validity of ii. for $n = N+1$ note that the the validity of (i) and (ii) for $n \leq N$ implies the exponential stability of the time varying system (3.7) and hence $|x_{N+1}| \leq ca^{N+1}|x_0| + c(1-a)^{-1}\kappa \leq c_3'\kappa$. Thus we have for $N \geq n_0$

$$|\theta_{N+1}^{\lambda} - \theta_N^{\lambda}| \leq |\lambda Q(x_{N+1})| \leq \lambda c |x_{N+1}|^2 \leq c(c_3'\kappa)^2\lambda = c_3\kappa^2\lambda. \tag{3.20}$$

Hence ii. is satisfied for $n = N+1$.

Finally, to prove i. for $n = N+1$ note that $|H(n,\theta,\omega)| \leq c\kappa^2$ by Lemma 3.3 and the established validity of iii. for $n \leq N+1$ implies $|\delta H(n)| \leq c\kappa^2\lambda \leq c\kappa^2$ for $n \leq N+1$. Now the first part of Theorem 1.4 applied for a finite horizontal implies that $|\theta_n^{\lambda} - y_n^{\lambda}| \leq c_0\alpha^{-1}4 \cdot c\kappa^2$ for $n \leq N+1$ and therefore if d is sufficiently large then $\theta_{N+1}^{\lambda} \epsilon D_{\theta}$, and thus the proof of the lemma is complete.

4 Appendix: Two Analytical Lemmas

Lemma 4.1 (cf. [5]) *Let $(G(t,y))$ be a function satisfying Condition 1.6 and let y_t the solution of (1.2). Further, let (x_t) be a continuously differentiable curve such that $x_s = y_s = \xi$. Then for $t > s$*

$$x_t - y_t = \int_s^t \frac{\partial}{\partial \xi} y(t,r,x_r)(\dot{x}_r - G(t,x_r)dr. \tag{4.1}$$

Proof: Consider the function

$$z_r = y(t,r,x_r).$$

Obviously the left hand side of (4.1) can be written as $z_t - z_s$. Write

$$\begin{aligned} z_t - z_s \quad &= \quad \int_s^t z_r' dr = \int_s^t (\frac{\partial}{\partial r} y(t,r,x_r) + \\ &\quad + \frac{\partial}{\partial \xi} y(t,r,x_r)\dot{x}_r)dr. \end{aligned}$$

Taking into account the equality

$$\frac{\partial}{\partial r} y(t,r,x_r) = -\frac{\partial}{\partial \xi} y(t,r,x_r) \cdot G(t,x_r)$$

(which simply follows from the identity $y(t,r,y_r) = y_t =$ const. with respect to r after differentiation) we get the lemma.

Let

$$Y(t,s,\xi) = \frac{\partial}{\partial \xi} y(t,s,\xi).$$

We have the following lemma:

Lemma 4.2 *Under Conditions 1.4 - 1.7 we have*

$$\|\frac{\partial}{\partial \xi} Y(t,s,\xi)\| \le L c_0^3 \alpha^{-1} e^{-\alpha(t-s)}.$$

Proof: We have

$$\frac{\partial}{\partial t} Y(t,s,\xi) = G_x(t, y(t,s,\xi)) Y(t,s,\xi), \qquad Y(s,s,\xi) = I. \tag{4.2}$$

Since $G_{xx}(t,x) = (\partial^2/\partial x^2)G(t,x)$ and $(\partial/\partial\xi)y(t,s,\xi)$ exist and are continuous in (t,x) and (t,ξ) respectively, we conclude that $Y_\xi(t,s,\xi) = (\partial/\partial\xi)Y(t,s,\xi)$ exists and is a continuous function of (t,ξ). To get $Y_\xi(t,s,\xi)$ we can differentiate (4.2) formally and obtain

$$\begin{aligned} \frac{\partial}{\partial t} Y_\xi(t,s,\xi) &= G_{xx}(t, y(t,s,\xi)) Y(t,s,\xi) Y(t,s,\xi) + \\ &\quad + G_x(t, y(t,s,\xi)) Y_\xi(t,s,\xi) \qquad Y_\xi(s,s,\xi) = 0. \end{aligned}$$

Since the operator norm of the first term is majorated by $Lc_0^2 e^{-2\alpha(t-s)}$ and since the time varying linear differential equation with transition matrix $G_x(t,y)$ is exponentially stable as indicated by Condition 1.7, we get from the identity

$$\int_0^t e^{-\alpha(t-r)} e^{-2\alpha r} dr = e^{-\alpha t} \int_0^t e^{-\alpha r} dr < \alpha^{-1} e^{-\alpha t}$$

the desired upper bound.

