Math. Control Signals Systems (1991) 4: 1–16

Mathematics of Control, Signals, and Systems

Tracking Randomly Varying Parameters: Analysis of a Standard Algorithm*

L. Guo,† L. Xia,† and J. B. Moore†

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, and some results on sample path deviations of the estimates.

Key words. Randomly varying parameters, Tracking error bounds, Kalman filter, Large deviations.

1. Introduction

Let us first define the signal model class and estimation algorithm.

Signal Model. Consider the following linear regression model:

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

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

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

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

(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 $\mathscr{F}_{k-1}$—the σ-algebra generated by $\{y_0, y_1, \ldots, y_{k-1}\}$.

* Date received: February 16, 1988. Date revised: December 6, 1988.

† Department of Systems Engineering, R. S. Phys. S, Australian National University, G.P.O. Box 4, Canberra, ACT 2601, Australia. L. Guo is now with Academica Sinica, Beijing, People's Republic of China.

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 covariance 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.

Estimation Algorithm (Kalman Filter). Consider the following estimation algorithm associated with (1.1):

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

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

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, Q must be chosen to be nonsingular to achieve a "short-memory" algorithm, meaning that the adaptation gain in (1.3a) does not diminish to zero.)

It is known that if the noise source $\{w_k, v_k\}$ 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.,

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

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 [M1] and later expanded on by various authors, e.g., Åström and Wittenmark [AW] and Kitagawa and Gersch [KG].

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 stochastic Lyapunov function technique, as well as the traditional martingale limit approach used in the convergence analyses of both the least-squares (e.g., [L], [M2], [LW], and [CG1]) and the stochastic-gradient (e.g., [CG2]) algorithms, cannot be directly used in the time-varying case. This is so even though (1.3) is the standard least-squares algorithm when $F = I$, $Q = 0$, and $R = 1$. Similar observations are also made in [MC]. (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., [AM1]), 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 for all k (e.g., [AM2]). This requirement turns out to be unrealistic for 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 asymptotic properties of the Kalman filter when it is used as a parameter estimator for stochastic linear regression models. The main contributions of the paper are the following:

(i) For the case where the parameters are generated from a stable linear model, i.e., (1.1b) with $|\lambda_i(F)| < 1$, upper bounds for the averaged tracking errors are established *without* any excitation requirement.
(ii) In the cases of drifting parameters, i.e., (1.1b) with $F = I$, and disturbed parameters, i.e., $\theta_k = \theta_0 + w_k$, averaged tracking-error bounds are also established under some excitation conditions on the regressors $\{\varphi_k\}$.
(iii) If instead of the averaged values, the tracking error itself is concerned, large deviations of the estimates may occur, and the rate for these deviations is also established.

2. Tracking-Error Bound

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

Let us first denote

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

and rewrite (1.3) as

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

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

The lower bound 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* $\mathrm{Sup}_k\, E\|w_k\|^{2+\varepsilon} < \infty$ *for some* $\varepsilon > 0$, *then*

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

and

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

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

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

Set $f_k = (F - K_k \varphi_k^\tau)\tilde{\theta}_k - K_k v_k$; then $\{f_k^\tau w_{k+1}\}$ is a martingale difference sequence with respect to the σ-algebra generated by $\{v_{i-1}, w_i, i \le k+1\}$, so the first assertion (2.3) follows from (2.5) and the orthogonality of f_k and w_{k+1}.

Now, by an estimate for the weighted sum of martingale difference sequences (e.g., [LW]), we know that

$$\sum_{i=1}^{n} f_i^\tau w_{i+1} = O\left(\left\{\sum_{i=1}^{n} \|f_i\|^2\right\}^{(1/2)+\eta}\right) \quad \text{a.s.} \qquad \text{for all} \quad \eta \in (0, \tfrac{1}{2}).$$

Consequently, it follows from (2.5) that

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \|\tilde{\theta}_i\|^2 &= \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \{\|f_i\|^2 + 2f_i^\tau w_{i+1} + \|w_{i+1}\|^2\} \\ &\ge \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \|w_{i+1}\|^2 \ge \operatorname{tr}(Q_w) \quad \text{a.s.,} \end{aligned}$$

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

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

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

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

$$P_k \ge Q > 0 \qquad \text{for any} \quad k \ge 1. \tag{2.7}$$

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 (2.6).

