INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING, VOL. 5, 155–174 (1991)

L_p-STABILITY OF ESTIMATION ERRORS OF KALMAN FILTER FOR TRACKING TIME-VARYING PARAMETERS

J. F. ZHANG, L. GUO AND H. F. CHEN

Institute of Systems Science, Academia Sinica, Beijing 100080, People's Republic of China

SUMMARY

The Kalman filtering algorithm, owing to its optimality in some sense, is widely used in systems and control, signal processing and many other fields. This paper presents a detailed analysis for the L_p-stability of tracking errors when the Kalman filter is used for tracking unknown time-varying parameters. The results of this paper differ from the previous ones in that the regression vector (in a linear regression model) or the output matrix (in state space terminology) is random rather than deterministic. The context is kept general so that, in particular, the time-varying parameter is allowed to be unbounded, and no assumption of stationarity or independence for signals is made.

KEY WORDS Stochastic systems Estimation Time-varying parameter Kalman filter

1. INTRODUCTION

Consider the time-varying linear model

$$y_k = \varphi_k^{\mathrm{T}}\theta_k + v_k. \quad \forall k \geqslant 0 \tag{1}$$

where y_k and v_k are the scalar output and noise respectively and φ_k and θ_k are the r-dimensional stochastic regression vector and the unknown time-varying parameter respectively. For simplicity of notation, denote the parameter variation at time instant k by w_k:

$$w_k \triangleq \theta_k - \theta_{k-1}, \quad \forall k \geqslant 1 \tag{2}$$

In the special case where v_k is a moving average process and φ_n consists of input-output data, i.e.

$$\varphi_k^{\mathrm{T}} = [y_{k-1} \ldots y_{k-s} \; u_{k-1} \ldots u_{k-t}]$$

with u_k being the input signal, then the linear model (1) is reduced to the ARMAX model with time-varying coefficients.

Tracking or estimating a system or a signal whose properties vary with time is a fundamental problem in system identification and signal processing. This problem has received considerable attention in the field of signal processing (see e.g. References 1–7), where most of the works are concerned with the study of the so-called least mean squares (LMS) algorithm or the normalized gradient algorithm, and usually some sort of stationarity and/or independence is required. In contrast to this, few precise studies have been done on the time-varying parameter-

This paper was recommended for publication by editor A. Benveniste

0890–6327/91/030155–20$10.00

Received 31 January 1990
Revised September 1990

tracking problem in the area of system identification, and most of the works are concentrated on the constant-parameter case, i.e. $w_k \equiv 0$ in (2) (see e.g. References 8–10).

Note that if we regard (1) and (2) as a state space model with state θ_k, then it is natural to use the Kalman filter to estimate the time-varying parameter θ_k (see e.g. References 6 and 11–14). The Kalman filter takes the form

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{P_k \varphi_k}{R + \varphi_k^{\mathrm{T}} P_k \varphi_k}(y_k - \varphi_k^{\mathrm{T}} \hat{\theta}_k) \tag{3}$$

$$P_{k+1} = P_k - \frac{P_k \varphi_k \varphi_k^{\mathrm{T}} P_k}{R + \varphi_k^{\mathrm{T}} P_k \varphi_k} + Q \tag{4}$$

where $P_0 \geqslant 0$, $R > 0$, $Q > 0$ and $\hat{\theta}_0$ are deterministic and can be arbitrarily chosen (here R and Q may be regarded as the *a priori* estimates for the variances of v_k and w_k respectively).

It is well known that if φ_k is $\mathscr{F}_{k-1}$-measurable, where $\mathscr{F}_{k-1} = \sigma\{y_i, i \leqslant k-1\}$, and $\{w_k, v_k\}$ is a Gaussian white noise process, then θ_k generated by (3) and (4) is the minimum variance estimate for θ_k and P_k is the estimation error covariance, i.e.

$$\hat{\theta}_k = \mathrm{E}(\theta_k \mid \mathscr{F}_{k-1}), \qquad P_k = \mathrm{E}(\tilde{\theta}_k \tilde{\theta}_k^{\mathrm{T}} \mid \mathscr{F}_{k-1}) \qquad (\tilde{\theta}_k = \theta_k - \hat{\theta}_k) \tag{5}$$

provided that $Q = \mathrm{E}(w_k w_k^{\mathrm{T}})$, $R = \mathrm{E}v_k^2$, $\hat{\theta}_0 = \mathrm{E}\theta_0$ and $P_0 = \mathrm{E}(\tilde{\theta}_0 \tilde{\theta}_0^{\mathrm{T}})$ (see e.g. References 15 and (16).

