Performance Analysis of General Tracking Algorithms

Lei Guo and Lennart Ljung, *Fellow, IEEE*

Abstract—**A general family of tracking algorithms for linear regression models is studied. It includes the familiar least mean square gradient approach, recursive least squares, and Kalman filter based estimators. The exact expressions for the quality of the obtained estimates are complicated. Approximate, and easy-to-use, expressions for the covariance matrix of the parameter tracking error are developed. These are applicable over the whole time interval, including the transient, and the approximation error can be explicitly calculated.**

I. Introduction

TRACKING is the key factor in adaptive algorithms of all kinds. In this contribution we shall study the special case where the underlying model is a linear regression, i.e., the observations are related by

$$y_k = \varphi_k^\tau \theta_k + v_k, \qquad k \geq 0. \tag{1}$$

Here y_k is an observation made at time k, and φ_k is a d-dimensional vector that is known at time k, v_k represents a disturbance, and the parameter vector θ_k describes how the components of φ_k relate to the observation y_k. It is the objective to estimate the vector θ_k from measurements $\{y_t, \varphi_t, t \leq k\}$.

Many technical problem formulations fit structure (1) by choosing φ_k and y_k appropriately. See, for example, [15] and [22].

To come up with good algorithms for estimating θ_k, it is natural to introduce some assumptions about the time-variation of this parameter vector. In general we may write

$$\theta_k = \theta_{k-1} + \gamma w_k \tag{2}$$

where γ is a scaling constant and w_k is an as yet undefined variable.

The tracking algorithms will provide us with an estimate

$$\hat{\theta}_k = \hat{\theta}_k(y^k, \varphi^k, \theta^k) \tag{3}$$

where superscript denotes the whole time history: $y^k = \{y_0, y_1, \cdots y_k\}$, etc.

Manuscript received February 18, 1994. Recommended by Past Associate Editor, B. Pasik-Duncan. This work was supported by the National Natural Science Foundation of China and the Swedish Research Council for Engineering Sciences (TFR).

L. Guo is with the Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100080, China.

L. Ljung is with the Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.

IEEE Log Number 9412439.

A prime question concerns the quality of such an estimate. We shall evaluate the quality in terms of the covariance matrix of the tracking error

$$\tilde{\theta}_k = \theta_k - \hat{\theta}_k. \tag{4}$$

This covariance matrix will be denoted by

$$\Pi_k^0 = E[\tilde{\theta}_k \tilde{\theta}_k^\tau] \tag{5}$$

where expectation will be taken over all relevant stochastic variables. A precise definition will be given later.

An exact expression for Π_k^0 will be very complicated—except in some trivial cases—and it will not be possible to derive it explicitly in closed form. The practical importance of having good tracking algorithms and estimates of their quality, however, still makes it vital to be able to work with Π_k^0.

For that reason, there is a quite substantial literature on the problem of how to approximate Π_k^0 with expressions Π_k that are simple to work with. This literature is—partly—surveyed in [1], [2], [12], and [20].

The current paper has the ambition to give a general result that subsumes and extends most of the earlier results.

Example 1.1—A Preview Example: Consider model (1) and (2) under the assumptions that

a) φ_k and θ_k are scalars;
b) $\{\varphi_k\}, \{v_k\}$, and $\{w_k\}$ are independent sequences of independent random variables with zero mean values and variances R_φ, R_v, and Q_w, respectively.
c) The fourth moment of φ_k is R_4.

Assume also that the estimate $\hat{\theta}_k$ is computed by the simple least mean square (LMS) algorithm

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu\varphi_k(y_k - \varphi_k\hat{\theta}_k). \tag{6}$$

This case is one—essentially the only one—where a simple exact expression for Π_k^0 can be calculated. Straightforward calculations give

$$\tilde{\theta}_{k+1} = (1 - \mu\varphi_k^2)\tilde{\theta}_k - \mu\varphi_k v_k + \gamma w_{k+1}. \tag{7}$$

Squaring and taking expectations gives

$$\Pi_{k+1}^0 = (1 - 2\mu R_\varphi + \mu^2 R_4)\Pi_k^0 + \mu^2 R_\varphi R_v + \gamma^2 Q_w. \tag{8}$$

This is a linear time-invariant difference equation for Π_k^0 and can be explicitly solved. In particular, if

$$|1 - 2\mu R_\varphi + \mu^2 R_4| < 1$$

the solution of (8) will converge to Π^* with

$$\Pi^* = \frac{1}{1-\mu R_4/(2R_\varphi)}\Pi, \quad \Pi = \frac{1}{2R_\varphi}\left[\mu R_\varphi R_v + \frac{\gamma^2}{\mu}Q_w\right]. \tag{9}$$

Simple manipulations then give

$$|\Pi^* - \Pi| \le \sigma(\mu)\Pi, \quad \sigma(\mu) = \left[\frac{R_4/(2R_\varphi)}{1-\mu R_4/(2R_\varphi)}\right]\mu.$$

Thus, Π^* can be well approximated by Π for small μ, since $\sigma(\mu) \to 0$ as $\mu \to 0$. □

Now, this example was particularly easy, primarily because of the assumed independence among $\{\varphi_k, v_k, w_k\}$ which makes φ_k and $\tilde{\theta}_k$ independent.

In more general cases we have to deal with dependence among $\{\varphi_k\}$, and that is actually at the root of the problem. Generally speaking, if $\{\varphi_k\}$ are weakly dependent, so should $\hat{\theta}_k$ and φ_k be, provided that $\hat{\theta}_k$ in (3) depends to a small extent on the "latest" φ_k, i.e., if the adaptation rate (μ in the example) is small and the error equation [(7) in the example] is stable.

The extra term caused by the dependence in the equation corresponding to (8) in the example should then have negligible influence. Indeed, it is the purpose of this contribution to establish this for a fairly general family of tracking algorithms. Despite the simple idea, it turns out to be surprisingly technically difficult to prove. This paper could be said to make the end of a series of results on performance analysis, starting with Theorem 1 in [12] and then followed by [14], [13], and [10]. There are many related, relevant results using other approaches. We may point to [2]–[6], [16], [18], and [20] and to the references in these books and papers.

The bottom line of the analysis is a result of the character

$$\|E[\tilde{\theta}_k\tilde{\theta}_k^\tau] - \Pi_k\| \le \sigma(\mu)\|\Pi_k\| \tag{10}$$

where $\sigma(\mu) \to 0$ as $\mu \to 0$, and μ is a measure of the adaptation rate in the algorithm, Π_k obeys a simple linear, deterministic, difference equation (like (8) without the term $\mu^2 R_4$).

The point with a result of character (10) is, clearly, that we arbitrarily can well approximate the actual tracking error covariance matrix with a simple expression that can be easily evaluated and analyzed. The essence of this paper does not lie in the expression for Π_k itself—it is not difficult to conjecture that such an approximation should be reasonable. Our contribution is rather to establish the connection in the explicit fashion (10) for a wide family of the most common tracking algorithms. One important step in achieving such results is to first establish that the underlying algorithm is exponentially stable. This is a major problem in itself, and a companion paper [9] is devoted to this step for the same family of algorithms.

The paper is organized as follows. In Section II the tracking algorithms are briefly described. Section III gives the main result: that (10) holds under the same general conditions for all algorithms in the family. There we also briefly discuss the practical consequences of the result. In the following section, a more general theorem is presented, which is the basis for the analysis. This theorem is more general and uses weaker but less explicit conditions. The proof of the main result is then given in Section V by showing that the general theorem can be applied to our family of algorithms. Notice that this analysis is of independent interest in that the conditions can be somewhat weakened in different ways for each individual algorithm.

II. The Family of Tracking Algorithms

We shall consider the general adaptation algorithm

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu L_k(y_k - \varphi_k^\tau\hat{\theta}_k), \qquad \mu \in (0,1) \tag{11}$$

where the gain L_k is chosen in some different ways.

Case 1—Least Mean Squares (LMS):

$$L_k = \varphi_k. \tag{12}$$

This is a standard algorithm [21], [22], and has been used in numerous adaptive signal processing applications.

Case 2—Recursive Least Squares (RLS):

$$L_k = P_k\varphi_k \tag{13}$$

$$P_k = \frac{1}{1-\mu}\left\{P_{k-1} - \mu\frac{P_{k-1}\varphi_k\varphi_k^\tau P_{k-1}}{1-\mu+\mu\varphi_k^\tau P_{k-1}\varphi_k}\right\} \tag{14}$$

$$P_0 > 0. \tag{15}$$

This gives an estimate $\hat{\theta}_k$ that minimizes

$$\sum_{t=1}^{k}(1-\mu)^{k-t}(y_t - \varphi_t^\tau\theta)^2$$

where $(1-\mu)$ is the "forgetting factor."

Case 3—Kalman Filter (KF) Based Algorithm:

$$L_k = \frac{P_{k-1}\varphi_k}{R + \mu\varphi_k^\tau P_{k-1}\varphi_k} \tag{16}$$

$$P_k = P_{k-1} - \frac{\mu P_{k-1}\varphi_k\varphi_k^\tau P_{k-1}}{R + \mu\varphi_k^\tau P_{k-1}\varphi_k} + \mu Q, \tag{17}$$

$$(R > 0, Q > 0). \tag{18}$$

Here R is a positive number, and Q is a positive definite matrix. The choice of L_k corresponds to a Kalman filter state estimation for (1) and (2) and is optimal in the *a posteriori* mean square sense if v_k and w_k are Gaussian white noises with covariance matrices R and Q, respectively, and if μ is chosen as γ in (2).

If $\{\varphi_k, y_k, \theta_k\}$ obey (1) and (2) and $\hat{\theta}_k$ is found using (11), we can write the estimation error $\tilde{\theta}_k$ as

$$\tilde{\theta}_{k+1} = (I - \mu F_k)\tilde{\theta}_k - \mu L_k v_k + \gamma w_{k+1},$$
$$F_k = L_k\varphi_k^\tau. \tag{19}$$

This is a purely algebraic consequence of (1), (2), and (11) and holds for whatever sequences v_k and w_k.