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

$$\sup_{k\ge 0} \|P_k\| \le b < \infty \quad \textit{a.s.} \tag{2.8}$$

Then, for K_k defined by (2.1),

$$\left\|\prod_{k=i}^{j-1} (F - K_k \varphi_k^\tau)\right\| \le \beta\alpha^{j-i} \quad \textit{a.s.,} \qquad \textit{for any} \quad j > i \ge 0, \tag{2.9}$$

where α and β are defined by

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

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

The proof is given in the Appendix. The key point here is the precise expressions for α and β in (2.10) and (2.11), which lead directly to the following 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 (albeit φ_k is random!). This fact is crucial in establishing the upper bound for the tracking errors in terms of mathematical expectations in the following.

We now establish an 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),

$$P_{k+1} \le \sum_{i=0}^{k} F^i Q (F^\tau)^i + F^{k+1} P_0 (F^\tau)^{k+1} \qquad \text{for any} \quad k \ge 0, \tag{2.12}$$

and hence

$$b \triangleq \left\| \sum_{i=0}^{\infty} F^i Q (F^\tau)^i \right\| + \sup_{k\ge 0} \|F^k P_0 (F^\tau)^k\| \tag{2.13}$$

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*

$$\limsup_{n\to 0} \frac{1}{n} \sum_{i=1}^{n} \|\tilde{\theta}_i\|^p \le \left(\frac{\beta}{1-\alpha}\right)^p \left[L_p(w) + \|F\| \left(\frac{b}{R}\right)^{1/2} L_p(v) \right]^p \tag{2.14}$$

and

$$\limsup_{n\to 0} E\|\tilde{\theta}_n\|^p \le \left(\frac{\beta}{1-\alpha}\right)^p \left[M_p(w) + \|F\| \left(\frac{b}{R}\right)^{1/2} M_p(v) \right]^p, \tag{2.15}$$

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*

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

$$L_p(w) \triangleq \limsup_{n\to 0} \left\{ \frac{1}{n} \sum_{i=1}^{n} \|w_i\|^p \right\}^{1/p} < \infty \quad a.s., \tag{2.17}$$

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

$$M_p(w) \triangleq \sup_i \{E\|w_i\|^p\}^{1/p} < \infty, \tag{2.18b}$$

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

Proof. By (2.5) we have

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

Applying Lemma 2.1 we see that

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

Applying the Minkowski inequality gives

$$\left(\sum_{k=0}^{n} \|\tilde{\theta}_{k+1}\|^p\right)^{1/p} \le \beta\|\tilde{\theta}_0\|\left(\sum_{k=0}^{n} \alpha^{p(k+1)}\right)^{1/p} + \beta\left\{\sum_{k=0}^{n}\left(\sum_{i=0}^{k} \alpha^{k-i}\|K_i v_i\|\right)^p\right\}^{1/p} + \beta\left\{\sum_{k=0}^{n}\left(\sum_{i=0}^{k} \alpha^{k-i}\|w_{i+1}\|\right)^p\right\}^{1/p}. \tag{2.20}$$

By the Hölder inequality it follows that $(1/p + 1/q = 1)$

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

and then

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

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

$$\|K_k\|^2 \le \|F\|^2 \frac{\varphi_k^\tau P_k^2 \varphi_k}{(R + \varphi_k^\tau P_k \varphi_k)^2} \le \|F\|^2 b \frac{\varphi_k^\tau P_k \varphi_k}{(R + \varphi_k^\tau P_k \varphi_k)^2} \le \frac{\|F\|^2 b}{R} \quad \text{for any} \quad k \ge 0, \tag{2.22}$$

which together with (2.16) and (2.21) yields

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

Similarly,

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

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

$$(E\|\tilde{\theta}_{k+1}\|^p)^{1/p} \le \beta\alpha^{k+1}(E\|\tilde{\theta}_0\|^p)^{1/p} + \beta\left\{E\left(\sum_{i=0}^{k}\alpha^{k-i}\|K_i v_i\|\right)^p\right\}^{1/p}$$
$$+ \beta\left\{E\left(\sum_{i=0}^{k}\alpha^{k-i}\|w_{i+1}\|\right)^p\right\}^{1/p}.$$

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 nonzero mean.