In studying asymptotic properties of the above algorithm, the primary issue is to establish boundedness (in some sense) of the tracking error $\tilde{\theta}_k$. This problem is obviously related to the stability theory of the Kalman filter, and the standard condition for such a stability (boundedness) is that the regression vector φ_k is deterministic and satisfies

$$\alpha I \leqslant \sum_{k=m+1}^{m+h} \varphi_k \varphi_k^{\mathrm{T}} \leqslant \beta I, \quad \forall m \tag{6}$$

for some positive constants α, β and integer h (see e.g. Reference 17).

As pointed out by Guo,[18] condition (6) is mainly a deterministic hypothesis; it excludes standard stochastic signals, including the Gaussian signal and white noise signals, and hence (6) is unsuitable for the stability study of the Kalman filter when φ_k is a random process. To the best of our knowledge, the first result which guarantees the stability of (3) and (4) and allows $\{\varphi_k\}$ to be a large class of stochastic processes appears in the recent work of Guo,[18] where it is assumed that $\{\varphi_k, \mathscr{F}_k\}$ is an adapted process ($\mathscr{F}_k$ is any family of non-decreasing σ-algebras) satisfying

$$\mathrm{E}\left(\sum_{k=m+1}^{m+h} \frac{\varphi_k \varphi_k^{\mathrm{T}}}{1 + \|\varphi_k\|^2} \middle| \mathscr{F}_m \right) \geqslant \alpha I, \quad \text{a.s.}, \quad \forall m \geqslant 0 \tag{7}$$

for some constant $\alpha > 0$ and integer h.

In this paper we will weaken condition (7) and generalize results in Reference 18, and at the same time provide new results and insights. The paper is organized as follows. In Section 2 we present our main results. Section 3 provides some lemmas and establishes the properties of $\{P_k\}$. The proof of the theorems is put into Section 4. Finally, in the Appendix the proof of auxiliary results which are used in the text is provided.

2. MAIN RESULTS

In the sequel the norm $\|X\|$ for a matrix X is defined as $\|X\| = \{\lambda_{\max}(XX^{\mathrm{T}})\}^{1/2}$, and $\lambda_{\max}(X)$ ($\lambda_{\min}(X)$) denotes the largest (smallest) eigenvalue of X. Let us first give a definition.

Definition 1

A random vector sequence $\{x_k, k \geqslant 0\}$ defined on the basic probability space $(\Omega, \mathscr{F}, P)$ is called L_p-stable $(p > 0)$ in the sample average sense if

$$\sup_{k \geqslant 0} \mathrm{E} \| x_k \|^p < \infty$$

and in the time average sense if

$$\sup_{k > 0} \frac{1}{k} \sum_{i=0}^{k} \| x_i \|^p < \infty, \quad p > 0, \quad \text{a.s.}$$

We now give the main condition that will be used in the paper.

Condition 1

$\{\varphi_k, \mathscr{F}_k\}$ is an adapted sequence (i.e. φ_k is $\mathscr{F}_k$-measurable for any k, where $\{\mathscr{F}_k\}$ is a family of non-decreasing σ-algebras) and there exists an integer $h > 0$ such that

$$\mathrm{E}\left(\sum_{k=m+1}^{m+h} \frac{\varphi_k \varphi_k^{\mathrm{T}}}{1 + \| \varphi_k \|^2} \,\middle|\, \mathscr{F}_m \right) \geqslant \frac{1}{\alpha_m} I, \quad \text{a.s.}, \quad \forall m \geqslant 0 \tag{8}$$

Here $\{\alpha_m, \mathscr{F}_m\}$ is an adapted non-negative sequence satisfying

$$\alpha_{m+1} \leqslant a\alpha_m + \eta_{m+1}, \quad \forall m \geqslant 0, \qquad M_0 \triangleq \mathrm{E}\alpha_0^{1+\delta} < \infty \tag{9}$$

where $\{\eta_m, \mathscr{F}_m\}$ is an adapted non-negative sequence,

$$\sup_{m \geqslant 0} \mathrm{E}(\eta_{m+1}^{1+\delta} \mid \mathscr{F}_m) \leqslant M, \quad \text{a.s.} \tag{10}$$

and where $a \in [0, 1)$, $0 < \delta < \infty$ and $0 \leqslant M < \infty$ are constants.

Remark 1

At first glance, Condition 1 looks rather complicated; however, it does have a clear meaning and is satisfied by a large class of stochastic signals. Note first that if in (9) we take $a = 0$ and $\eta_m \equiv \alpha^{-1}$ for some $\alpha > 0$, then (8) reduces to (7), which is weaker than (6), and hence Condition 1 includes (6) and (7). Next note that the sequence $\{\alpha_k\}$ required in (8) may not be bounded in the sample path and hence the matrix on the LHS of (8) may not be uniformly positive definite, so (8) is really weaker than (7).