If we introduce stochastic assumptions about $\{v_k\}$ and $\{w_k\}$, we can use (19) to express the covariance matrix $E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^\tau]$. That will be quite complex, however, primarily due to the dependence between $\{L_k, \varphi_k, \tilde{\theta}_k\}$. The basic approximating expression will instead be based on the following expression

$$\Pi_{k+1} = (I - \mu G_k)\Pi_k(I - \mu G_k)^\tau + \mu^2 R_v(k) M_k + \gamma^2 Q_w(k+1) \tag{20}$$

where $G_k = EF_k, M_k = EL_k L_k^\tau, R_v(k) = Ev_k^2$ and $Q_w(k) = Ew_k w_k^\tau$. As follows from Example 1.1, this would be the correct expression for the covariance matrix of $\tilde{\theta}_{k+1}$, if v_k and w_k were white noises and $L_k\varphi_k^\tau$ was independent of $\tilde{\theta}_k$, and if a term of size $\mu^2\Pi_k$ was neglected.

Indeed, we shall prove that (20) provides a good approximation of the true covariance matrix in the sense that (10) holds. Note that Π_k obeys a simple linear difference equation and can easily be calculated and examined.

III. The Main Result

A. The Assumptions

We shall now consider algorithm (11) with either of the three choices of the gain L_k, discussed in the previous section. For the analysis we shall impose some conditions on the involved variables. These are of the following character:

C1) The regressors $\{\varphi_k\}$ span the regressor space (to ensure that the whole parameter vector θ can be estimated).

C2) The dependence between the regressors φ_k and $(\varphi_i, v_{i-1}, w_i)$ decays to zero as the time distance $(k-i)$ tends to infinity.

C3) The measured error v_k and the parameter drift w_k are of white noise character.

In more exact terms, the three assumptions take the following form.

P1: Let $S_t = E[\varphi_t\varphi_t^\tau]$, assume that there exist constants $h>0$ and $\delta>0$ such that

$$\sum_{t=k+1}^{k+h} S_t \geq \delta I, \qquad \forall k.$$

P2: Let $\mathcal{G}_k = \sigma\{\varphi_k\}, \mathcal{F}_k = \sigma\{\varphi_i, v_{i-1}, w_i, i \leq k\}$. Assume that $\{\varphi_k\}$ is weakly dependent (ϕ-mixing) in the sense that there is a function $\phi(m)$ with $\phi(m) \to 0$, as $m \to \infty$, such that

$$\sup_{A\in\mathcal{G}_{k+m}, B\in\mathcal{F}_k} |P(A|B) - P(A)| \leq \phi(m) \qquad \forall k, \forall m. \tag{21}$$

Also, assume that there is a constant $c_\varphi > 0$ such that $\|\varphi_k\| \leq c_\varphi$ a.s., $\forall k$.

P3: Let $\mathcal{F}_k$ be the σ-algebra defined in P2, and assume that

$$E[v_k|\mathcal{F}_k] = 0, \quad E[w_{k+1}|\mathcal{F}_k] = E[w_{k+1}v_k|\mathcal{F}_k] = 0$$
$$E[v_k^2|\mathcal{F}_k] = R_v(k), \quad E[w_k w_k^\tau] = Q_w(k)$$

$$\sup_k\{E[|v_k|^r|\mathcal{F}_k] + E\|w_k\|^r\} \leq M, \quad \text{for some} \quad r>2, M>0.$$

B. The Result

Now, let Π_k be defined by the following linear, deterministic difference equation

$$\Pi_{k+1} = (I - \mu R_k S_k)\Pi_k(I - \mu R_k S_k)^\tau + \mu^2 R_v(k) R_k S_k R_k + \gamma^2 Q_w(k+1) \tag{22}$$

where $S_k = E[\varphi_k\varphi_k^\tau]$, and R_k is defined as follows.

LMS Case:

$$R_k = I. \tag{23}$$

RLS Case:

$$R_k = R_{k-1} - \mu R_{k-1}S_k R_{k-1} + \mu R_{k-1}, \qquad (R_0 = P_0). \tag{24}$$

KF Case:

$$R_k = R_{k-1} - \mu R_{k-1}S_k R_{k-1} + \mu Q/R, \qquad (R_0 = P_0/R). \tag{25}$$

We then have the following main result.

Theorem 3.1: Consider any of the three basic algorithms in Section II. Assume that P1, P2, and P3 hold. Let Π_k be defined as above. Then $\forall\mu \in (0, \mu^*), \forall k \geq 1$

$$\|E[\tilde{\theta}_k\tilde{\theta}_k^\tau] - \Pi_k\| \leq c\sigma(\mu)\left[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^k\right] \tag{26}$$

where $\sigma(\mu) \to 0$ (as $\mu \to 0$), which is defined by

$$\sigma(\mu) \triangleq \min_{m\geq 1}\{\sqrt{\mu}m + \phi(m)\} \tag{27}$$

and $\phi(m)$ was defined in P2, and $\alpha \in (0,1), \mu^* \in (0,1), c>0$ are constants which may be computed using properties of $\{\varphi_k, v_k, w_k\}$.

The proof is given in Section V. Let us now discuss the conditions used in the above theorem.

C. The Degree of Approximation

First of all, it is clear that the quantity $\sigma(\mu)$ plays an important role. The faster it tends to zero, the better approximation is obtained. The rate by which it tends to zero is according to (27) a reflection of how fast $\phi(m)$ (that is, the dependence among the regressors) tends to zero as m increases. For example, if the regressors are m-dependent, so that φ_k and φ_ℓ are independent for $|k-\ell|>m$, then $\phi(n) = 0$ for $n>m$ and $\sigma(\mu)$ will behave like $\sqrt{\mu}$. Also, if the dependence is exponentially decaying ($\phi(m) \approx Ce^{-\alpha m}$), then we can find that

$$\sigma(\mu) < C\mu^{0.5-\delta}$$

for arbitrarily small, positive δ. This gives a good picture of typical decay rates of σ.

D. Persistence of Excitation: Condition P1

Condition P1 is quite natural and weak, just requiring the regressor covariance matrix to add up to full rank over a given time span of arbitrary length. It has been known to be a necessary condition (in a certain sense) for boundedness of $E\|\tilde{\theta}_k\|^2$ generated by LMS (cf., [8]); it is also known to be the minimum excitation condition needed for the stability analysis of RLS (cf., [10]).

E. Boundedness and ϕ-Mixing of the Regressors: Condition P2

Condition P2 requires boundedness and ϕ-mixing of the regressors. Although such conditions are standard ones in the literature (e.g., [11]), they can still be considered as restrictive. As seen in several of the results in Section V, both ϕ-mixing and boundedness can be weakened considerably when we deal with specific algorithms.

It may also be remarked that when $\{\varphi_k\}$ is unbounded, we can modify the algorithm and make Theorem 3.1 hold true: Introduce the normalized signal

$$(\overline{y}_k, \overline{\varphi}_k, \overline{v}_k) = \frac{1}{\sqrt{1+\|\varphi_k\|^2}}(y_k, \varphi_k, v_k).$$

Then we have from (1)

$$\overline{y}_k = \theta_k^\tau \overline{\varphi}_k + \overline{v}_k.$$

Thus, $\{\theta_k\}$ may be estimated based on this normalized linear regression. In this case, Theorem 3.1 can be applied to this case if only S_k and $R_v(k)$ in (22)–(25) are replaced by $E[\varphi_k\varphi_k^\tau/1+\|\varphi_k\|^2]$ and $E[1/1+\|\varphi_k\|^2]R_v(k)$, respectively.

F. The Parameter Drift Model: Condition P3

There are two things to mention around Condition P3. First, we note that the martingale difference property of w_k essentially means that the true parameters, according to model (2), are assumed to be a random walk. Although this model is quite standard, it has also been criticized as being too restrictive. We believe that a random walk model, in the context of slow adaptation (small μ), captures the tracking behavior of the algorithm very well. This is, in a sense, a worst case analysis, since the future behavior of the model is unpredictable.

We may also note that time-varying covariances $Q_w(k)$ and $R_v(k)$ are allowed. Several of the special model drift cases described in [12] are therefore covered by P3. Other drift models, where the driving noise is colored, can be put into a similar Kalman filter framework. To cover also that case with our techniques, however, requires more work.

Condition P3 also introduces assumptions about higher moments than P2. We remark that if we only assume that $\{v_k\}$ and $\{w_k\}$ are bounded in e.g., mean square sense, then upper bounds for the mean square tracking errors can be established (cf., [7] and [8]). The strengthened assumption in P3 allows us to obtain performance values much more accurate than upper bounds.

G. The Practical Use of the Theorem

The practical consequences of Theorem 3.1 is that a very simple algorithm, the linear, deterministic difference equation (22), will describe the tracking behavior. This equation is quite easy to analyze. In fact, there is an extensive literature on such analysis, in particular for the special case of LMS. Among many references, we may refer to [12] for a survey of such results. In essence, all these results capture the dilemma between tracking error (Π is large because μ is small) and the noise sensitivity (Π is large because μ is large) and may point to the best compromises between these requirements.

For example, under weak stationarity of the regressors

$$S_k \equiv S$$

we find that R_k will converge to $\tilde{R}$ as $k \to \infty$, where $\tilde{R} = I$ in the LMS case, $\tilde{R} = S^{-1}$ in the RLS case and for the KF case we have to solve

$$\tilde{R}S\tilde{R} = Q/R$$

for $\tilde{R}$. Inserted into (22) this gives the following stationary values Π for the tracking error covariance matrix (neglecting the term $\mu^2\Pi$)

$$\begin{aligned} LMS:&\quad S\Pi + \Pi S = \mu R_v S + \frac{\gamma^2}{\mu}Q_w \\ RLS:&\quad \Pi = \frac{1}{2}\left[\mu R_v S^{-1} + \frac{\gamma^2}{\mu}Q_w\right] \\ KF:&\quad \tilde{R}S\Pi + \Pi(\tilde{R}S)^\tau = \mu R_v Q/R + \frac{\gamma^2}{\mu}Q_w. \end{aligned}$$

Note that if we have $Q = Q_w$ and $R = R_v$, then the latter equation can be solved as

$$\Pi = \frac{R}{2}\left(\mu + \frac{\gamma^2}{\mu}\right)\tilde{R}.$$

From these expressions the trade-offs between tracking ability and noise sensitivity are clearly visible.

