TRACKING RANDOMLY VARYING PARAMETERS

Analysis of a Standard Algorithm *

L. Guo, L. Xia, J.B. Moore

Department of Systems Engineering , R.S.Phys.S, Australian National University,
G.P.O. Box 4, Canberra, ACT 2601 , Australia.

* Work partially supported by D.S.T.O. Australia.

Abstract

In linear stochastic system identification, when the unknown parameters are randomly time varying and can be represented by a Markov model, a natural estimation algorithm to use is the Kalman filter. In seeking an understanding of the properties of this algorithm, existing Kalman filter theory yields useful results only for the case where the noises are Gaussian with covariances precisely known. In other cases, the stochastic and unbounded nature of the regression vector (which is regarded as the output gain matrix in state space terminology) precludes application of standard theory. Here we develop asymptotic properties of the algorithm. In particular, we establish the tracking error bounds for the unknown randomly varying parameters.

I. Introduction

Let us first define the signal model class and estimation algorithm.
Signal model: Consider the following linear regression model

$$y_k = \varphi_k^\tau \theta_k + v_k , k \geq 0 \tag{1.1a}$$

$$\theta_{k+1} = F\theta_k + w_{k+1} , \ E\|\theta_0\|^2 < \infty \tag{1.1b}$$

with θ_k viewed as time varying unknown parameters with a Markov model representation. The noise sources $\{w_k\}$ and $\{v_k\}$ are mutually independent and also independent themselves, with zero mean and covariance.

$$E[w_{k+1}w_{k+1}^\tau] = Q_w \geq 0 , E[v_{k+1}v_{k+1}^\tau] = R_v > 0 \tag{1.2}$$

(Generalization of our theory to the case of time varying covariances is straightforward). The measurement y_k is assumed scalar, and the regression vector φ_k is stochastic and belongs to F_{k-1}— the σ-algebra generated by $\{y_0, y_1, ...y_{k-1}\}$.

Much of the work done in stochastic system identification has been concerned with identifying the parameters θ_k in (1.1) for the case when $\theta_k = \theta_0$ is constant, that is, when $F = I$ and the covariances of w_k is zero. Typically, φ_k is viewed as the regression vector of an ARMAX model and least squares identification of θ_0 is applied. When θ_k is time varying, one natural approach to use is to model θ_k as in (1.1b) where all eigenvalues of F lie in or on the unit circle, i.e., $|\lambda_i(F)| \leq 1$, for all i. In such cases, the natural performance criterion is tracking error bounds.

Estimation Algorithm (Kalman filter) : Consider the following estimation algorithm associated with (1.1) as

$$\hat{\theta}_{k+1} = F\hat{\theta}_k + \frac{FP_k\varphi_k}{R+\varphi_k^\tau P_k\varphi_k}(y_k - \varphi_k^\tau\hat{\theta}_k) \tag{1.3a}$$

$$P_{k+1} = FP_kF^\tau - \frac{FP_k\varphi_k\varphi_k^\tau P_kF^\tau}{R+\varphi_k^\tau P_k\varphi_k} + Q , \tag{1.3b}$$

where $P_0 \geq 0$, $Q > 0$ and $R > 0$ as well as $\hat{\theta}_0$ are deterministic and can be arbitrarily chosen (here Q and R may be regarded as a priori estimates for Q_w and R_v respectively. We stress that even if Q_w is singular, here Q must be chosen as nonsingular to achieve a short memory algorithm. That is the adaptation gain in (1.3a) does not diminish to zero.).

It is known that if the noise source $\{ w_k^\tau , v_k \}$ is a Gaussian white noise sequence, then $\hat{\theta}_k$ generated by (1.3) is the best estimate for θ_k, and P_k is the estimation error covariance, i.e.

$$\hat{\theta}_k = E[\theta_k | F_{k-1}] , P_k = E[\tilde{\theta}_k\tilde{\theta}_k^\tau | F_{k-1}],$$

provided that $Q = Q_w$, $R = R_v$, $\hat{\theta}_0 = E[\theta_0]$ and $P_0 = E[\tilde{\theta}_0\tilde{\theta}_0^\tau]$, where $\tilde{\theta}_k$ is the estimation error:

$$\tilde{\theta}_k = \theta_k - \hat{\theta}_k \tag{1.4}$$

This remarkable result was first observed by Mayne [1] and expanded on by various authors e.g. Astrom and Wittenmark [2], Kitagawa and Gersch [3].