Remark 2.3. We have recently applied the property (2.15), with $p = 4 + \delta$ for some $\delta > 0$, to adaptive control problems [GM], and it appears that the nontrivial stochastic adaptive control problem considered in [MC] 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 for $\{P_k\}$ 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, take $\varphi_k = 0$ for all $k \ge 0$, then, by (1.3b),

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

Nevertheless, we have the following results.

Lemma 2.2. *Assume that there exists a strictly increasing sequence of* random *integers $\{t_n\}$ with $t_0 = 0$, $d \triangleq \sup_n(t_n - t_{n-1}) < \infty$ a.s., and a* random *constant $\delta > 0$ such that, for any $k \ge 1$,*

$$\sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i\varphi_i^\tau}{1 + \|\varphi_i\|^2} \ge \delta I \quad a.s. \tag{2.26}$$

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

$$\sup_k \|P_k\| \le \|P_0\| + (1 + d)\|Q\| + (R + d^2\|Q\|)/\delta \quad (\triangleq b) \quad a.s. \tag{2.27}$$

The proof of this lemma is given in the Appendix. Condition (2.26) can be regarded as requiring certain kinds of excitations; thus divergence phenomena as in (2.25) may be explained as arising from lack of excitation.

Remark 2.5. Lemma 2.2 can be generalized to the case where $F \neq I$, and a similar bound is obtained. Results like the following theorem (Theorem 2.3) are also available. However, the matrix on the left-hand side of (2.26) will involve the matrix F in general.

Remark 2.6. Let $\{t_n\}$ be a sequence of increasing *random* integers with $t_0 = 0$, $\sup_n(t_n - t_{n-1}) < \infty$. If

$$\inf_k \lambda_{\min}(k) > 0 \quad \text{a.s.} \tag{2.28}$$

and

$$\sup_k \frac{\lambda_{\max}(k)}{\lambda_{\min}(k)} < \infty \quad \text{a.s.,} \tag{2.29}$$

where $\lambda_{\max}(k)$ and $\lambda_{\min}(k)$ denote the maximum and minimum eigenvalues of the matrix $\sum_{i=t_{k-1}+1}^{t_k} \varphi_i \varphi_i^\tau$, respectively, then (2.26) holds.

It is interesting to compare conditions (2.26) or (2.28)–(2.29) with the standard persistence of excitation condition used in the analysis of short-memory adaptive-control algorithms in the literature (e.g., [ABJ]). That is, there exist constants $0 < \delta_1 \le \delta_2 < \infty$ and $N < \infty$ such that

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

This implies that $\{\varphi_k\}$ is a bounded sequence. Clearly, condition (2.28)–(2.29) 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, take φ_k as a scalar, linear function: $\varphi_k = ck$, $c \neq 0$. Then, clearly (2.30) fails, while (2.28)–(2.29) still holds because $\lambda_{\max}(k)$ and $\lambda_{\min}(k)$ coincide in this case.

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

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

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 right-hand side of (2.29) is deterministic. ■

Remark 2.7. An alternative excitation condition which involves the use of conditional expectation, may also be used to establish the tracking error bounds. To be precise, assume that there exist two *deterministic* constants $\delta > 0$ and $N < \infty$ such that

$$E\left\{\sum_{i=mN}^{(m+1)N-1} \frac{\varphi_i\varphi_i^\tau}{1+\|\varphi_i\|^2} | \mathscr{E}_{mN-1}\right\} \geq \delta I \quad \text{a.s.,} \qquad \text{for all} \quad m \geq 0,$$

where $\{\mathscr{E}_k\}$ is any nondecreasing family of σ-algebras such that φ_k is $\mathscr{E}_k$-measurable for any k. Then finite upper bounds for both $E\|\tilde{\theta}_k\|^2$ and $(1/n)\sum_{i=1}^n \|\tilde{\theta}_i\|$ can also be established if $\{w_k, v_k\}$ satisfies (2.16)–(2.18) for some $p > 4$. For details see [G].