IV. A General Theorem

In this section, we shall present a general theorem on performance of tracking algorithm (11) when the gain L_k is not specified, from which our main result Theorem 3.1 will follow. The general theorem has weaker, but less explicit, assumptions. From now on the treatment and discussion will be more technical. The main line of thought in the proofs, however, follows the outline given after Example 1.1 in Section I.

A. Notations

The following notations will be used in the remainder of the paper. These notations are the same as in the companion paper [9].

a) The minimum and maximum eigenvalues of a matrix X are denoted by $\lambda_{\min}(X)$ and $\lambda_{\max}(X)$, respectively, and

$$\|X\| \triangleq \{\lambda_{\max}(XX^\tau)\}^{1/2}$$
$$\|X\|_p \triangleq \{E(\|X\|^p)\}^{1/p}, \qquad p \geq 1.$$

b) Let $x = \{x_k(\mu), k \geq 1\}$ be a random sequence parameterized by $\mu \in (0,1)$. Denote

$$\mathcal{L}_p(\mu^*) = \{x: \sup_{\mu \in (0,\mu^*]} \sup_{k \geq 1} \|x_k(\mu)\|_p < \infty\}. \quad (28)$$

c) Let $F = \{F_k(\mu)\}$ be any (square) matrix random process parameterized by $\mu \in (0,1)$. For any $p \geq 1, \mu^* \in (0,1)$, define

$$\mathcal{S}_p(\mu^*) = \left\{ F: \left\| \prod_{j=i+1}^{k} (I - \mu F_k(\mu)) \right\|_p \leq M(1-\mu\alpha)^{k-i}, \forall \mu \in (0,\mu^*], \forall k \geq i \geq 0, \text{ for some } M > 0 \text{ and } \alpha \in (0,1) \right\}$$

similarly

$$\mathcal{S}(\mu^*) = \left\{ F: \left\| \prod_{j=i+1}^{k} (I - \mu E[F_k(\mu)]) \right\| \leq M(1-\mu\alpha)^{k-i}, \forall \mu \in (0,\mu^*], \forall k \geq i \geq 0, \text{ for some } M > 0, \text{and } \alpha \in (0,1) \right\}.$$

In what follows, it will be convenient to introduce the sets

$$\mathcal{S}_p \triangleq \bigcup_{\mu^* \in (0,1)} \mathcal{S}_p(\mu^*), \quad \mathcal{S} \triangleq \bigcup_{\mu^* \in (0,1)} \mathcal{S}(\mu^*). \quad (29)$$

We may call these stability sets. They are related to the stability of random equation (19) and deterministic equation (20), respectively. For simplicity, we shall sometimes suppress the parameter (μ) in $F_k(\mu)$, when there is no risk of confusion.

d) For scalar random sequences $a = (a_k, k \geq 0)$, we set

$$\mathcal{S}^0(\lambda) = \left\{ a: a_k \in [0,1], E \prod_{j=i+1}^{n} (1 - a_j) \leq M\lambda^{k-i}, \forall k \geq i \geq 0, \text{for some } M > 0 \right\}.$$

Also

$$\mathcal{S}^0 \triangleq \bigcup_{\lambda \in (0,1)} \mathcal{S}^0(\lambda). \quad (30)$$

e) Let $p \geq 1$ and let $x \triangleq \{x_i\}$ be any random process. Set

$$\mathcal{M}_p = \left\{ x: \left\| \sum_{i=m+1}^{m+n} x_i \right\|_p \leq C_p n^{1/2}, \forall n \geq 1, m \geq 0, \text{ for some } C_p \text{ depending only on } p \text{ and } x \right\}.$$

As is known, for example from [10], martingale difference sequence, ϕ- and α-mixing sequences, and linear processes (a process generated from a white noise source via a linear filter with absolutely summable impulse response) are all in the set $\mathcal{M}_p$.

In particular, when $\{x_i\}$ is a martingale difference sequence, by the Burkholder inequality we have $(p > 1)$

$$\left\| \sum_{i=m+1}^{m+n} x_i \right\|_p \leq (B_p x_p^*) n^{1/2}, \qquad \forall n \geq 1, m \geq 0 \quad (31)$$

where $x_p^* \triangleq \sup_k \|x_k\|_p$ and B_p is a constant depending on p only (cf., [11]). (This fact will be frequently used in the sequel without explanations.)

f) Let $\{A_k\}$ be a matrix sequence, $b_k \geq 0, \forall k \geq 0$. Then by $A_k = O(b_k)$ we mean that there exists a constant $M < \infty$ such that

$$\|A_k\| \leq M b_k, \qquad \forall k \geq 0.$$

The constant M may be called the ordo-constant. Throughout the sequel, the ordo-constant does not depend on μ, even if $\{A_k\}$ or $\{b_k\}$ does.

B. Assumptions

We will first show, given the exponential stability of the homogenous part of (19) and a certain weak dependence property of the adaptation gains, how the tracking performance can be analyzed, and then we present more detailed discussions on such properties.

In the sequel, unless otherwise stated, $\mathcal{F}_k$ denotes the σ-algebra generated by $\{\varphi_i, w_i, v_{i-1}, i \leq k\}$, and $\{F_k\}$ is defined in (19).

To establish the general theorem, we need the following assumptions:

A1) (Exponential stability) There are $\mu^* \in (0,1)$, and $p \geq 2$ such that

$$\{F_k\} \in \mathcal{S}_p(\mu^*) \bigcap \mathcal{S}(\mu^*).$$

A2) (Weak dependence) There is a real number $q \geq 3$ together with a bounded function $\phi(m,\mu) \geq 0$, with

$$\lim_{\substack{m \to \infty \\ \mu \to 0}} \phi(m,\mu) = 0$$

(taking first m to infinity and then μ to zero) such that $\forall m, \forall k, \forall \mu \in (0,\mu^*]$

$$\|E[F_k|\mathcal{F}_{k-m}] - E[F_k]\|_q \leq \phi(m,\mu).$$

A3) $L_i \in \mathcal{F}_i, \forall i \geq 1$, and there is $\mu^* \in (0,1)$ such that

$$\{L_i\} \in \mathcal{L}_r(\mu^*), \quad \{F_i\} \in \mathcal{L}_{2q}(\mu^*)$$

with $r = (1/2 - 1/p - 3/2q)^{-1}$, and with p and q defined as in A1) and A2).

A4) For all $k \geq 1$ we have

$$E[v_k|\mathcal{F}_k] = 0, \quad E[w_{k+1}|\mathcal{F}_k] = E[w_{k+1}v_k|\mathcal{F}_k] = 0$$
$$E[v_k^2|\mathcal{F}_k] = R_v(k), \quad E[w_{k+1}w_{k+1}^\tau] = Q_w(k+1)$$

$$E[|v_k|^r|\mathcal{F}_k] + E[\|w_{k+1}\|^r] \leq M < \subset, \qquad \forall k \geq 1$$

for deterministic quantities $R_v(k), Q_w(k+1)$ and M, where r is defined as in A3).

The key conditions are A1) and A2). In general, A1) can be guaranteed by a certain type of stochastic persistence of excitation condition, which is studied in the companion paper [9], while A2) can be guaranteed by imposing a certain weak dependence condition on the regressor $\{\varphi_i\}$. More detailed discussions will be given later. At the moment, we just remark that if A1) and A2) hold for all $p \geq 1$ and all $q \geq 1$, then in A3) and A4), the number r needs only to satisfy $r > 2$.

C. The General Theorem

Now, recursively define a matrix sequence $\{\hat{\Pi}_k\}$ as follows

$$\begin{aligned}\hat{\Pi}_{k+1} &= (I - \mu E[F_k])\hat{\Pi}_k(I - \mu E[F_k])^\tau \\ &\quad + \mu^2 R_v(k) E[L_k L_k^\tau] + \gamma^2 Q_w(k+1) \qquad (32)\end{aligned}$$

where $\hat{\Pi}_0 = E[\tilde{\theta}_0 \tilde{\theta}_0^\tau]$, and $R_v(k)$ and $Q_w(k+1)$ are defined in Assumption A4). Note that this definition is very close to the definition of Π_k in (22). We now have a result that is the "mother-theorem" of Theorem 3.1.

Theorem 4.1: Let Assumptions A1)–A4) hold. Let the tracking error $\tilde{\theta}_k$ be defined by (11) [or (19)], and let $\hat{\Pi}_k$ be defined by (32). Then $\forall \mu \in (0, \mu^*], \forall k \geq 1$

$$\|E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^\tau] - \hat{\Pi}_{k+1}\| \leq c\sigma(\mu)\left[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^k\right]$$

where $c > 0$ and $\alpha \in (0,1)$ are constants and $\sigma(\mu)$ is a function that tends to zero as μ tends to zero. It is defined by

$$\sigma(\mu) \triangleq \min_{m \geq 1}\{\sqrt{\mu} m + \phi(m, \mu)\}.$$

The proof is given in Appendix A.

Next, we show that under more conditions, the expression for $\hat{\Pi}_k$ in (11) can be further simplified.

Corollary 4.1: Under the conditions of Theorem 4.1, if $F_k = P_k \varphi_k \varphi_k^\tau$ with $\|\varphi_k\|_{2t} = O(1), \|F_k\|_t = O(1)$, for some $t > 1$, and if there are some function $\delta(\mu)$, tending to zero as μ tends to zero, and some deterministic sequence $\{R_k\}$ such that

$$\|P_k - R_k\|_s = O(\delta(\mu)) \quad \forall k, \forall \mu \in (0, \mu^*], \quad s = (1 - t^{-1})^{-1}$$

then we have $(\forall \mu \in (0, \mu^*], \forall k \geq 1)$

$$\begin{aligned}&\|E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^\tau] - \Pi_{k+1}\| \\ &\quad \leq c[\sigma(\mu) + \delta(\mu)]\left[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^k\right] \qquad (33)\end{aligned}$$

for some constants $c > 0$ and $\alpha \in (0,1)$, where Π_k is recursively defined by

$$\begin{aligned}\Pi_{k+1} &= (I - \mu R_k S_k)\Pi_k(I - \mu R_k S_k)^\tau \\ &\quad + \mu^2 R_v(k) R_k S_k R_k^\tau + \gamma^2 Q_w(k+1) \qquad (34)\end{aligned}$$

with $S_k = E[\varphi_k \varphi_k^\tau]$ and $\Pi_0 = \hat{\Pi}_0$.