In the nongaussian case, however, the properties of (1.3) applied to (1.1) have not been well studied. The reasons for this may be explained as follows: a). In the time varying case, there is no almost sure parameter convergence. Also, the successful stochastic Lyapunov function technique, as well as the martingale limit approach used in least squares (LS) convergence analysis (e.g.,Ljung [4], Moore [5], Lai and Wei [6], and Chen and Guo [7]), fail in the present case. This is so even though (1.3) is the standard LS algorithm when $F = I$, $Q = 0$ and $R = 1$. Similar observations are also made by Meyn and Caines [8]; b). The algorithm (1.3) is a Kalman filter when $Q = Q_w$ and $R = R_v$. It is optimal in a linear minimum variance sense when φ_k is deterministic (e.g. Anderson and Moore [9]) , and not stochastic as here. Thus, the stochastic nature of the regressors precludes applicability of the useful properties of the Kalman filter, even when Q_w and R_v are precisely known; c).The existing theory for time varying linear systems usually requires that the system output gain matrix (i.e., φ_k , in the present case) is bounded in k (e.g. Anderson and Moore [10]). This requirement turns out to be unrealistic in applying the theory to general adaptive control and identification problems. This is especially so in the stochastic case, because φ_k may contain the past system inputs and outputs, and the system noise may be unbounded. Hence, the unbounded nature of the regressors $\{\varphi_k\}$ also precludes the direct application of the standard theory.

In this paper, we establish tracking error bounds for the case of randomly varying parameters. The main concern is with the following three cases:

(i). Parameters generated from a stable linear model, i.e., (1.1b) with $|\lambda_i(F)| < 1$, for any i;

(ii).Drifting parameters, i.e., (1.1b) with $F = I$; and

(iii). Disturbed parameters ,i.e., $\theta_k = \theta_0 + w_k$.

II. Tracking Error Bound

In the sequel, we denote $\lambda_{max}(A)$ and $\lambda_{min}(A)$ the maximum and minimum eigenvalues of a matrix A respectively, and $\|A\| = \{ \lambda_{max}(AA^\tau)\}^{1/2}$ its norm, so that $\|A\| = \lambda_{max}(A)$ when A is symmetric and nonnegative definite.

Let us first denote

$$K_k = FP_k\varphi_k(R + \varphi_k^\tau P_k\varphi_k)^{-1} \quad (2.1)$$

and rewrite (1.3) as

$$\hat{\theta}_{k+1} = F\hat{\theta}_k + K_k (y_k - \varphi_k^\tau\hat{\theta}_k), \quad (2.2a)$$

$$P_{k+1} = (F - K_k\varphi_k^\tau) P_k (F - K_k\varphi_k^\tau)^\tau + K_kRK_k^\tau + Q, \quad (2.2b)$$

The lower bounds to the tracking error is relatively straightforward by combining (1.1b) and (2.2a), indeed, we have

Theorem 2.1. Consider the signal model (1.1) and algorithm (1.3), if

$$\sup_k E\|w_k\|^{2+\varepsilon} < \infty \text{ for some } \varepsilon > 0 \text{, then}$$

$$\inf_k E\| \tilde{\theta}_k\|^2 \geq tr(Q_w), \quad (2.3)$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \|\tilde{\theta}_i\|^2 \geq tr(Q_w), \text{ a.s.}, \quad (2.4)$$

where Q_w and $\tilde{\theta}_k$ are defined by (1.2) and (1.4) respectively.

Proof. By (1.1) and (2.2a), the error equation is

$$\tilde{\theta}_{k+1} = (F - K_k\varphi_k^\tau)\tilde{\theta}_k - K_kv_k + w_{k+1} \quad (2.5)$$

Set

$$f_k = (F - K_k\varphi_k^\tau)\tilde{\theta}_k - K_kv_k$$

then $\{f_k^\tau w_{k+1}\}$ is a martingale difference sequence with respect to the σ-algebra generated by $\{ v_{i-1}, w_i, i \leq k+1\}$, so the first assertion (2.3) follows from (2.5) and the orthogonality of f_k and w_{k+1} immediately.

Now, by an estimation for the weighted sum of martingale difference sequences (e.g. Chen and Guo [11], pp. 848), we know that

$$\sum_{i=1}^{n} f_i^\tau w_{i+1} = O(\{ \sum_{i=1}^{n} \|f_i\|^2 \}^{(1/2)+\eta}), \text{ a.s.}$$

for any $\eta > 0$. Consequently, by taking $\eta < \frac{1}{2}$, we have

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\|f_i\|^2 + 2f_i^\tau w_{i+1})$$

$$= \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\|f_i\|^2 \{1 + O([\sum_{i=1}^{n}\|f_i\|^2]^{\eta-(1/2)})\} \geq 0$$

From this inequality and (2.5) it follows that

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\| \tilde{\theta}_i \|^2 = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}\{\|f_i\|^2 + 2f_i^\tau w_{i+1} + \|w_{i+1}\|^2\}$$

$$\geq \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}\|w_{i+1}\|^2 \geq tr(Q_w), \text{ a.s.}$$

which is the second assertion (2.4). Hence the proof is complete. #

The upper bounds for the tracking error depend on the stability of the equation

$$\xi_{k+1} = (F - K_k\varphi_k^\tau) \xi_k, \ k \geq 0, \quad (2.6)$$

as can be seen from (2.5), which we will show depend on the bounds of $\{P_k\}$.

A lower bound to P_k is easy to get, since from (2.2b):

$$P_k \geq Q > 0, \quad \text{for any } k \geq 1. \quad (2.7)$$

However, upper bounds for $\{P_k\}$ are far from obvious for general F and $\{\varphi_k\}$. Let us first see the role played by the upper bound of $\{P_k\}$ in the stability of the equation (2.6).