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

$$\theta_n = \theta_{n-1} + w_n = \theta_0 + S_n, \qquad S_n = \sum_{i=1}^n w_i. \tag{2.32}$$

Consequently, by a result in [DV, p. 751],

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

Hence with probability 1, the averaged value of parameters

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

has at least a *divergence* rate of $n/\log\log n$ as $n \to \infty$. However, from (2.31) it is known that the averaged value of the *tracking error* $(1/n)\sum_{i=1}^n \|\tilde{\theta}_i\|^2$ is bounded. This shows that the estimation algorithm (1.3) can indeed perform the nontrivial 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

$$\theta_k = \theta_0 + w_k \tag{2.33}$$

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* $F = I$. *Assume that conditions of Lemma* 2.2 *apply. Then*

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

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 $\tilde{\theta}_{k+1} = (I - K_k\varphi_k^\tau)\tilde{\theta}_k - K_k v_k - K_k\varphi_k^\tau w_k$. Note that by (A.1) in the Appendix, $K_k\varphi_k^\tau$ is bounded by $\|K_k\varphi_k^\tau\| \le b/a$. Hence, an argument like that used in the proof of (2.14) leads to the desired result (2.34). ■

3. Deviations of the Estimates

A major difference between the algorithm (1.3) and the standard least-squares algorithm is the introduction of a nonzero term Q in (1.3b), which prevents P_k from tending to zero, and hence guarantees (1.3) to be a short-memory algorithm. This is why the estimator (1.3) can track time-varying parameters in a certain sense. However, due to the short-memory property of the algorithm, large deviations of the estimates may occur. Indeed, in almost all cases, for a fixed $L > 0$, there is a probability of 1 that the norm of the estimates has exceeded L after a sufficiently large time. In the work of [CFM] the mean exit-time problem for a stationary Markov chain, which is produced by a constant gain Robbins–Monro algorithm, is studied. Related problems for adaptive control algorithms are also considered in [BAG]. Here we are chiefly concerned with the asymptotic properties of the estimation error $\tilde{\theta}_k$ itself, rather than its averaged values as in the previous section, especially the possible divergence rates of $\tilde{\theta}_k$.

It is easy to see that in the study of large deviations for stochastic algorithms, the case where the noises are possibly unbounded is of major interest. Henceforth we make the following assumption:

(A) The noise sources $\{w_k\}$ and $\{v_k\}$ are mutually independent i.i.d. random sequences with zero mean and nonzero variances as defined in (1.2), and the tail probability of $\{w_k\}$ does not vanish "too fast" in the sense that

$$\liminf_{x\to\infty} \exp\{hx^2\} P(\|w_1\| \ge x) \ne 0$$

for some constant $h > 0$.

Obviously, this assumption includes a large class of random sequences, in particular the gaussian case. By the Borel–Cantelli lemma, it is easy to obtain the following result.

Lemma 3.1. *If* $\{w_k\}$ *satisfies conditions in assumption* (A), *then*

$$\limsup_{n\to\infty} \frac{\|w_n\|}{(\log n)^{1/2}} \ge \frac{1}{h^{1/2}} \quad a.s.$$

We also need the following result which can be proven by a technique similar to that in [CT, p. 135].

Lemma 3.2. *Let $\{x_i\}$ be an i.i.d. sequence and let $\{\tau_i\}$ be any sequence of finite stopping times such that $1 \le \tau_i < \tau_{i+1}$, a.s., for $i \ge 1$, with $\{x_i\}$ and $\{\tau_i\}$ independent. Then $\{x_{\tau_i}\}$ is also an i.i.d. sequence, having the same distribution as x_1.*

Theorem 3.1. *Consider the signal model* (1.1) *with the noises $\{w_n\}$ and $\{v_n\}$ satisfying assumption* (A). *If* either $|\lambda_i(F)| < 1$, *for all i*, or $F = I$ *but with conditions in Lemma 2.2 satisfied, then the estimation error $\tilde{\theta}_k = \theta_k - \hat{\theta}_k$ produced by* (1.3) *has the following property*:

$$P\left\{\limsup_{n\to\infty} \frac{\|\tilde{\theta}_n\|}{(\log n)^{1/2}} \ge \frac{c}{h^{1/2}}\right\} = 1, \tag{3.1}$$

where $c = [1 + \|F\|(1 + b/a)]^{-1}$, $a = \lambda_{\min}(Q)$, b is given by (2.13) *or* (2.27) *according to $|\lambda_i(F)| < 1$ or $F = I$, and h is given in assumption* (A).