Proof: By Theorem 4.1, we need only to show that

$$\|\hat{\Pi}_{k+1} - \Pi_{k+1}\| = O\left(\delta(\mu)\left[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^k\right]\right).$$

This can be derived by straightforward calculations based on the equations for $\hat{\Pi}_k$ and Π_k, and hence Corollary 4.1 is true.

Remark: If in Condition A2)

$$\phi(m, \mu) = O(\phi(m) + \delta(\mu)), \quad \delta(\mu) = \min_{m \geq 1}[\sqrt{\mu} m + \phi(m)]$$

then $\sigma(\mu)$ defined in Theorem 4.1 satisfies $\sigma(\mu) = O(\delta(\mu))$. This will be the case for RLS and KF algorithms in Theorem 3.1, as can be seen from Section V.

The following result also follows directly from Theorem 4.1.

Corollary 4.2: If, in addition to the conditions of Theorem 4.1, $R_v(k) \equiv R_v, Q_w(k) \equiv Q_w$, and there are F, G, and a function $\delta(\mu)$, tending to zero as μ tends to zero, such that $\forall \mu \in (0, \mu^*]$

$$\|EF_k - F\| + \|E(L_k L_k^\tau) - G\| \leq \delta(\mu), \qquad \forall k$$

then for some $\alpha \in (0,1)$ and for all $\mu \in (0, \mu^*], k \geq 1$

$$\begin{aligned}E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^\tau] &= \Pi + O\left([\sigma(\mu) + \delta(\mu)]\left[\mu + \frac{\gamma^2}{\mu}\right]\right) \\ &\quad + O((1-\alpha\mu)^k) \qquad (35)\end{aligned}$$

where Π satisfies the following Lyapunov equation

$$F\Pi + \Pi F^\tau = \mu R_v G + \frac{\gamma^2}{\mu} Q_w. \qquad (36)$$

Now denote

$$\overline{R}_v = R_v \int_0^\infty e^{-Ft} G e^{-F^\tau t}, \quad \overline{Q}_w = \int_0^\infty e^{-Ft} Q_w e^{-F^\tau t}$$

the solution to Lyapunov equation (36) can be expressed as

$$\Pi = \mu \overline{R}_v + \frac{\gamma^2}{\mu}\overline{Q}_w$$

in which there is a reminiscence of the results obtained in the simple example discussed in Section I [see (9)].

D. Discussion on the Assumptions

Now, let us discuss the key assumptions A1) and A2).

First, Assumption A1) has been studied in the companion paper [9], and here we only give some results concerning $\{F_k\} \in \mathcal{S}$, which will be used shortly in the next section.

Proposition 4.1: Let $\{G_k\}$ be a random matrix process, possibly dependent on μ, with the property

$$E\|G_k\| \leq \delta(\mu), \quad \text{for all small } \mu \text{ and all } k \qquad (37)$$

where $\delta(\mu) \to 0$ as $\mu \to 0$.

Then $\{F_k\} \in \mathcal{S} \Leftrightarrow \{F_k + G_k\} \in \mathcal{S}$.

Proof: (Sufficiency) Recursively define $(\forall x: \|x\| = 1)$

$$x_{k+1} = (I - \mu E[F_k + G_k])x_k, \quad \forall k \geq m, x_m = x.$$

Then

$$\begin{aligned}x_{k+1} &= (I - \mu E(F_k))x_k - \mu E(G_k)x_k \\ &= \prod_{i=m}^{n}[I - \mu E(F_i)]x_m \\ &\quad - \sum_{i=m}^{n} \mu \prod_{j=i+1}^{n}[I - \mu E(F_j)]E(G_i)x_i.\end{aligned}$$

Consequently, similar to the proof of Theorem 3.1 in [9], by the Gronwall inequality we have

$$\|x_{n+1}\| \le 2M(1-\mu\alpha)^{n-m+1} \times \left\{1+\sum_{i=m}^{n}\prod_{j=i+1}^{n}(1+\mu E\|G_j\|)\cdot \mu E\|G_i\|\right\}.$$

From this and condition (37), it is not difficult to convince oneself that $\{F_k+G_k\} \in \mathcal{S}$.

(Necessity) By using the fact proved above and noting that $F_k=(F_k+G_k)-G_k$, we know that $\{F_k\}\in\mathcal{S}$. This completes the proof. #

The following useful result follows from Proposition 4.1 immediately.

Proposition 4.2: Let $F_k = P_k H_k$ and the following conditions be satisfied:

i) $\{H_k\} \in \mathcal{L}_t(\mu^*), \mu^* \in (0,1), t \ge 1$.
ii) $\|P_k - \overline{P}_k\|_s \le \delta(\mu), \forall \mu \in (0,\mu^*]$, where $\delta(\mu) \to 0$ as $\mu \to 0, s=(1-t^{-1})^{-1}$, and $\{\overline{P}_k\}$ is a deterministic process.

Then $\{F_k\} \in \mathcal{S} \Leftrightarrow \{\overline{P}_k H_k\} \in \mathcal{S}$.

Proof: The result follows directly from Proposition 4.1, if we note that

$$F_k = \overline{P}_k H_k + (P_k - \overline{P}_k)H_k.$$

#

We now turn to discuss the weak dependence condition A2).

Example 4.1: Let $\{\varphi_i\}$ satisfy (21), and $L(\cdot)$: $R^d \to R^{d\times d}$ be a real matrix function with $\|L(\varphi(k))\|_q = O(1)$, for some $1 \le q \le \infty$. Then we have the following inequality (cf., [19])

$$\|E[L(\varphi_k)|\mathcal{F}_{k-m}] - EL(\varphi_k)\|_q = O([\phi(m)]^{1-(1/q)}) \qquad \forall k, m. \tag{38}$$

Hence, if $F_k = L(\varphi_k)$, then Condition A2) holds.

Note that when $\{\varphi_k\}$ satisfies Condition P2 in Section III, we have by taking $q=\infty$ in (38)

$$\|E[\varphi_k\varphi_k^\tau|\mathcal{F}_{k-m}] - E\varphi_k\varphi_k^\tau\|_\infty = O(\phi(m)). \tag{39}$$

This fact will be used in the next section in the proof of Theorem 3.1. #

Example 4.2: Let $\{\varphi_k\}$ be generated by

$$x_k = Ax_{k-1} + B\xi_k, \qquad \text{(A stable)}$$
$$\varphi_k = Cx_k + \xi_k$$

where $\{\xi_j, j\ge k+1\}$ and $\{v_{j-1}, w_j, j \le k\}$ are independent, and $\{\xi_j\}$ is an independent sequence. Assume that

$$\sup_k E\|\xi_k\|^{(b+1)q} < \infty, \qquad \text{for some} \quad b\ge 0, q \ge 1.$$

Then for any function $L(\cdot)$: $R^d \to R^{d\times d}$, with

$$\|L(x)-L(x')\| \le M(\|x\|+\|x'\|+1)^b\|x-x'\|, \qquad \forall x, x'$$

there is a constant $\lambda \in (0,1)$ such that (cf., [14]) $\forall m \ge 0$ $\forall k \ge 0$

$$\|E[L(\varphi_{k+m})|\mathcal{F}_k] - EL(\varphi_{k+m})\|_q = O(\lambda^m).$$

Hence, if $F_k = L(\varphi_k)$, then again, Condition A2) holds. #

The following simple result will be useful in the sequel.

Proposition 4.3: Let $F_k = P_k L(\varphi_k)$, and the following two conditions hold:

i) There is a bounded deterministic matrix sequence $\{\overline{P}_k\}$ and a function $\delta(\mu)$ tending to zero as μ tends to zero, such that

$$\|P_k - \overline{P}_k\|_s \le \delta(\mu), \quad \forall \mu \in (0,\mu^*], \qquad \text{for some } s>1.$$

ii) There is a number $r>1$ such that $\|L(\varphi_k)\|_r = O(1)$, together with a function $\phi(m)$ tending to zero as m tends to infinity, such that

$$\|E[L(\varphi_{k+m})|\mathcal{F}_k] - EL(\varphi_{k+m})\|_q \le \phi(m), \quad \forall k, \forall m, \quad (q=(r^{-1}+s^{-1})^{-1}).$$

Then Condition A2) holds with $\phi(m,\mu) = O(\phi(m)+\delta(\mu))$.

Proof: The result follows directly from the following identity

$$\begin{aligned} E[F_{k+m}|\mathcal{F}_k] - EF_{k+m} &= [(P_{k+m} - \overline{P}_{k+m})L(\varphi_{k+m})|\mathcal{F}_k] \\ &\quad - E[(P_{k+m}-\overline{P}_{k+m})L(\varphi_{k+m})] \\ &\quad + \overline{P}_{k+m}\{E[L(\varphi_{k+m})|\mathcal{F}_k] - EL(\varphi_{k+m})\}. \end{aligned}$$

#

V. Analysis of the Basic Algorithms

In this section, we shall show that, for the basic LMS, RLS, and KF algorithms, Conditions A1)–A3) can be guaranteed by imposing some explicit (stochastic excitation and weak dependence) conditions on the regressors $\{\varphi_k\}$ and at the same time prove Theorem 3.1.

A. Analysis of LMS

For the LMS defined by (11) and (12), let us introduce the following two kinds of weak dependence conditions:

L1) Condition P2 is satisfied but with the boundedness condition on $\{\varphi_k\}$ relaxed to the following: There exist positive constants ε, δ, M, and K such that

$$E\exp\left\{\sum_{j=i+1}^{n}\varepsilon\|\varphi_j\|^{2+\delta}\right\} \le M\exp\{K(n-i)\} \qquad \forall n \ge i \ge 0.$$

L1′) The random process $F_k \triangleq \varphi_k\varphi_k^\tau$ has the following expansion

$$F_k = \sum_{j=0}^{\infty} A_j Z_{k-j} + D_k, \quad \sum_{j=0}^{\infty}\|A_j\| < \infty$$

where $\{Z_k\}$ is an independent process such that $\{Z_j, j \ge k+1\}$ and $\{v_{j-1}, w_j, j\le k\}$ are independent and satisfies

$$\sup_k E\exp\{\alpha\|Z_k\|^{1+\delta}\} < \infty, \text{ for some } \alpha>0, \delta>0$$

and where $\{D_k\}$ is a bounded deterministic process.