Lemma 2.1. Assume that there exists a random constant b such that

$$\sup_{k\geq 0} \| P_k \| \leq b < \infty, \text{ a.s.}, \quad (2.8)$$

then for K_k defined by (2.1),

$$\| \prod_{k=i}^{j-1}(F - K_k\varphi_k^\tau) \| \leq \beta\, \alpha^{j-i}, \text{ a.s., for any } j > i \geq 0, \quad (2.9)$$

where α and β are defined by

$$\alpha = \frac{\|F\|(a+b)b^{1/2}}{[a^3 + \|F\|^2(a+b)^2b]^{1/2}}, \quad (2.10)$$

$$\beta = [b/a]^{1/2}, \quad a = \lambda_{min}(Q). \quad (2.11)$$

The proof is given in Appendix A. The precise expressions of α and β in (2.10) and (2.11) lead directly to the following important observation.

Remark 2.1. If b, the upper bound of P_k, is a deterministic constant, then the exponential bounds claimed in (2.9) are also deterministic.

This fact is very crucial in establishing the upper bound for the tracking errors in terms of mathematical expectations in the sequel.

Let us now proceed to establish the upper bound for the tracking errors by considering different parameter models separately.

A. Parameters Generated from a Stable Model.

In this case, $|\lambda_i(F)| < 1$ for all i, then by (1.3b),

$$P_{k+1} \leq FP_kF^\tau + Q$$
$$\leq F^2P_{k-1}(F^\tau)^2 + FQF^\tau + Q \leq$$
$$\leq \sum_{i=0}^{k} F^iQ(F^\tau)^i + F^{k+1}P_0(F^\tau)^{k+1}, \text{ for any } k \geq 0, \quad (2.12)$$

and hence

$$b \triangleq \| \sum_{i=0}^{\infty} F^iQ(F^\tau)^i \| + \sup_{k\geq 0} \| F^kP_0(F^\tau)^k\|, \quad (2.13)$$

can serve as a finite deterministic upper bound for $\{P_k\}$ since P_0 is deterministic. This enables us to establish the following results.

Theorem 2.2. Consider the signal model (1.1) with $|\lambda_i(F)| < 1$, for any i, and the estimation algorithm (1.3). Then

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\theta}_i\|^p \leq (\frac{\beta}{1-\alpha})^p [L_p(w) + \|F\|(b/R)^{1/2} L_p(v)]^p, \quad (2.14)$$

and

$$\limsup_{n\to\infty} E \|\tilde{\theta}_n\|^p \leq (\frac{\beta}{1-\alpha})^p [M_p(w) + \|F\|(b/R)^{1/2} M_p(v)]^p, \quad (2.15)$$

here $\tilde{\theta}_n = \theta_n - \hat{\theta}_n$, b, α and β are given by (2.13), (2.10) and (2.11) respectively, and $p > 1$ is any real number such that

$$L_p(v) \triangleq \limsup_{n\to\infty} \{ \frac{1}{n}\sum_{i=1}^{n}\| v_i \|^p \}^{1/p} < \infty, \text{ a.s.} \quad (2.16)$$

$$L_p(w) \triangleq \limsup_{n\to\infty} \{ \frac{1}{n}\sum_{i=1}^{n}\| w_i \|^p \}^{1/p} < \infty, \text{ a.s.} \quad (2.17)$$

$$M_p(v) \triangleq \sup_i \{ E\| v_i \|^p \}^{1/p} < \infty, \quad (2.18a)$$

$$M_p(w) \overset{\Delta}{=} \sup_i \{ E\| w_i \|^p \}^{1/p} < \infty, \qquad (2.18b)$$

and $E \| \theta_0 \|^p < \infty$.

Proof. By (2.5) we have

$$\tilde{\theta}_{k+1} = \prod_{j=0}^{k} (F-K_j\varphi_j^\tau)\, \tilde{\theta}_0 + \sum_{i=0}^{k} [\prod_{j=i+1}^{k} (F-K_j\varphi_j^\tau)] (-K_i v_i + w_{i+1})$$

Applying Lemma 2.1 we see that

$$\| \tilde{\theta}_{k+1} \| \le \beta\, \alpha^{k+1} \| \tilde{\theta}_0 \| + \beta \sum_{i=0}^{n} \alpha^{k-i} (\|K_i v_i\| + \|w_{i+1}\|), \qquad (2.19)$$

then applying the Minkowski inequality gives

$$(\sum_{i=0}^{n} \| \tilde{\theta}_{i+1} \|^p)^{1/p} \le \beta \| \tilde{\theta}_0 \| (\sum_{k=0}^{n} \alpha^{p(k+1)})^{1/p} +$$

$$\beta \{ \sum_{k=0}^{n} (\sum_{i=0}^{k} \alpha^{k-i} \|K_i v_i\|)^p \}^{1/p} + \beta \{ \sum_{k=0}^{n} (\sum_{i=0}^{k} \alpha^{k-i} \|w_{i+1}\|)^p \}^{1/p}, \qquad (2.20)$$