Proof. By the error equation (2.5) and the upper bound for $K_k\varphi_k^\tau$ as provided by (A.1) in the Appendix, it follows that

$$\begin{aligned}
&\limsup_{k\to\infty} \frac{c}{(\log k)^{1/2}} \|w_{k+1} - K_k v_k\| \\
&\le \limsup_{k\to\infty} \frac{c}{(\log k)^{1/2}} \|\tilde{\theta}_{k+1} - (F - K_k\varphi_k^\tau)\tilde{\theta}_k\| \\
&\le \limsup_{k\to\infty} c\left[\frac{\log(k+1)}{\log k}\right]^{1/2} \left\{\frac{\|\tilde{\theta}_{k+1}\|}{[\log(k+1)]^{1/2}}\right\} \\
&\quad + c\limsup_{k\to\infty} \|F - K_k\varphi_k^\tau\| \frac{\|\tilde{\theta}_k\|}{[\log k]^{1/2}} \\
&\le \limsup_{k\to\infty} c\frac{\|\tilde{\theta}_{k+1}\|}{[\log(k+1)]^{1/2}} + c\|F\|\left(1 + \frac{b}{a}\right)\limsup_{k\to\infty} \frac{\|\tilde{\theta}_k\|}{[\log k]^{1/2}} \\
&\le \limsup_{k\to\infty} \frac{\|\tilde{\theta}_k\|}{[\log k]^{1/2}}.
\end{aligned} \tag{3.2}$$

Now, let us define a sequence of increasing stopping times $\{\tau_i\}$ as

$$\tau_{k+1} = \inf\{n > \tau_k : \|v_n\| \le 2(R_v)^{1/2}\}, \qquad \tau_1 = 1. \tag{3.3}$$

Since

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} v_i^2 = R_v \quad \text{a.s.},$$

it is easy to see that $\{\tau_i\}$ is a sequence of finite stopping times and is independent of $\{w_k\}$. We now show that

$$\liminf_{n\to\infty} \frac{n}{\tau_n} > 0 \quad \text{a.s.} \tag{3.4}$$

Otherwise, if (3.4) did not hold, we would have the following contradiction:

$$R_v = \lim_{n\to\infty} \frac{1}{\tau_n} \sum_{i=1}^{\tau_n} v_i^2 = \limsup_{n\to\infty} \frac{1}{\tau_n} \sum_{i=1}^{\tau_n} v_i^2$$

$$\geq \limsup_{n\to\infty} \frac{1}{\tau_n} \sum_{i=1}^{\tau_n} v_i^2 I_{[|v_i| > 2(R_v)^{1/2}]} \geq 4R_v \limsup_{n\to\infty} \frac{1}{\tau_n} \sum_{i=1}^{\tau_n} I_{[|v_i| > 2(R_v)^{1/2}]}$$

$$\geq 4R_v \limsup_{n\to\infty} \frac{1}{\tau_n} [\tau_n - n] \geq 4R_v \left[1 - \liminf_{n\to\infty} \frac{n}{\tau_n} \right] = 4R_v,$$

and so (3.4) holds.

Now, by (3.2)–(3.4) and the upper bound for K_k provided by (2.22), it follows that

$$\limsup_{k\to\infty} \frac{\|\tilde{\theta}_k\|}{[\log k]^{1/2}} \geq \limsup_{k\to\infty} \frac{c}{(\log k)^{1/2}} \left\{ \|w_{k+1}\| - \|F\| \left(\frac{b}{R}\right)^{1/2} \|v_k\| \right\}$$

$$\geq \limsup_{k\to\infty} \frac{c}{[\log \tau_k)^{1/2}} \left\{ \|w_{\tau_k+1}\| - 2\|F\| \left(\frac{bR_v}{R}\right)^{1/2} \right\}$$

$$\geq c \limsup_{k\to\infty} \frac{\|w_{\tau_k+1}\|}{(\log \tau_k)^{1/2}} = c \limsup_{k\to\infty} \left(\frac{\log k}{\log \tau_k}\right)^{1/2} \frac{\|w_{\tau_k+1}\|}{(\log k)^{1/2}}$$

$$\geq c \liminf_{k\to\infty} \left(\frac{\log k}{\log \tau_k}\right)^{1/2} \limsup_{k\to\infty} \frac{\|w_{\tau_k+1}\|}{(\log k)^{1/2}} \geq \frac{c}{h^{1/2}} \quad \text{a.s.}$$