Theorem 5.1: Let Conditions P1 and P3 be satisfied. If either L1) or L1′) above holds, then Conditions A1)–A4) of Theorem 4.1 hold (for all $p \geq 1, q \geq 1$) and Theorem 3.1 is true for the LMS case.

Proof: First, in the LMS case, Conditions P1 and L1) [or L1′)] ensure that Condition A1) of Theorem 4.1 holds for all $p \geq 1$ (cf., [9, Theorem 3.3]). Next, when L1) holds, by Example 4.1 we know that Condition A2) is true for all $q \geq 1$. Also, when L1′) holds, by the assumed independency we have for all $q \geq 1$

$$\|E[F_k|\mathcal{F}_{k-m}] - EF_k\|_q = \left\|\sum_{j=m}^{\infty}[A_jZ_{k-j} - EA_jZ_{k-j}]\right\|_q = O\left(\sum_{j=m}^{\infty}\|A_j\|\right), \quad \forall m \geq 1.$$

Hence A2) holds again for all $q \geq 1$.

Moreover, Conditions A3) and A4) hold obviously in the present case. Finally, by (39), the result of Theorem 3.1 (in the LMS case) follows directly from Theorem 4.1. This completes the proof.

B. Analysis of RLS

For the RLS algorithm defined by (11), (13), and (14), let us introduce the following two kinds of excitation conditions:

R1) There exist constants $h > 0, c > 0, \delta > 0$ such that

$$P\left\{\lambda_{\min}\left(\sum_{i=k+1}^{k+h}\varphi_i\varphi_i^\tau\right) \geq c|\mathcal{F}_k\right\} > \delta, \quad \forall k.$$

R1′) There exists $h > 0$ such that

$$\sup_k E\left[\lambda_{\min}\left(\sum_{i=k+1}^{k+h}\varphi_i\varphi_i^\tau\right)\right]^{-t} < \infty, \quad \forall t \geq 1.$$

The following weak dependence condition will also be used:

R2) There exists a number $t \geq 5$, such that $\|\varphi_k\|_{4t} = O(1)$, and that

$$\|E[\varphi_k\varphi_k^\tau|\mathcal{F}_{k-m}] - E\varphi_k\varphi_k^\tau\|_{2t} \leq \phi(m), \quad \forall k, m$$

where $\phi(m) \to 0$ as $m \to \infty$.

Detailed discussions and investigations on the above first two conditions can be found in [10] and [17]. It has been shown in [10] that if Condition P1 and (21) in Section III hold, then R1) is true. Also, if $\{\varphi_k\}$ is generated by a linear state-space model as in Example 4.2, then R1′) can be verified (cf., [17]). Moreover, Condition R2) has been discussed in the last section.

Theorem 5.2: Let Condition R1 [or R1′)] and R2) above be satisfied. Then Conditions A1)–A3) hold (for any $p < 2t, q < t$), and Theorem 3.1 is true for the RLS case.

Proof: First, note that

$$\prod_{j=i+1}^{k}(I - \mu F_j) = (1-\mu)^{k-i}P_{k+1}P_{i+1}^{-1}, \quad \forall k \geq i \tag{40}$$

and

$$P_k^{-1} = (1-\mu)P_{k-1}^{-1} + \mu\varphi_k\varphi_k^\tau. \tag{41}$$

From this and Condition R2) it follows that

$$\|P_k^{-1}\|_{2t} = O(1), \quad \forall \mu \in (0,1). \tag{42}$$

Also, by Theorem 1 in [10], there is $\mu^* \in (0,1)$ such that

$$\{P_k\} \in \mathcal{L}_s(\mu^*), \quad \forall s \geq 1. \tag{43}$$

Combining (40), (42), and (43), we get

$$\{F_k\} \in \mathcal{S}_p, \quad \forall p < 2t. \tag{44}$$

Now, define ($\overline{P}_0 = P_0$)

$$\overline{P}_k^{-1} = (1-\mu)\overline{P}_{k-1}^{-1} + \mu E(\varphi_k\varphi_k^\tau). \tag{45}$$

Since either R1) or R1′) implies P1 (cf., [10]) by a similar (actually simpler) argument as that used for the proof of (43), we know that $\|\overline{P}_k\| = O(1)$. We next prove that

$$\|P_k^{-1} - \overline{P}_k^{-1}\|_{2t} = O(\delta(\mu)), \quad \delta(\mu) = \min_{m\geq 1}\{\sqrt{\mu}m + \phi(m)\}. \tag{46}$$

First, by (41) and (45)

$$P_k^{-1} - \overline{P}_k^{-1} = \mu\sum_{i=1}^{k}(1-\mu)^{k-i}[\varphi_i\varphi_i^\tau - E\varphi_i\varphi_i^\tau]. \tag{47}$$

For any fixed $m \geq 1$, by denoting

$$\delta_j(i) = E[\varphi_i\varphi_i^\tau|\mathcal{F}_{i-j}] - E[\varphi_i\varphi_i^\tau|\mathcal{F}_{i-j-1}], \quad 0 \leq j \leq m-1$$

we have

$$\varphi_i\varphi_i^\tau - E\varphi_i\varphi_i^\tau = \sum_{j=0}^{m-1}\delta_j(i) + \{E[\varphi_i\varphi_i^\tau|\mathcal{F}_{i-m}] - E[\varphi_i\varphi_i^\tau]\}. \tag{48}$$

Now, since for each j, the sequence $\{\delta_j(i), i \geq 1\}$ is a martingale difference, we can apply Lemma A.2 in Appendix A to each such $\{\delta_j(i), i \geq 1\}$ to obtain

$$\mu\left\|\sum_{i=1}^{k}(1-\mu)^{k-i}\sum_{j=0}^{m-1}\delta_j(i)\right\|_{2t} = O(\sqrt{\mu}m). \tag{49}$$

Also, by our assumption

$$\mu\left\|\sum_{i=1}^{k}(1-\mu)^{k-i}\{E[\varphi_i\varphi_i^\tau|\mathcal{F}_{i-m}] - E[\varphi_i\varphi_i^\tau]\}\right\|_{2t} \leq \phi(m). \tag{50}$$

Hence, (46) follows from (47)–(50) immediately.

Similar to the proof of (44), it is evident that

$$\{\overline{P}_k\varphi_k\varphi_k^\tau\} \in \mathcal{S}. \tag{51}$$

Now

$$\|P_k - \overline{P}_k\| \le \|P_k\| \cdot \|P_k^{-1} - \overline{P}_k^{-1}\| \cdot \|\overline{P}_k\|.$$

From this, (43), and (46) it follows that

$$\|P_k - \overline{P}_k\|_s = O(\delta(\mu)). \qquad \forall s < 2t, (\text{ for small } \mu). \tag{52}$$

Hence, by Proposition 4.2 and (51), we know that $\{F_k\} \in \mathcal{S}$. This in conjunction with (44) verifies Condition A1).

Now, by (52) and R2) from Proposition 4.3 it is evident that Condition A2) holds for any $q < t$.

To prove A3), first note that for any $q < t$, (44) implies

$$\{F_k\} \in \mathcal{L}_{2q}(\mu^*), \quad \text{for some} \quad \mu^* > 0.$$

So we need only to prove that

$$\{L_i\} \in \mathcal{L}_r(\mu^*). \quad \text{for} \quad r > \left(\frac{1}{2} - \frac{1}{2t} - \frac{3}{2t}\right)^{-1} = \frac{2t}{t-4}.$$

This is true since by (43) and $\|\varphi_k\|_{4t} = O(1)$

$$\{L_i\} = \{P_i\varphi_i\} \in \mathcal{L}_r(\mu^*), \qquad \forall r < 4t$$

and since $4t > 2t/t - 4$. Hence A3) holds.

Thus, by taking $t = \infty$ in the above argument, we see that Conditions A1) and A2) hold for all $p \ge 1$ and all $q \ge 1$. Hence Theorem 4.1 can be applied to prove Theorem 3.1 for the RLS case, while the expression for Π_k will follow from Corollary 4.1 if we can prove that

$$\|P_k - R_k\|_s = O(\delta(\mu)), \quad s = \frac{t}{t-1} \tag{53}$$

where P_k and R_k are, respectively, defined by (14) and (24). Furthermore, by (52), it is clear that (53) will be true if

$$\|R_k - \overline{P}_k\| = O(\delta(\mu))$$

holds. This can be verified, however, by using the definitions for R_k and $\overline{P}_k$ (see Appendix B). Hence the proof is complete.

C. Analysis of the KF Algorithm

Among the three basic algorithms described in Section II, the KF algorithm defined by (11), (16), and (17) is the most complicated one to analyze. Let us now introduce the following two conditions on stochastic excitation and weak dependence:

K1) There are constants $h > 0$ and $\lambda \in (0,1)$ (independent of δ) such that

$$\left\{\frac{\lambda_k}{1 + b_{kh+1}}\right\} \in \mathcal{S}^0(\lambda)$$

where $\mathcal{S}^0(\lambda)$ is defined by (30), and λ_k and b_k are defined as follows: ($\mathcal{G}_k$ is as before the sigma-algebra generated by $\{\varphi_i, i \le k\}$)

$$\lambda_k \triangleq \lambda_{\min}\left\{E\left[\frac{1}{1+h}\sum_{i=kh+1}^{(k+1)h}\frac{\varphi_i\varphi_i^\tau}{1+\|\varphi_i\|^2}\middle|\mathcal{G}_{kh}\right]\right\},$$

$$b_k = (1-\delta)b_{k-1} + \delta(\|\varphi_k\|^2 + 1). \qquad \delta \in (0,1).$$

K2) There exists a number $t \ge 7$ together with a function $\phi(m) \to 0$ (as $m \to \infty$) such that $\|\varphi_k\|_{4t} = O(1)$, and that

$$\|E[\varphi_k\varphi_k^\tau|\mathcal{F}_{k-m}] - E\varphi_k\varphi_k^\tau\|_t \le \phi(m) \qquad \forall k, m.$$

Remark 5.2: If Conditions P1 and P2 are satisfied, then both K1) and K2) above hold (cf., [10]). When P2 is replaced by, for example, the situation discussed in Example 4.2, then again, both K1) and K2) can be verified (cf., [8]).