now, by the Holder inequality it follows that ($1/p + 1/q = 1$) :

$$(\sum_{i=0}^{k} \alpha^{k-i} \| K_i v_i \|)^p = \{ \sum_{i=0}^{k} \alpha^{(k-i)/q} [\alpha^{(k-i)/p} \| K_i v_i \|] \}^p$$

$$\le (\sum_{i=0}^{k} \alpha^{k-i})^{p/q} (\sum_{i=0}^{k} \alpha^{k-i} \| K_i v_i \|^p) \le (\frac{1}{1-\alpha})^{p/q} \sum_{i=0}^{k} \alpha^{k-i} \| K_i v_i \|^p$$

and then

$$\sum_{k=0}^{n} (\sum_{i=0}^{k} \alpha^{k-i} \| K_i v_i \|)^p \le (\frac{1}{1-\alpha})^{p/q} \sum_{i=0}^{n} \sum_{k=i}^{n} \alpha^{k-i} \| K_i v_i \|^p$$

$$\le (\frac{1}{1-\alpha})^{(p/q)+1} \sum_{i=0}^{n} \| K_i v_i \|^p = (1-\alpha)^{-p} \sum_{i=0}^{n} \| K_i v_i \|^p \qquad (2.21)$$

Let us now consider the upper bound for K_i. Since b is an upper bound for $\|P_k\|$, then by (2.1),

$$\| K_k \|^2 \le \|F\|^2 \frac{\varphi_k^\tau P_k^2 \varphi_k}{(R + \varphi_k^\tau P_k \varphi_k)^2} \le \|F\|^2\, b \frac{\varphi_k^\tau P_k \varphi_k}{(R + \varphi_k^\tau P_k \varphi_k)^2}$$

$$\le \|F\|^2 b / R, \text{ for any } k \ge 0, \qquad (2.22)$$

which together with (2.16) and (2.21) yields

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} (\sum_{i=0}^{k} \alpha^{k-i} \|K_i v_i\|)^p \le [\frac{\|F\|(b/R)^{1/2}}{1-\alpha}]^p [L_p(v)]^p, \qquad (2.23)$$

Similarly,

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} (\sum_{i=0}^{k} \alpha^{k-i} \|w_{i+1}\|)^p \le (\frac{1}{1-\alpha})^p [L_p(w)]^p \qquad (2.24)$$

Finally, the first result (2.14) follows from (2.20), (2.23) and (2.24).

Let us now consider (2.15). The inequality corresponding to (2.20) can also be derived by the Minkowski inequality and takes the form :

$$(E\|\tilde{\theta}_{k+1}\|^p)^{1/p} \le \beta\, \alpha^{k+1} (E \| \tilde{\theta}_0 \|^p)^{1/p} +$$

$$+ \beta \{ E(\sum_{i=0}^{k} \alpha^{k-i} \| K_i v_i \|)^p \}^{1/p} + \beta \{ E(\sum_{i=0}^{k} \alpha^{k-i} \| w_{i+1} \|)^p \}^{1/p}$$

From this, a similar argument as used in the proof of (2.14) leads to (2.15) because in this case the constants b, α and β are all deterministic. This completes the proof. #

Remark 2.2. From the proof of Theorem 2.2 we see that the independence assumptions made on the noise sequences $\{w_k\}$ and $\{v_k\}$ are not really used, indeed, Theorem 2.2 holds for any random sequences $\{w_k\}$ and $\{v_k\}$ satisfying (2.16)-(2.18). In particular, w_k, which appeared in the parameter model (1.1b) may have non-zero mean.

Remark 2.3. We have recently applied the property (2.15) with $p = 4 + \delta$ for some $\delta > 0$, to adaptive control problems (Guo and Meyn [12]), and it appears that the non-trivial stochastic adaptive control problem considered by Meyn and Caines [8] can be generalized to the case where the noises are nongaussian with unknown covariances.

Remark 2.4. Observe that there is no excitation requirement to achieve the bounds of the theorem. Of course, from (1.3b) and the matrix inversion lemma,

$$P_{k+1} = F [(P_k)^{-1} + \varphi_k R^{-1} \varphi_k^\tau]^{-1} F^\tau + Q$$

and it is clear that the greater the excitation of φ_k, the smaller is P_{k+1} in norm and the lower are the tracking error bounds (α,β are smaller).

B. Drifting Parameters.

In this case, $F = I$, and similar arguments as used in (2.12) for the boundedness proof of $\{P_k\}$ fail. Moreover, it turns out that it is impossible to establish the upper bounds for P_k without further assumptions on the regressors $\{\varphi_k\}$. To see this, let us take $\varphi_k = 0$, for all $k \ge 0$, then by (1.3b),

$$P_{k+1} = P_k + Q = P_0 + (k+1)Q \xrightarrow[k \to \infty]{} \infty. \qquad (2.25)$$

Nevertheless, we have the following results.