Acknowledgement

The author wishes to thank the NSF and the University of Kansas for their financial support to present this paper at the Workshop on Stochastic Theory and Adaptive Control that was held at the University of Kansas, 26-28 September 1991; and to Zsuzsanna Vágó for her help in preparing this paper.

References

[1] K.J. Åström and B. Wittenmark. On self-tuning regulators, *Automatica* **9** (1973), 185–199.

[2] A. Benveniste, M. Metivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations.* Berlin: Springer-Verlag, 1990.

[3] P.E. Caines. *Linear Stochastic Systems.* New York: Wiley, 1988.

[4] D.P. Djereveckii and A.L. Fradkov. *Applied Theory of Discrete Adaptive Control Systems.* Moscow: Nauka, 1981. (In Russian).

[5] S. Geman. Some averaging and stability results for random differential equations, *SIAM Journal of Applied Mathematics* **36** (1979), 87–105.

[6] L. Gerencsér. Parameter tracking of time varying continuous–time linear stochastic systems. In *Modelling, Identification and Robust Control,* (C.I. Byrnes and A. Lindquist, eds.), pages 581–595. Amsterdam: North Holland, 1986.

[7] L. Gerencsér. Pathwise stability of random differential equation. Preprint 19, Department of Mathematics, Chalmers University of Technology, 1986.

[8] L. Gerencsér. On a class of mixing processes, *Stochastics* **26** (1989), 165–191.

[9] L. Gerencsér. Exponential stability of time varying differential equations, 1991, Manuscript.

[10] L. Gerencsér. Fixed gain estimation and tracking. In *Stochastic Theory and Adaptive Control. Proc. of a Workshop held in Lawrence,* (T.E. Duncan and B. Pasik-Duncan, eds.), pages 198–209. New York: Springer-Verlag, 1992.

[11] L. Gerencsér. Predictive stochastic complexity associated with fixed gain estimators. In *Proceedings of the 2-nd European Control Conference, Groningen,* (C. Praagman J.W. Niueuwenhuis and H.L. Trentelman, eds.), pages 1673–1677. 1993.

[12] L. Gerencsér. SUMMARY: Fixed gain off-line estimators of ARMA parameters, *J. of Mathematical Systems, Estimation and Control* **4(2)** (1994), 249-252. Full length paper appeared in **6(1)** (1996), 53-78.

[13] L. Gerencsér. Rate of convergence of the LMS algorithm, *Systems and Control Letters*, 1994. To appear.

[14] L. Gerencsér and J. Baikovicius. Change point detection using stochastic complexity. In *Proceedings of the 9th IFAC-IFORS Symposium on Identification and System Parameter Estimation* (L. Keviczky and Cs. Bányász, eds.), pages 395–400. Oxford: Pergamon Press, 1991.

[15] L. Gerencsér and J. Baikovicius. A computable criterion for model selection for linear stochastic systems. In *Proceedings of the 9th IFAC-IFORS Symposium on Identification and System Parameter Estimation*, (L. Keviczky and Cs. Bányász, eds.), pages 389–394. Oxford: Pergamon Press, 1991.

[16] L. Gerencsér and Zs. Vágó. Fixed gain estimation in continuous time, 1994. To appear in *Acta Applicaende Mathematica.*

[17] Ph. Hartman. *Ordinary Differential Equations.* New York: Wiley and Sons, Inc., 1964.

[18] H.J. Kushner and A.Schwartz. Weak convergence and asymptotic properties of adaptive filters with constant gains, *IEEE. Trans. on Information Theory* **30(2)** (1984), 177–182.

[19] L. Ljung. Analysis of recursive stochastic algorithms, *IEEE Trans. Aut. Cont* **22** (1977), 551–575.

[20] L. Ljung and T. Söderström. *Theory and practice of recursive identification.* Cambridge: The MIT Press, 1983.

[21] S.P. Meyn and P.E. Caines. A new approach to stochastic adaptive control, *IEEE Trans.on Aut.Control* **32** (1987), 220–226.

[22] L.S. Pontryagin. *Ordinary Differential Equations.* Moscow: Nauka, 1970.

[23] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific Publisher, 1989.

[24] F. Szigeti. A differential-algebraic condition for controllability and observability of time varying linear systems, 1992. Submitted.

[25] G. Zames and L.Y. Wang. Local-Global Double Algebras for Slow H^∞ Adaptation: Part I–Inversion and Stability, *IEEE Trans. Aut. Cont.* **36** (1991), 130–142.

Computer and Automation Institute, Hungarian Academy of Sciences H-1111, Budapest Kende u 13-17, Hungary. Also affiliated with the Department of Applied Mathematics and Informatics, the University of Pécs.

Communicated by Clyde F. Martin