Theorem 5.3: Let Conditions K1) and K2) above be satisfied. Then Conditions A1)–A3) of Theorem 4.1 hold (for any $p < 2t, q < 4t/7$), and Theorem 3.1 is true for the KF case.

The proof is prefaced by several lemmas. First, we need some results proved in the companion paper ([9, Theorem 3.5, Lemmas 7.1 and 7.2]), which are collected into the following lemma for convenience of reference.

Lemma 5.1: For the KF algorithm defined by (11), (16), and (17), let Condition K1) be satisfied, and $\|\varphi_k\|_{4t} = O(1)$ for some $t \ge \frac{1}{2}$. Then

i) $\{P_k\} \in \mathcal{L}_s(\mu^*)$. $\quad \forall \mu^* \in (0,1), \forall s \ge 1$

ii) $\{P_k^{-1}\} \in \mathcal{L}_{2t}(\mu^*)$, $\quad \forall \mu^* \in (0,1)$

iii) $\{F_k\} \in \mathcal{S}_p$, $\quad \forall p < 2t$.

To apply Proposition 4.3, our main objective is then to show that $\{P_k\}$ can be "approximated" by a deterministic process $\{\overline{P}_k\}$ defined by

$$\overline{P}_k = [\overline{P}_{k-1}^{-1} + \mu R^{-1}E(\varphi_k\varphi_k^\tau)]^{-1} + \mu Q,$$
$$\overline{P}_0 = P_0. \tag{54}$$

First, by a similar (actually simplified) argument as that for Lemma 5.1, the following lemma can be established (details are not repeated).

Lemma 5.2: Assume that $\|\varphi_k\|_2 = O(1)$ and that there is an integer $h > 0$ such that

$$\inf_k \lambda_{\min}\left\{\sum_{i=k+1}^{k+h} E(\varphi_i\varphi_i^\tau)\right\} > 0. \tag{55}$$

Then the following three properties hold

i) $\|\overline{P}_k\| = O(1)$, $\quad \forall \mu \in (0,1)$.

ii) $\|\overline{P}_k^{-1}\| = O(1)$. $\quad \forall \mu \in (0,1)$.

iii) $\{(\overline{P}_k - \mu Q)R^{-1}\varphi_k\varphi_k^\tau\} \in \mathcal{S}(\mu^*)$, $\quad \forall \mu^* \in (0,1)$.

The last assertion iii) above corresponds to that in Lemma 5.1, because by (16), (17), $F_k \triangleq L_k\varphi_k^\tau$ can be rewritten as

$$F_k = (P_k - \mu Q)R^{-1}\varphi_k\varphi_k^\tau. \tag{56}$$

To be able to estimate the "distance" between P_k and $\overline{P}_k$, we need some auxiliary results.

Define $\forall k \ge i$

$$\Phi(k+1, i) = (I - \mu F_k)\Phi(k, i), \quad \Phi(i,i) = I, \tag{57}$$
$$\Psi(k+1, i) = (I - \mu\overline{F}_k)\Psi(k, i), \quad \Psi(i,i) = I \tag{58}$$

where $\overline{F}_k \triangleq (\overline{P}_k - \mu Q)R^{-1}E(\varphi_k\varphi_k^\tau)$.

Lemma 5.3: Let $\{e_k\} \in \mathcal{M}_r, r \geq 1$

$$\{\overline{F}_i\} \in \mathcal{S}(\mu^*),$$
$$\sup_{k \geq i \geq 0} \|\Phi(k,i)\|_q = O(1), \qquad \forall \mu \in (0, \mu^*],$$
$$\{F_i\} \in \mathcal{L}_t(\mu^*).$$

Then for $s \triangleq (r^{-1} + q^{-1} + t^{-1})^{-1}$

$$\left\| \sum_{i=0}^{k} \Phi(k+1, i+1) e_i \Psi^\tau(k+1, i+1) \right\|_s = O(\mu^{-(1/2)}, \qquad \forall \mu \in (0, \mu^*].$$

The proof is essentially the same as that for Lemma A.2 in Appendix A, and hence details are not repeated.

Lemma 5.4: Let Condition K1) hold, and $\|\varphi_k\|_{4t} = O(1)$, $t \geq 1$. If $\{e_k\} \in \mathcal{M}_r$, then for $s < (r^{-1} + 3/4t)^{-1}$

$$\left\| \sum_{i=0}^{k} \Phi(k+1, i+1) e_i \Psi^\tau(k+1, i+1) \right\|_s = O(\mu^{-(1/2)}, \quad \text{for small } \mu.$$

Proof: We need only to verify the conditions of Lemma 5.3. First of all, K1) implies $\{\lambda_k\} \in \mathcal{S}^0(\lambda), \lambda \in (0,1)$. So by Theorem 2.2 in [8]

$$\inf_k \lambda_{\min}\left\{ \sum_{i=kh_0+1}^{(k+1)h_0} E\lambda_i \right\} > 0, \qquad \text{for some} \quad h_0 > 0$$

which implies that there exists $h_1 > 0$, such that

$$\inf_k \lambda_{\min}\left\{ \sum_{i=kh_1+1}^{(k+1)h_1} E\left[\frac{\varphi_i \varphi_i^\tau}{1 + \|\varphi_i\|^2} \right] \right\} > 0$$

hence, we have

$$\inf_k \lambda_{\min}\left\{ \sum_{i=kh_1+1}^{(k+1)h_1} E[\varphi_i \varphi_i^\tau] \right\} > 0.$$

Therefore, (55) is true and Lemma 5.2 is applicable.

Now, by Theorem 2.4 in [8]

$$\|\Phi(k,i)\| \leq \{\|P_k\| \cdot \|P_i^{-1}\|\}^{1/2}.$$

So, by Lemma 5.1

$$\sup_{k \geq i \geq 0} \|\Phi(k,i)\|_q = O(1), \qquad \forall q < 4t. \tag{59}$$

Moreover, by (56) and Lemma 5.1–i)

$$\{F_i\} \in \mathcal{L}_p(\mu^*), \qquad \forall p < 2, t. \tag{60}$$

Now, by (59), (60), and Lemma 5.2–iii), it follows from Lemma 5.3 that the desired result is true. #

Lemma 5.5: Let $\{P_k\}$ and $\{\overline{P}_k\}$ be, respectively, defined by (17) and (54). Then

$$\begin{aligned} P_k - \overline{P}_k = &(I - \mu F_k)(P_{k-1} - \overline{P}_{k-1})(I - \mu \overline{F}_k)^\tau \\ &+ \mu R^{-1}(P_k - \mu Q)[E(\varphi_k \varphi_k^\tau) \\ &- \varphi_k \varphi_k^\tau](\overline{P}_k - \mu Q) \end{aligned}$$

where F_k is defined by (56) and $\overline{F}_k$ is defined as in (58).

Proof: Denote

$$Q_k \triangleq [P_{k-1}^{-1} + \mu R^{-1} \varphi_k \varphi_k^\tau]$$
$$\overline{Q}_k \triangleq [\overline{P}_{k-1}^{-1} + \mu R^{-1} E(\varphi_k \varphi_k^\tau)].$$

Then by (17) and (54)

$$\begin{aligned} P_k - \overline{P}_k &= Q_k^{-1} - \overline{Q}_k^{-1} = Q_k^{-1}[\overline{Q}_k - Q_k]\overline{Q}_k^{-1} \\ &= Q_k^{-1}\{\overline{P}_{k-1}^{-1} - P_{k-1}^{-1} + \mu R^{-1}[E(\varphi_k \varphi_k^\tau) \\ &\quad - \varphi_k \varphi_k^\tau]\}\overline{Q}_k^{-1} \\ &= Q_k^{-1} P_{k-1}^{-1}(P_{k-1} - \overline{P}_{k-1})\overline{P}_{k-1}^{-1}\overline{Q}_k^{-1} \\ &\quad + \mu R^{-1} Q_k^{-1}[E(\varphi_k \varphi_k^\tau) - \varphi_k \varphi_k^\tau]\overline{Q}_k^{-1}. \end{aligned} \tag{61}$$

But, it is not difficult to verify that (via the matrix inverse formula)

$$Q_k^{-1} P_{k-1}^{-1} = I - \mu F_k$$

and

$$\overline{P}_{k-1}^{-1} \overline{Q}_k^{-1} = I - \mu \overline{F}_k^\tau.$$

Hence, the desired result follows from (61) directly.

Lemma 5.6: Let Conditions K1) and K2) be satisfied. Then

$$\|P_k - \overline{P}_k\|_s = O(\delta(\mu)), \qquad \forall s < \frac{4t}{5}$$

holds for all small μ, where $\delta(\mu) = \min_{m \geq 1}\{\sqrt{\mu} m + \phi(m)\}$.