Here we have used (3.4) and Lemmas 3.1 and 3.2. This proves the theorem. ■

The above theorem tells us that there is a subsequence of $\{\|\tilde{\theta}_k\|\}$ which diverges at least at the rate of $O(\{\log n\}^{1/2})$. However, we note that under the conditions of the theorem, the divergence rate of such a subsequence is also dominated by that of the noise sequence, because, by (2.19),

$$\max_{0\leq k\leq n} \|\tilde{\theta}_k\| = O\left(\max_{0\leq k\leq n} \{\|v_k\| + \|w_k\|\} \right).$$

Remark 3.1. Various generalizations and variations of the above theorem are readily obtained. In particular, we point out that if in assumption (A) the conditions on $\{w_k\}$ and those on $\{v_k\}$ are interchanged, then under a mild regularity condition on the regressors $\{\varphi_k\}$, results like Theorem 3.1 still hold. This can be proved by an approach similar to that in [BAG].

4. 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 needed 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 this 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. Finally, deviation phenomena of the estimates have been studied here, which will provide valuable information for adaptive controller design for time-varying systems.

Appendix

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):

$$\begin{aligned}\|K_k\varphi_k^\tau\| &\le \frac{\|F\|\,\|P_k\|\,\|\varphi_k\|^2}{R + \varphi_k^\tau P_k \varphi_k} \\ &\le \frac{\|F\|b\|\varphi_k\|^2}{[\lambda_{\min}(Q)\|\varphi_k\|^2]} \qquad \text{(by (2.7))} \\ &\le \|F\|\frac{b}{a}. \end{aligned} \tag{A.1}$$

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

$$\|F_k\| \le \|F\|\left(1 + \frac{b}{a}\right). \tag{A.2}$$

Now consider the following inequalities. By (A.2), (2.2b), and the matrix inversion lemma,

$$\begin{aligned} P_k^{-1} - F_k^\tau P_{k+1}^{-1} F_k &= P_k^{-1} - F_k^\tau [F_k P_k F_k^\tau + K_k R K_k^\tau + Q]^{-1} F_k \\ &\ge P_k^{-1} - F_k^\tau [F_k P_k F_k^\tau + Q]^{-1} F_k \\ &= [P_k + P_k F_k^\tau Q^{-1} F_k P_k]^{-1} \\ &\ge [P_k + (P_k)^{1/2}\|(P_k)^{1/2} F_k^\tau Q^{-1} F_k (P_k)^{1/2}\|(P_k)^{1/2}]^{-1} \\ &\ge \left[P_k + \|F\|^2\left(1 + \frac{b}{a}\right)^2 \frac{b}{a} P_k\right]^{-1} \\ &= \left[1 + \|F\|^2\left(1 + \frac{b}{a}\right)^2 \frac{b}{a}\right]^{-1} P_k^{-1}. \end{aligned}$$

Consequently, by definition (2.10) for α,

$$F_k^\tau P_{k+1}^{-1} F_k \le \alpha^2 P_k^{-1} \qquad \text{for any} \quad k \ge 0.$$

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

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

$$\le b\alpha^{2(j-i)}\|P_i^{-1}\| \le \left(\frac{b}{a}\right)\alpha^{2(j-i)} \qquad \text{for any} \quad j > i \ge 0. \qquad \blacksquare$$

Proof of Lemma 2.2. Clearly, if the result holds for any deterministic sequences $\{\varphi_k\}$ and $\{t_k\}$ and deterministic constant δ, then the stochastic case can be proved by applying the result for each sample path. So, without loss of generality, we assume that all the quantities appearing in the lemma are deterministic in the following analysis.