Lemma 2.2. Consider that there exists a strictly increasing sequence of random integers $\{t_n\}$ with $t_0 = 0$, $d \overset{\Delta}{=} \sup_k (t_n - t_{n-1}) < \infty$, a.s., and random constants $\delta > 0$, $M < \infty$ such that for any $k \ge 1$,

$$\lambda_{min}(k) \ge \delta, \text{ a.s.} \qquad (2.26)$$

and

$$\lambda_{max}(k) / \lambda_{min}(k) \le M, \text{ a.s.}, \qquad (2.27)$$

where $\lambda_{max}(k)$ and $\lambda_{min}(k)$ denote the maximum and minimum eigenvalues of the matrix

$$\sum_{i=t_{k-1}+1}^{t_k} \varphi_i \varphi_i^\tau \qquad (2.28)$$

respectively. Then $\{P_k\}$ defined by (1.3b) with $F = I$, has the following upper bound:

$$\sup_k \|P_k\| \le \|P_0\| + R/\delta + [1 + (1+M)d] \|Q\| < \infty, \text{ a.s.} \qquad (2.29)$$

The proof of this lemma is given in Appendix B. The conditions (2.26)-(2.27) can be regarded as certain kinds of excitations, thus, the divergence phenomena as in (2.25) may be explained as lack of excitation of $\{\varphi_k\}$.

It is interesting to compare the conditions (2.26)-(2.27) with the standard persistence of excitation condition used in the analysis of *short memory* adaptive control algorithms in the literature (e.g. Anderson et al,[13]). That is, there exist constants $0 < \delta_1 \le \delta_2 < \infty$, and $N < \infty$ such that

$$\delta_1 I \le \sum_{i=k}^{k+N} \varphi_i \varphi_i^\tau \le \delta_2 I, \text{ for any } k \ge 0. \qquad (2.30)$$

This implies that $\{\varphi_k\}$ is a bounded sequence. Clearly, Condition (2.26)-(2.27) is weaker than (2.30), and it means that $\lambda_{max}(k)$ and $\lambda_{min}(k)$ may grow at the same rate, and does not necessarily mean that $\{\varphi_k\}$ is bounded. As an example, let us take φ_k as a scalar slope function : $\varphi_k = c\,k$, $c \neq 0$, then, clearly (2.30) fails, while (2.26)-(2.27) still holds because $\lambda_{max}(k)$ and $\lambda_{min}(k)$ coincide in this case. A related but different excitation condition to (2.26)-(2.27) has been introduced and studied in (Chen and Guo,[14]) for the analysis of short memory gradient algorithms when the regressors $\{\varphi_k\}$ are possibly unbounded.

Remark 2.5. Lemma 2.2 can be generalized to the case where $F \neq I$, and a similar bound as in (2.29) is also achieved. Similar results as in the following Theorem 2.3 are also available. However, in this case, the matrix given by (2.28), which are used in defining $\lambda_{min}(k)$ and $\lambda_{max}(k)$, will involve the matrix F in general.

Theorem 2.3. Consider the signal model (1.1) with F = I, and the estimation algorithm (1.3), consider also that the conditions in Lemma 2.2 apply. Then

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \|\tilde{\theta}_i\|^p \leq \left(\frac{\beta}{1-\alpha}\right)^p [L_p(w) + (b/R)^{1/2} L_p(v)]^p \quad (2.31)$$

Here $\tilde{\theta}_k = \theta_k - \hat{\theta}_k$, α and β are defined by (2.10) and (2.11) with F = I and with the upper bound b for $\{P_k\}$ given in (2.29). Also, $L_p(w)$, $L_p(v)$ and $p > 1$ are defined in (2.16)-(2.17).

Proof. The proof is actually the same as that for (2.14). Note that the result (2.15) is also achieved in the present case provided that the quantity on the R.H.S. of (2.29) is deterministic. #

As an example, let us now consider the i.i.d. noise case, and without loss of generality assume that θ_k is one dimensional. More precisely, let $\{w_k\}$ be i.i.d. random variables with mean zero and variance $\sigma^2 > 0$. Putting F = I in (1.1b) we get

$$\theta_n = \theta_{n-1} + w_n = \theta_0 + S_n ,\ S_n = \sum_{i=1}^{n} w_i \quad (2.32)$$

Consequently, by Strassen's invariance principle (Strassen, [15]),

$$\limsup_{n\to\infty} \left(\sum_{i=1}^{n} S_i^2\right) / (n^2 \log\log n) = 8\sigma^2/\pi^2 \cdot \text{a.s.},$$

On the other hand, by a result of Donsker and Varadhan [16, pp.751],

$$\liminf_{n\to\infty} \left(\sum_{i=1}^{n} S_i^2\right) (\log\log n) / n^2 = \sigma^2/4 \cdot \text{a.s.} .$$

Hence with probability 1, the averaged value of parameters

$$\frac{1}{n}\sum_{i=1}^{n} (\theta_i)^2 \sim \frac{1}{n}\left(\sum_{i=1}^{n} S_i^2\right),\ n \to \infty$$

fluctuate in the interval

$$\left[\left(\frac{1}{4} + o(1)\right) \sigma^2 n / \log\log n ,\ (8\pi^{-2} + o(1))\, \sigma^2 n \log\log n \right]$$

as $n \to \infty$. Thus from this and the result (2.31) we see that the estimation algorithm (1.3) can indeed perform the non-trivial task of tracking rapidly varying parameters in the long run average sense.