Proof: Set

$$H_k = R^{-1}(P_k - \mu Q)$$
$$X_k = (\varphi_k \varphi_k^\tau)(\overline{P}_k - \mu Q).$$

Then by (57), (58) and Lemma 5.5 we have $(\forall k \geq 1, \forall m \geq 1)$

$$\begin{aligned} P_k - \overline{P}_k &= \mu \sum_{i=1}^{k} \Phi(k+1, i+1) H_i [EX_i - X_i] \\ &\quad \cdot \Psi^\tau(k+1, i+1) \\ &= \mu \sum_{i=1}^{k} \Phi(k+1, i+1)[H_{i-m} \\ &\quad + (H_i - H_{i-m})][EX_i - X_i]\Psi^\tau(k+1, i+1). \end{aligned} \tag{62}$$

Note that by (17), $\forall i \geq 1, \forall m \geq 1$

$$\begin{aligned} H_i - H_{i-m} &= R^{-1}(P_i - P_{i-m}) \\ &= \mu R^{-1} \sum_{k=i-m+1}^{i} \left[-\frac{P_{k-1}\varphi_k \varphi_k^\tau P_{k-1}}{R + \mu \varphi_k^\tau P_{k-1} \varphi_k} + Q \right]. \end{aligned} \tag{63}$$

From this and Lemma 5.1-i), we have $\forall m \geq 1$

$$\|H_i - H_{i-m}\|_p = O(\mu m), \qquad \forall p < 2t, \mu \in (0,1). \tag{64}$$

Hence, by Lemma 5.2-iii), (59), and the Hölder inequality, for any $s < 4t/5$

$$\left\| \sum_{i=1}^{k} \Phi(k+1, i+1)(H_i - H_{i-m}) \times [EX_i - X_i]\Psi^\tau(k+1, i+1) \right\|_s = O\left(\mu m \sum_{i=1}^{k} (1-\mu\alpha)^{(k-i)}\right), \quad \text{for some } \alpha \in (0,1)) = O(m), \quad \forall m \geq 1. \tag{65}$$

Now, consider the following decomposition

$$\begin{aligned} X_i - EX_i &= \sum_{j=0}^{m-1} \delta_j(i) + \{E[X_i|\mathcal{G}_{i-m}] - EX_i\}, \\ \delta_j(i) &\triangleq E[X_i|\mathcal{G}_{i-j}] - E[X_i|\mathcal{G}_{i-j-1}], \\ & 0 \leq j \leq m-1 \\ \mathcal{G}_i &= \sigma\{\varphi_s, s \leq i\} \end{aligned} \tag{66}$$

and note that $\{H_{i-m}\delta_j(i), \mathcal{G}_{i-j}, i \geq 1\}$ is a martingale difference for each $0 \leq j \leq m-1$, and that

$$\sup_{i,j} \|H_{i-m}\delta_j(i)\|_p = O(1), \qquad \forall p < 2t, 1 \leq j \leq m-1$$

then by Lemma 5.4 we have for small $\mu > 0$ and for $s < 4t/5$

$$\sum_{j=0}^{m-1} \left\| \sum_{i=1}^{k} \Phi(k+1, i+1)[H_{i-m}\delta_j(i)]\Psi^\tau(k+1, i+1) \right\|_s = O(\mu^{-(1/2)} m). \tag{67}$$

Next, by Condition K2), (59), and Lemma 5.2-iii), it follows that for any $s < 4t/5$

$$\left\| \sum_{i=1}^{k} \Phi(k+1, i+1) H_{i-m}\{E[X_i|\mathcal{G}_{i-m}] - EX_i\} \times \Psi^\tau(k+1, i+1) \right\|_s = O(\mu^{-1}\phi(m)). \tag{68}$$

Finally, substituting (65)–(68) into (62) we see that for small μ and any $m \geq 1$

$$\|P_k - \overline{P}_k\|_s = O(\sqrt{\mu} m + \phi(m)), \qquad \forall s < \frac{4t}{5}.$$

Hence Lemma 5.6 is true. #

Proof of Theorem 5.3: First, we prove Condition A1). By Lemma 5.6, Lemma 5.2-iii), and (56) from Proposition 4.2 we know that $\{F_k\} \in \mathcal{S}$. Consequently, by Lemma 5.1, we have

$$\{F_k\} \in \mathcal{S}_p \bigcap \mathcal{S}, \qquad \forall p < 2t \tag{69}$$

which verifies Condition A1).

Next, by Proposition 4.3 and Lemma 5.6, it can be seen that A2) holds for any $q < 4t/7$.

Finally, to verify Condition A3), we note that by Lemma 5.1-i)

$$\{L_i\} \in \mathcal{L}_r(\mu^*), \qquad \forall r < 4t, \quad \mu^* \in (0,1)$$

and

$$\{F_i\} \in \mathcal{L}_{2q}(\mu^*), \qquad \forall q < \frac{4t}{7}, \quad \mu^* \in (0,1).$$

But $(1/2 - 1/2t - 3/2 \cdot 2/t)^{-1} < 4t$, hence Condition A3) is also true.

Thus, Theorem 4.1 can be applied to prove Theorem 3.1 (by taking $t = \infty$), while the expression for Π_{t+1} will follow from Corollary 4.1 if we can show that

$$\|P_k R^{-1} - R_k\|_s = O(\delta(\mu)), \qquad \forall s < \frac{4t}{5}.$$

This, however, follows directly from Lemma 5.6 since it can be verified (See Appendix B) that

$$\|R_k - \overline{P}_k R^{-1}\| = O(\mu).$$

This completes the proof of Theorem 5.3.

VI. Conclusions

In this article, we have presented a number of results by which the true covariance matrix of the parameter tracking error can be approximated by a matrix that can be computed by a much simpler equation. As mentioned above, there is a considerable literature on this problem. We may point to the following aspects of how the results of this paper extends and contains earlier ones.

- The approximation in Theorems 3.1 and 4.1 is explicit. It involves the true error covariance matrix and the approximating one. The result is not asymptotic. (It is of interest, however, only for small gains μ.) It is applicable both during the transient and over infinite time horizons.
- We have treated the whole family of the most commonly used tracking algorithms in one general result (Theorem 3.1).
- This includes what appears to be the first formal treatment of the Kalman Filter as a tracking algorithm in this respect.
- For the LMS case, the regressors are not assumed to be bounded and/or independent (Theorem 5.1).
- For the RLS case, the Riccati equation for calculating Π_k is simpler than those earlier used [See (24)].
- We have also, in Theorem 4.1, given a general result on the tracking error under quite weak assumptions. Together with the results of Section V, this may serve as a tool kit for building specialized results for particular algorithms.

The basic result is quite easy to understand, and its practical implications for dealing with the key features of tracking algorithms are quite important. The basis for the important compromise between tracking ability and noise sensitivity lies in these expressions. Nevertheless, the analysis and the proof of the result turn out to be surprisingly technically complicated.

APPENDIX A
PROOF OF THEOREM 4.1

We first prove several lemmas.

Lemma A.1: Let $\alpha \in (0,1)$ be a constant. Then $\forall \mu \in (0,1)$

i) $\sup_{k\geq 0}(1-\alpha\mu)^k\sqrt{k} = O(\mu^{-(1/2)})$

ii) $\sum_{k=0}^{\infty}(1-\alpha\mu)^k k = O(\mu^{-2})$

iii) $\sum_{k=0}^{\infty}(1-\alpha\mu)^k\sqrt{k} = O(\mu^{-(3/2)})$

where the "O" constant depends only on α.

Proof: Denote $\beta = e^{-\alpha} < 1$, then we have

$$(1-\alpha\mu)^k \leq \beta^{k\mu}, \qquad \forall \mu \in (0,1).$$

Now, note that $\sup_{x\geq 0}\beta^x\sqrt{x} < \infty$, then replace x by $k\mu$ gives assertion i). To prove ii), observe that

$$\sum_{k=0}^{\infty}\beta^{k\mu}k = \frac{\partial}{\partial\mu}\left(\sum_{k=0}^{\infty}\frac{1}{\log\beta}\beta^{k\mu}\right) = \frac{1}{\log\beta}\frac{\partial}{\partial\mu}\left(\frac{1}{1-\beta^{\mu}}\right) \leq \frac{1}{(\beta\log\beta)^2}\frac{1}{\mu^2}.$$

Hence ii) is also true. Finally, iii) follows from ii) via the Schwarz inequality.

Now, define

$$\Phi(k+1,i) = (I-\mu F_k)\Phi(k,i),$$
$$\Phi(i,i) = I, \qquad \forall k \geq i \geq 0.$$

Lemma A.2: Let $\{e_k\} \in \mathcal{M}_r, r \geq 1$, and $\{F_k\} \in \mathcal{S}_p(\mu^*) \cap \mathcal{L}_t(\mu^*), p \geq 1, t \geq 1, \mu^* \in (0,1)$. Then for $s \triangleq (r^{-1}+p^{-1}+t^{-1})^{-1}$

$$\left\|\sum_{i=0}^{k}\Phi(k+1,i+1)e_i\right\|_s = O(\mu^{-(1/2)}), \qquad \forall \mu \in (0,\mu^*].$$

Proof: Set $S(k,i) = \Sigma_{j=i}^k e_j$, then by summation by parts

$$\begin{aligned}\sum_{i=0}^{k}\Phi(k+1,i+1)e_i &= \sum_{i=0}^{k}\Phi(k+1,i+1)[S(k,i)-S(k,i+1)]\\ &= \Phi(k+1,1)S(k,0)+\sum_{i=1}^{k}[\Phi(k+1,i+1)\\ &\quad -\Phi(k+1,i)]S(k,i)\\ &= \Phi(k+1,1)S(k,0)+\mu\sum_{i=1}^{k}\Phi(k+1,i+1)F_iS(k,i).\end{aligned}$$

So by the Hölder inequality and Lemma A.1

$$\begin{aligned}\left\|\sum_{i=0}^{k}\Phi(k+1,i+1)e_i\right\|_s &\leq \|\Phi(k+1,1)\|_p\cdot\|S(k,0)\|_r\\ &\quad +\mu\sum_{i=1}^{k}\|\Phi(k+1,i+1)\|_p\cdot\|F_i\|_t\cdot\|S(k,i)\|_r.\\ &= O([1-\alpha\mu]^k\sqrt{k})+O\left(\mu\sum_{i=0}^{k}(1-\alpha\mu)^{k-i}\sqrt{k-i}\right)\\ &= O(\mu^{-(1/2)})+O(\mu^{-(1/2)}) = O(\mu^{-(1/2)}),\\ &\quad \forall\mu\in(0,\mu^*].\end{aligned} \tag{A.1}$$

#

Note that by (19)

$$\tilde{\theta}_{k+1} = \Phi(k+1,0)\tilde{\theta}_0+\mu\sum_{i=0}^{k}\Phi(k+1,i+1)\cdot[-\mu L_iv_i+\gamma w_{i+1}].$$

Applying Lemma A.2 we immediately have the following.

Lemma A.3: Let $\{L_iv_i\} \in \mathcal{M}_r$, $\{w_i\} \in \mathcal{M}_r$, $\{F_k\} \in \mathcal{S}_p(\mu^*) \cap \mathcal{L}_t(\mu^*)$, $r \geq 1$, $p \geq 1, t \geq 1$, $\mu^* \in (0,1)$. Then for $s \triangleq (r^{-1}+p^{-1}+t^{-1})^{-1}$, and $\forall\mu \in (0,\mu^*]$

$$\|\tilde{\theta}_{k+1}\|_s = O\left(\sqrt{\mu}+\frac{\gamma}{\sqrt{\mu}}+(1-\alpha\mu)^k\right)$$

holds for all $k \geq 1$, where $\alpha \in (0,1)$ is a constant. #

Now, let us introduce a new sequence $\{\overline{\theta}_k\}$

$$\begin{aligned}\overline{\theta}_{k+1} &= (I-\mu E[F_k])\overline{\theta}_k-\mu L_kv_k+\gamma w_{k+1},\\ \overline{\theta}_0 &= \tilde{\theta}_0.\end{aligned} \tag{A.2}$$

Similar to the proof above, we have the following lemma.