Let us further illustrate Condition 1 by the following examples.

Example 1

Let $\{\varphi_k\}$ be an r-dimensional ϕ-mixing process; that is, there is a deterministic sequence $\{\phi(h), h \geqslant 0\}$ such that

(i) $\qquad \phi(h) \to 0 \quad \text{as } h \to \infty$

(ii) $\qquad \sup_{A \in \mathscr{F}_{s+h}^{\infty},\, B \in \mathscr{F}_0^{s}} | P(A \mid B) - P(A) | \leqslant \phi(h), \quad \forall s \geqslant 0, \quad \forall h \geqslant 0$

where, for any non-negative integers $s \geqslant 0$ and $h \geqslant 0$, $\mathscr{F}_0^s \triangleq \sigma\{\varphi_k, 0 \leqslant k \leqslant s\}$ and $\mathscr{F}_{s+h}^{\infty} \triangleq \sigma\{\varphi_k, s+h \leqslant k < \infty\}$.

Suppose further that

$$\inf_k \lambda_{\min} \mathrm{E}(\varphi_k \varphi_k^{\mathrm{T}}) > 0 \quad \text{and} \quad \sup_k \mathrm{E} \| \varphi_k \|^4 < \infty \tag{11}$$

Then Condition 1 holds with $\mathscr{F}_m = \mathscr{F}_0^m$.

The proof of this example is given in the Appendix. We remark that any h-dependent random process (including moving average processes of order h) is ϕ-mixing. This kind of condition has been previously used, for example, by Eweda and Macchi,[3] Macchi[4] and Kushner[19] in their study of the LMS algorithm.

Example 2

Let $\{\varphi_k\}$ be the output of the linear stochastic model

$$x_k = A x_{k-1} + B \xi_k, \quad \forall k \geqslant 1, \qquad \mathrm{E} \| x_0 \|^5 < \infty \tag{12}$$

$$\varphi_k = C x_k + \zeta_k, \quad \forall k \geqslant 0 \tag{13}$$

where $A \in R^{n \times n}$, $B \in R^{n \times q}$ and $C \in R^{r \times n}$ are deterministic matrices, A is stable and (A, B, C) is output-controllable in the sense of Fortmann and Hitz.[20]

Suppose that $\{\xi_k\}$ and $\{\zeta_k\}$ are independent processes which are also mutually independent and satisfy

$$\mathrm{E}\xi_k = 0, \qquad \mathrm{E}\zeta_k = 0 \tag{14}$$

$$\mathrm{E}(\xi_k \xi_k^{\mathrm{T}}) \geqslant \varepsilon I, \quad \forall k \geqslant 0 \tag{15}$$

$$\mathrm{E}(\| \xi_k \|^{4(1+\mu)} + \| \zeta_k \|^4) \leqslant M < \infty, \quad \forall k \geqslant 0 \tag{16}$$

for some constants $\varepsilon > 0$, $\mu > 0$ and $M > 0$. Then Condition 1 is fulfilled.

The proof is given in the Appendix. It is worth noting that it is generally hard to show that $\{\varphi_k\}$ defined in Example 2 satisfies condition (7) unless the noise process $\{\xi_k, \zeta_k\}$ is assumed to be bounded (see Reference 18, Example 2). We now present the main results of the paper.

Theorem 1

Consider the time-varying model (1) and (2). Suppose that $\{v_k, w_k\}$ is a stochastic sequence which satisfies for some $p > 0$ and $\beta > 1$

$$\sigma_p \triangleq \sup_{k \geqslant 0} \mathrm{E}\{ Z_k^p [\log(e + Z_k)]^{\beta + 3p/2} \} < \infty \tag{17}$$

$$\mathrm{E}\{ \| \tilde{\theta}_0 \|^p [\log(e + \| \tilde{\theta}_0 \|)]^{p/2} \} < \infty \tag{18}$$

where $Z_k = \| v_k \| + \| w_{k+1} \|$, $\tilde{\theta}_0 = \theta_0 - \hat{\theta}_0$ and v_k, w_k, θ_0 and $\hat{\theta}_0$ are given by (1)–(4) respectively. Then under Condition 1 the estimation error $\{\theta_k - \hat{\theta}_k, k \geqslant 0\}$ generated by (3) and (4) is L_p-stable in the sample average sense and

$$\limsup_{k \to \infty} \mathrm{E} \| \theta_k - \hat{\theta}_k \|^p \leqslant A [\sigma_p \log^{1 + 3p/2}(e + \sigma_p^{-1})] \tag{19}$$

where A is a finite constant dependent on h, a, M, M_0 and δ only.