Let us consider another situation.

C. Disturbed Parameters.

By disturbed parameters we mean that the parameters can be modeled by

$$\theta_k = \theta_0 + w_k \quad (2.33)$$

with unknown θ_0 and noise $\{w_k\}$. This case, is not a specialization of (1.1b), but can still be studied by use of the theory developed.

Theorem 2.4. Consider the signal model (1.1a) with parameters described by (2.33), and the algorithm (1.3) with F = I. Consider also that conditions of lemma 2.2 apply. Then

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \|\tilde{\theta}_i\|^p \leq \left(\frac{\beta}{1-\alpha}\right)^p [(b/a)L_p(w) + (b/R)^{1/2} L_p(v)]^p \quad (2.34)$$

where $\tilde{\theta}_k = \theta_0 - \hat{\theta}_k$, $a = \lambda_{min}(Q)$, and the constants α, β, b, $L_p(w)$ and $L_p(v)$ are all the same as those in Theorem 2.3.

Proof. With $\tilde{\theta}_k = \theta_0 - \hat{\theta}_k$ and F = I, the error equation (2.5) is now changed to

$$\tilde{\theta}_{k+1} = (I - K_k\varphi_k^\tau)\tilde{\theta}_k - K_k v_k + K_k\varphi_k^\tau w_{k+1}$$

Note that by (A1) in Appendix A , $K_k\varphi_k^\tau$ is bounded by

$$\|K_k\varphi_k^\tau\| \leq b/a .$$

Hence, a similar argument as used in the proof of (2.14) leads to the desired result (2.34). The details are not repeated here. #

III. Conclusions

When the Kalman filter is applied to estimation of randomly varying parameters, our results show that it has quite reasonable tracking properties ---- even in the nongaussian case when it is not an optimal filter.

If the parameters are generated from a stable model, we have seen that there is no restriction on the regressors to achieve tracking error bounds. The bounds obtained have application for adaptive controller analysis.

If the parameters are drifting, as when the parameter model is unstable, the theory of the paper shows that the regressors must be suitably exciting to achieve tracking error bounds. For the case of parameters disturbed by noise, there is again an excitation requirement to achieve tracking error bounds.

In a companion paper, the bounds studied here are applied in an adaptive control context.

Appendix A

Proof of Lemma 2.1:

We first establish the upper bound for $K_k\varphi_k^\tau$ as follows (note that φ_k may be unbounded),

$$\begin{aligned} \| K_k\varphi_k^\tau \| &\leq \|F\| \|P_k\| \|\varphi_k\|^2 / (R + \varphi_k^\tau P_k \varphi_k) \\ &\leq \|F\|\, b\, \|\varphi_k\|^2 / [\lambda_{min}(Q) \|\varphi_k\|^2], \quad \text{(by (2.7))}, \\ &\leq \|F\| \frac{b}{a} \end{aligned} \quad (A1)$$

Let us then denote for simplicity $F_k = F - K_k\varphi_k^\tau$. An upper bound for F_k is

$$\|F_k\| \leq \|F\| \left(1 + \frac{b}{a}\right) \quad (A2)$$

Now consider the following inequalities. By (A2), (2.2b) and the matrix inversion Lemma,

$$\begin{aligned} &P_k^{-1} - F_k^\tau P_{k+1}^{-1} F_k \\ &= P_k^{-1} - F_k^\tau [F_k P_k F_k^\tau + K_k R K_k^\tau + Q]^{-1} F_k \\ &\geq P_k^{-1} - F_k^\tau [F_k P_k F_k^\tau + Q]^{-1} F_k \end{aligned}$$

$$= [P_k + P_k F_k^\tau Q^{-1} F_k P_k]^{-1}$$

$$\geq [P_k + (P_k)^{1/2} \| (P_k)^{1/2} F_k^\tau Q^{-1} F_k (P_k)^{1/2} \| (P_k)^{1/2}]^{-1}$$

$$\geq [P_k + \|F\|^2 (1 + \frac{b}{a})^2 \frac{b}{a} P_k]^{-1}$$

$$= [1 + \|F\|^2 (1 + \frac{b}{a})^2 \frac{b}{a}]^{-1} P_k^{-1} .$$

Consequently, by the definition (2.10) for α :

$$F_k^\tau P_{k+1}^{-1} F_k \leq \alpha^2 P_k^{-1} , \qquad \text{for any } k \geq 0.$$

Thus, noting (2.7) and (2.8), and repeatedly using this inequality, we get

$$\| \prod_{k=i}^{j-1} F_k \|^2 \leq b \| (\prod_{k=i}^{j-1} F_k)^\tau P_j^{-1} (\prod_{k=i}^{j-1} F_k) \|$$

$$\leq b \alpha^{2(j-i)} \|P_i^{-1}\| \leq (\frac{b}{a}) \alpha^{2(j-i)}, \text{for any } j > i \geq 0.$$

#

Appendix B

Proof of Lemma 2.2.