Let us first establish the upper bound for the subsequence $\{P_{t_n+1}\}$. To this end, we introduce an auxiliary stochastic system

$$x_{k+1} = x_k + Q^{1/2}\eta_k^1, \tag{A.3}$$

$$z_k = \varphi_k^{\tau} x_k + (R)^{1/2}\eta_k^2, \tag{A.4}$$

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 $\text{var}(x_0) = P_0$ and x_0 is independent of $\{\eta_k^{1\tau}, \eta_k^2\}$.

Denote by $\hat{x}_k$ the estimate for x_k based on $\{z_0, \ldots, z_k\}$ which is given by the Kalman filter. Then it is well known (e.g., [AM1]) that P_k defined by (1.3b) (or (2.2b) with $F = I$) can be represented by

$$P_{k+1} = \Sigma_k + Q \qquad \text{for any} \quad k \ge 0, \tag{A.5}$$

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:

$$\hat{x}_{t_k}^* = W^{-1}(k)\sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i z_i}{1 + \|\varphi_i\|^2}, \qquad W(k) \triangleq \sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i\varphi_i^{\tau}}{1 + \|\varphi_i\|^2}.$$

Note that by (A.3) and (A.4),

$$\begin{aligned} x_{t_k} - \hat{x}_{t_k}^* &= W^{-1}(k)\left\{W(k)x_{t_k} - \sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i z_i}{1 + \|\varphi_i\|^2}\right\} \\ &= W^{-1}(k)\left\{\sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i\varphi_i^{\tau}}{1 + \|\varphi_i\|^2}\sum_{j=i+1}^{t_k} Q^{1/2}\eta_{j-1}^1 - \sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i (R)^{1/2}\eta_i^2}{1 + \|\varphi_i\|^2}\right\} \\ &= I_1(k) + I_2(k). \end{aligned} \tag{A.6}$$

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

$$S_i \triangleq \sum_{j=t_{k-1}+1}^{i} \frac{\varphi_j\varphi_j^{\tau}}{1 + \|\varphi_j\|^2}.$$

By interchanging the order of summation we have

$$\sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i\varphi_i^{\tau}}{1 + \|\varphi_i\|^2}\sum_{j=i+1}^{t_k} Q^{1/2}\eta_{j-1}^1 = \sum_{j=t_{k-1}+1}^{t_k-1} S_j Q^{1/2}\eta_j^1.$$

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

$$\begin{aligned}\|EI_1(k)I_1^\tau(k)\| &\le \left\| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k-1} S_i Q S_i W^{-1}(k) \right\| \\ &\le \left\| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k-1} (S_i)^{1/2} S_i (S_i)^{1/2} W^{-1}(k) \right\| \|Q\| \\ &\le \left\| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k-1} S_i W^{-1}(k) \right\| \|S_{t_k}\| \|Q\| \\ &\le \left\| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k-1} S_{t_k} W^{-1}(k) \right\| (t_k - t_{k-1}) \|Q\| \\ &\le (t_k - t_{k-1})^2 \|W^{-1}(k)\| \|Q\| \le d^2 \|Q\|/\delta, \end{aligned} \tag{A.7}$$

while for $I_2(k)$ we have

$$\begin{aligned}\|EI_2(k)I_2^\tau(k)\| &= \left\| W^{-1}(k) \sum_{i=t_{k-1}+1}^{t_k} \frac{\varphi_i R \varphi_i^\tau}{(1 + \|\varphi_i\|^2)^2} W^{-1}(k) \right\| \\ &\ge R \|W^{-1}(k)\| \le R\delta^{-1}. \end{aligned} \tag{A.8}$$

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

$$\|E(x_{t_k} - \hat{x}_{t_k}^*)(x_{t_k} - \hat{x}_{t_k}^*)^\tau\| \le R + d^2\|Q\|)/\delta.$$

From this and the optimality of the Kalman filter

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

the following upper bound for P_{t_k+1} follows by noting (A.5):

$$\|P_{t_k+1}\| \le \|Q\| + (R + d^2\|Q\|)/\delta \qquad \text{for all} \quad k \ge 1.$$