Clearly, if the result holds for any deterministic sequences $\{\varphi_k\}$ and $\{t_k\}$ and deterministic constants δ and M, then the stochastic case can be proved by applying the result for each sample path. So, without lose of generality, we can assume that all the quantities appearing in the lemma are deterministic in the following proof.

Let us first establish the upper bound for the subsequence $\{P_{t_n+1}\}$

To this end, we introduce an auxiliary stochastic system

$$x_{k+1} = x_k + Q^{1/2} \eta_k^1 \qquad (A3)$$

$$z_k = \varphi_k^\tau x_k + (R)^{1/2} \eta_k^2 \qquad (A4)$$

where $\{ \eta_k^{1\tau} , \eta_k^2 \}$ is an i.i.d. Gaussian random sequence with zero mean and unity covariance. Assume further that $var(x_0) = P_0$ and x_0 is independent of $\{\eta_k^{1\tau}, \eta_k^2 \}$.

Denote $\hat{x}_k$ the estimation for x_k based on $\{z_0, ..., z_k\}$ which is given by the Kalman filter, then it is well known that (e.g. Anderson and Moore [9]) P_k defined by (1.3b) (or (2.2b) with F = I) can be represented by

$$P_{k+1} = \Sigma_k + Q, \quad \text{for any } k \geq 0. \qquad (A5)$$

where

$$\Sigma_k = E(x_k - \hat{x}_k)(x_k - \hat{x}_k)^\tau .$$

Let us consider another linear estimate $\hat{x}_n^*$ for x_n at time $n = t_k$ as follows

$$\hat{x}_{t_k}^* = w^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k} \varphi_i z_i , \quad W(k) \triangleq \sum_{i=t_{k-1}+1}^{t_k} \varphi_i \varphi_i^\tau .$$

Note that by (A3) and (A4),

$$x_{t_k} - \hat{x}_{t_k}^* = W^{-1}(k) \{ W(k) x_{t_k} - \sum_{i=t_{k-1}+1}^{t_k} \varphi_i z_i \}$$

$$= W^{-1}(k) \{ \sum_{i=t_{k-1}+1}^{t_k} \varphi_i \varphi_i^\tau \sum_{j=i+1}^{t_k} Q^{1/2} \eta_{j-1}^1 - \sum_{i=t_{k-1}+1}^{t_k} \varphi_i (R)^{1/2} \eta_i^2 \}$$

$$= I_1(k) + I_2(k) . \qquad (A6)$$

We now proceed to estimate the covariances of $I_1(k)$ and $I_2(k)$ as follows.

Denote

$$S_i \triangleq \sum_{j=t_{k-1}+1}^{i} \varphi_j \varphi_j^\tau , \; S_{t_{k-1}} \triangleq 0 .$$

$$T_i \triangleq \sum_{j=i}^{t_k} Q^{1/2} \eta_{j-1}^1 , \; T_{t_k+1} \triangleq 0 .$$

By summation by parts we have

$$\sum_{i=t_{k-1}+1}^{t_k} \varphi_i \varphi_i^\tau \sum_{j=i+1}^{t_k} Q^{1/2} \eta_{j-1}^1 = \sum_{i=t_{k-1}+1}^{t_k} (S_i - S_{i-1}) T_{i+1}$$

$$= \sum_{i=t_{k-1}+1}^{t_k - 1} S_i (T_{i+1} - T_{i+2}) + S_{t_k} T_{t_k+1} - S_{t_{k-1}} T_{t_{k-1}+2}$$

$$= \sum_{i=t_{k-1}+1}^{t_k - 1} S_i Q^{1/2} \eta_i^1 .$$

Then by orthogonality of $\{\eta_i^1\}$ and monotonocity of $\{S_i\}$:

$$\|E I_1(k) I_1^\tau(k)\| \leq \| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k - 1} S_i Q S_i W^{-1}(k) \|$$

$$\leq \| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k - 1} (S_i)^{1/2} S_i (S_i)^{1/2} W^{-1}(k) \| \|Q\|$$

$$\leq \| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k - 1} S_i W^{-1}(k) \| \lambda_{max}(S_{t_k}) \|Q\|$$

$$\leq \| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k - 1} S_{t_k} W^{-1}(k) \| \lambda_{max}(k) \|Q\|$$

$$\leq (t_k - t_{k-1}) \| W^{-1}(k) \| \lambda_{max}(k) \|Q\|$$

$$\leq d \|Q\| \lambda_{max}(k) / \lambda_{min}(k) \leq d M \|Q\| , \qquad (A7)$$

while for $I_2(k)$ we have

$$\|E I_2(k) I_2^\tau(k)\| = \| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k} \varphi_i R \varphi_i^\tau W^{-1}(k) \|$$

$$= R \|W^{-1}(k)\| \leq R \delta^{-1} . \qquad (A8)$$

Thus by the orthognality of $I_1(k)$ and $I_2(k)$ from (A6)---(A8) we get

$$\| E (x_{t_k} - \hat{x}_{t_k}^*) (x_{t_k} - \hat{x}_{t_k}^*)^\tau \| \leq R\delta^{-1} + dM\|Q\|.$$

From this and the optimality of the Kalman filter

$$\Sigma_{t_k} = E(x_{t_k} - \hat{x}_{t_k})(x_{t_k} - \hat{x}_{t_k})^\tau \leq E (x_{t_k} - \hat{x}_{t_k}^*)(x_{t_k} - \hat{x}_{t_k}^*)^\tau$$

the following upper bound for P_{t_k+1} follows by noting (A5) :

$$\| P_{t_k+1} \| \leq R/\delta + (1 + dM) \|Q\| , \text{for all } k \geq 1.$$

To complete the proof, we have to establish the upper bound for $\{P_n\}$.

Since $\{t_k\}$ is a sequence of strictly increasing integers, $t_k \longrightarrow \infty$, as $k \longrightarrow \infty$, then for any integer $n \geq t_1 + 1$, there exists a integer $k \geq 1$ such that

$t_k+1 \leq n \leq t_{k+1}$.

From this and the following inequality (by(1.3b) with F = I) :

$$P_{k+1} \leq P_k + Q \text{ ,for any } k \geq 0 \qquad \text{(A9)}$$

we obtain

$$\|P_n\| \leq \|P_{t_k+1}\| + (n - t_k)\|Q\|$$

$$\leq R/\delta + (1+dM)\|Q\| + (t_{k+1} - t_k)\|Q\|$$

$$\leq R/\delta + [1 + d(M+1)]\|Q\| , n \geq t_1+1, \qquad \text{(A10)}$$

while for the case where $n \leq t_1$, by (A9),

$$\|P_n\| \leq \|P_0 + t_1 Q\| = \|P_0 + (t_1 - t_0)Q\|$$

$$\leq \|P_0\| + d\,\|Q\| \qquad \text{(A11)}$$

Finally, the desired result follows by combining (A10) and (A11). #

References

[1] D.Q. Mayne, "Optimal Non-stationary Estimation of the Parameters of a Linear System With Gaussian Inputs", J. Electron. Contr. Vol. 14, p101, 1963.

[2] K.J. Astrom and B. Wittenmark "Problems of Identification and Control", J. Math. Anal. Applic. 34, pp 90-113, 1971.

[3] G. Kitagawa and W. Gersh, " A Smoothness Priors Time Varying AR Coefficient Modelling of Non-stationary Covariance Time Series" IEEE Trans. AC Vol. AC-30, pp48-56, 1985.

[4] L. Ljung, "Consistency of the Least - Squares Identification Method" IEEE Trans. AC Vol. AC-21, pp 779-781, 1976.

[5] J.B. Moore, "On Strong Consistency of Least Squares Identification Algorithm", Automatica ,Vol. 14 pp 505-509. 1978.

[6] T. L. Lai and C. Z. Wei "Least Squares Estimation in Stochastic Regression Model with Applications to Identification and Control of Dynamic Systems", Ann. Statist., 10, pp 154-166, 1982.

[7] H.F. Chen and L. Guo, "Convergence Rate of Least - Squares Identification and Adaptive Control for Stochastic Systems", Int. J. Control Vol.44, No. 5, pp 1459-1476, 1986.

[8] S.P. Meyn and P. E. Caines, "A New Approach to Stochastic Adaptive Control", IEEE Trans. AC. Vol. AC-32, pp 220-226, 1986

[9] B.D.O. Anderson and J.B. Moore, "Optimal Filtering", Prentice-Hall, Englewood Cliffs, NJ. 1979.

[10] B.D.O. Anderson and J.B. Moore,"Detectability and Stability of Discrete-time Linear Systems", SIAM J. on Control and Optimization ,Vol. 19, No. 1, pp 20-32, 1981.

[11] H.F. Chen and L. Guo, "Optimal Adaptive Control and Consistent Parameter Estimates for ARMAX Model with Quadratic Cost", SIAM J. on Control and Optimization Vol. 25, No. 4, pp 845-867, 1987.

[12] L. Guo and S.P. Meyn, "Short Memory Stochastic Adaptive Control", Submitted for publication.

[13] B.D.O. Anderson et al, "Stability of Adaptive Systems: Passivity and Averaging Analysis" Boston, Mass:MIT Press, 1986.

[14] H. F. Chen and L. Guo, "Adaptive Control via Consistent Estimation for Deterministic Systems", Int. J. Control ,Vol. 45, No. 6, pp2183-2202, 1987.

[15] V. Strassen, "An Invariance Principle for the Law of Iterated Logarithm", Z. Wahrsch. Gebiete 3, pp 211-226, 1964.

[16] M. D. Donsker and S. R. S. Varadhan, "On Laws of the Iterated Logarithm for Local Times" Comm. Pure. Appl. Math 30, pp 707-753, 1977.