To complete the proof, we have to establish the upper bound for $\{P_n\}$. Since $\{t_k\}$ is a sequence of strictly increasing integers, for any integer $n \ge t_1 + 1$, there exists an integer $k \ge 1$ such that $t_k + 1 \le n \le t_{k+1}$. From this and the following inequality (by (1.3b) with $F = I$),

$$P_{k+1} \le P_k + Q \qquad \text{for any} \quad k \ge 0, \tag{A.9}$$

we obtain

$$\begin{aligned}\|P_n\| &\le \|P_{t_k+1}\| + (n - t_k)\|Q\| \le \|P_{t_k+1}\| + (t_{k+1} - t_k)\|Q\| \\ &\le (1 + d)\|Q\| + (R + d^2\|Q\|)/\delta, \qquad n \ge t_1 + 1, \end{aligned} \tag{A.10}$$

while for the case where $n \le t_1$, by (A.9),

$$\|P_n\| \le \|P_0 + t_1 Q\| = \|P_0 + (t_1 - t_0)Q\| \le \|P_0\| + d\|Q\|.$$

The desired result follows by combining this with (A.10). ■

References

[ABJ] B. D. O. Anderson, R. R. Bitmead, C. R. Johnson, Jr., P. V. Kokotovic, Kosut, I. M. Y. Mareels, L. Praly, and B. D. Riedle, *Stability of Adaptive Systems: Passivity and Averaging Analysis*, MIT Press, Cambridge, MA, 1986.

[AM1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[AM2] B. D. O. Anderson and J. B. Moore, Detectability and stability of discrete-time linear systems, *SIAM J. Control Optim.*, **19** (1981), 20–32.

[AW] K. J. Åström and B. Wittenmark, Problems of identification and control, *J. Math. Anal. Appl.* **34** (1971), 90–113.

[BAG] R. R. Bitmead, B. D. O. Anderson, and L. Guo, Distributional properties of adaptive estimators, *Proceedings of the 1988 IFAC Workshop on Robust Adaptive Control*, Newcastle, Australia, pp. 20–25.

[CG1] H. F. Chen and L. Guo, Convergence rate of least-squares identification and adaptive control for stochastic systems, *Internat. J. Control*, **44** (1986), 1459–1476.

[CG2] H. F. Chen and L. Guo, Adaptive control with recursive identification for stochastic linear systems, in *Control and Dynamic Systems—Advances in Theory and Applications* (C. T. Leondes, ed.), Vol. 26, pp. 277–327, Academic Press, New York, 1987.

[CT] Y. S. Chow and H. Teicher, *Probability Theory: Independency, Interchangability and Martingales*, Springer-Verlag, New York, 1978.

[CFM] M. Cottrell, J. C. Fort, ad G. Malgouyres, Large deviations and rare events in the study of stochastic algorithms, *IEEE Trans. Automat. Control*, **28** (1983), 907–920.

[DV] M. D. Donsker and S. R. S. Varadhan, On laws of the iterated logarithm for local times, *Comm. Pure Appl. Math*, **30** (1977), 707–753.

[G] L. Guo, Estimating time-varying parameters by a Kalman-filter-based algorithm: Stability and convergence, 1988, submitted for publication.

[GM] L. Guo and S. P. Meyn, Adaptive control for time-varying systems: A combination of martingale and Markov chain techniques, *Internat. J. Adaptive Control Signal Process*, **3** (1989), 1–14.

[KG] G. Kitagawa and W. Gersh, A smoothness priors time varying AR coefficient modelling of non-stationary covariance time series, *IEEE Trans. Automat. Control*, **30** (1985), 48–56.

[LW] T. L. Lai and C. Z. Wei, Least squares estimation in stochastic regression model with applications to identification and control of dynamic system, *Ann. Statist.*, **10** (1982), 154–166.

[L] L. Ljung, Consistency of the least-squares identification method, *IEEE Trans. Automat. Control*, **21** (1976), 779–781.

[M1] D. Q. Mayne, Optimal nonstationary estimation of the parameters of a linear system with Gaussian input, *J. Electron. Control*, **14** (1963), 101.

[MC] S. P. Meyn and P. E. Caines, A new approach to stochastic adaptive control, *IEEE Trans. Automat. Control*, **32** (1986), 220–226.

[M2] J. B. Moore, On strong consistency of least squares identification algorithm, *Automatica*, **14** (1978), 505–509.