Lemma A.4: Let $\{L_iv_i\} \in \mathcal{M}_r, \{w_i\} \in \mathcal{M}_r$, $r \geq 1$, $\{F_i\} \in \mathcal{S}(\mu^*)$, $\mu^* \in (0,1)$. Then

$$\|\overline{\theta}_{k+1}\|_r = O\left(\sqrt{\mu}+\frac{\gamma}{\sqrt{\mu}}+(1-\alpha\mu)^k\right), \qquad \forall\mu\in(0,\mu^*]$$

where $\alpha \in (0,1)$ is a constant (without loss of generality, it may be taken as the same as that in Lemma A.3). #

Proof of Theorem 4: By (A.2) and Condition A4) it is evident that

$$\hat{\Pi}_{k+1} = E[\overline{\theta}_{k+1}\overline{\theta}_{k+1}^{\tau}], \qquad \forall k \geq 0.$$

Hence by Schwarz inequality

$$\begin{aligned}&\|E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^{\tau}]-\hat{\Pi}_{k+1}\|\\ &\quad = \|E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^{\tau}-\overline{\theta}_{k+1}\overline{\theta}_{k+1}^{\tau}]\|\\ &\quad = \|E[(\tilde{\theta}_{k+1}-\overline{\theta}_{k+1})\tilde{\theta}_{k+1}^{\tau}+\overline{\theta}_{k+1}(\tilde{\theta}_{k+1}^{\tau}-\overline{\theta}_{k+1}^{\tau})]\|\\ &\quad \leq \|\tilde{\theta}_{k+1}-\overline{\theta}_{k+1}\|_2\cdot[\|\tilde{\theta}_{k+1}\|_2+\|\overline{\theta}_{k+1}\|_2].\end{aligned} \tag{A.3}$$

Denote

$$\varepsilon_k(\alpha) = \sqrt{\mu}+\frac{\gamma}{\sqrt{\mu}}+(1-\alpha\mu)^k. \tag{A.4}$$

By Lemmas A.3 and A.4, we know that

$$\|\tilde{\theta}_{k+1}\|_2+\|\overline{\theta}_{k+1}\|_2 = O(\varepsilon_k(\alpha)). \tag{A.5}$$

So we need only to consider the first term on the right-hand side of (A.3).

By (19) and (A.2), it is seen that

$$\tilde{\theta}_{k+1}-\overline{\theta}_{k+1} = (I-\mu E[F_k])(\tilde{\theta}_k-\overline{\theta}_k)+\mu(E[F_k]-F_k)\tilde{\theta}_k.$$

Define

$$\Psi(k+1,i) = (I - \mu E[F_k])\Psi(k,i), \quad \Psi(i,i) = I, \quad \forall k \geq i.$$

Then we have ($\forall k \geq 0$)

$$\begin{aligned}\tilde{\theta}_{k+1} - \bar{\theta}_{k+1} &= \mu \sum_{i=0}^{k} \Psi(k+1,i+1)(E[F_i] - F_i)\tilde{\theta}_i \\ &= \mu \sum_{i=0}^{m-1} \Psi(k+1,i+1)(E[F_i] - F_i)\tilde{\theta}_i \\ &\quad + \mu \sum_{i=m}^{k} \Psi(k+1,i+1)(E[F_i] - F_i) \\ &\quad \times \{\tilde{\theta}_{i-m} + (\tilde{\theta}_i - \tilde{\theta}_{i-m})\} \end{aligned} \tag{A.6}$$

where $m = m(\mu)$ which is defined by

$$m(\mu) = \operatorname*{argmin}_{m \geq 1}[\sqrt{\mu} m + \phi(m,\mu)]. \tag{A.7}$$

Note that

$$\sqrt{\mu} m(\mu) \leq \sqrt{\mu} m(\mu) + \phi(m(\mu),\mu) \leq \sqrt{\mu} + \phi(1,\mu)$$

which implies that

$$m(\mu) \leq 1 + \frac{\phi(1,\mu)}{\sqrt{\mu}} \leq \frac{c}{\sqrt{\mu}}, \qquad \forall \mu \in (0,1)$$

for some constant $c > 0$. Consequently, for any $\alpha \in (0,1)$

$$(1-\alpha\mu)^{-m(\mu)} \leq (1-\alpha\mu)^{-c/\sqrt{\mu}} \to 1, \qquad \text{as} \quad \mu \to 0$$

hence $(1-\alpha\mu)^{-m(\mu)}, \mu \in (0,1)$, is a bounded function for any $\alpha \in (0,1)$. In the sequel, we will frequently use this fact without explanations and will also drop the variable μ in $m(\mu)$ in what follows.

Now, denote

$$s = \left(\frac{1}{r} + \frac{1}{p} + \frac{1}{2q}\right)^{-1}. \tag{A.8}$$

Then by (A.4), Lemma A.3, and Conditions A1)–A4) we have $\forall \mu \in (0,\mu^*]$

$$\|\tilde{\theta}_{k+1}\|_s = O(\varepsilon_k(\alpha))). \tag{A.9}$$

Note that the number s defined by (A.8) satisfies $[s^{-1} + (2q)^{-1}]^{-1} > 2$, we have by (A.9)

$$\mu \left\| \sum_{i=0}^{m-1} \Psi(k+1,i+1)(E[F_i] - F_i)\tilde{\theta}_i \right\|_2 = O(\mu m \varepsilon_k(\alpha)). \tag{A.10}$$

So, we need only to consider the last term in (A.6) for $k \geq m$.

Note that by (19), $\forall i \geq m$

$$\tilde{\theta}_i - \tilde{\theta}_{i-m} = \sum_{j=i-m}^{i-1} [-\mu F_j \tilde{\theta}_j - \mu L_j v_j + \gamma w_{j+1}]. \tag{A.11}$$

So by denoting

$$u = \left(\frac{1}{s} + \frac{1}{2q}\right)^{-1} \tag{A.12}$$

and applying the Hölder inequality to (A.11) and by noting that $\{-\mu L_j v_j + \gamma w_{j+1}\} \in \mathcal{M}_r \subset \mathcal{M}_u$, we have for any $i \geq m$

$$\begin{aligned}&\|\tilde{\theta}_i - \tilde{\theta}_{i-m}\|_u \\ &\leq \sum_{j=i-m}^{i-1} \mu \|F_j\|_{2q} \cdot \|\tilde{\theta}_j\|_s \\ &\quad + \left\| \sum_{j=i-m}^{i-1} (-\mu L_j v_j + \gamma w_{j+1}) \right\|_u \\ &= O\left(\sum_{j=i-m}^{i-1} \mu \varepsilon_j(\alpha)\right) + O(\sqrt{m}[\mu + \gamma]) \\ &= O\left(\sqrt{\mu} m \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right]\right) + O(\mu m (1-\alpha\mu)^{i-m}). \end{aligned} \tag{A.13}$$

for $\mu \in (0,\mu^*]$.

By (A8) and (A.12) and the definition of r in Condition A3), it is readily verified that $1/2 = 1/2q + 1/u$. Hence, by (A.13) and Condition A1) we obtain ($\forall k \geq m$)

$$\begin{aligned}&\mu \left\| \sum_{i=m}^{k} \Psi(k+1,i+1)(E[F_i] - F_i)(\tilde{\theta}_i - \tilde{\theta}_{i-m}) \right\|_2 \\ &\leq \mu \sum_{i=m}^{k} \|\Psi(k+1,i+1)\| \\ &\quad \cdot \|E[F_i] - F_i\|_{2q} \cdot \|\tilde{\theta}_i - \tilde{\theta}_{i-m}\|_u \\ &= O\left(\sqrt{\mu} m \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right]\right) \\ &\quad + O(\mu^2 m (k-m)(1-\alpha\mu)^{k-m}) \\ &= O\left(\sqrt{\mu} m \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right]\right) \\ &\quad + O(\mu^2 m \sup_{k \geq 0}\{k(1-\alpha\mu)^k\}) \\ &= O\left(\sqrt{\mu} m \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right]\right) \end{aligned} \tag{A.14}$$

where for the last relationship we have used Lemma A.1-i).

Now, set for $j \geq 0, i \geq m$

$$\delta_j(i) \triangleq E[F_i | \mathcal{F}_{i-j}] - E[F_i | \mathcal{F}_{i-j-1}].$$

Then $\forall i \geq m$

$$F_i - EF_i = \sum_{j=0}^{m-1} \delta_j(i) + E[F_i | \mathcal{F}_{i-m}] - EF_i. \tag{A.15}$$

For any fixed $0 \leq j \leq m-1$, denote $e_i = \delta_j(i)\tilde{\theta}_{i-m}$, then it is obvious that $\{e_i, \mathcal{F}_{i-j}, i \geq m\}$ is a martingale difference sequence and that by (A.9) and the fact that $1/2 \geq 1/2q + 1/s$

$$\|e_i\|_2 \leq 2\|F_i\|_{2q} \cdot \|\tilde{\theta}_{i-m}\|_s = O(\varepsilon_{i-m}(\alpha)). \qquad i \geq m.$$

Consequently, denote $S(k,i) \triangleq \Sigma_{j=i}^{k} e_j$, we have for any $k \geq i > m$

$$\|S(k,i)\|_2 = \left\{\sum_{j=i}^{k} Ee_j^2\right\}^{1/2} = O\left(\sqrt{k-i+1}(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}) + (1-\alpha\mu)^{i-m}\mu^{-1/2}\right).$$

Hence similar to (A.1) we have by (A.4) and Lemma A.1

$$\begin{aligned}\mu\left\|\sum_{i=m}^{k}\Psi(k+1,i+1)\sum_{j=0}^{m-1}\delta_j(i)\tilde{\theta}_{i-m}\right\|_2 &\leq \mu\sum_{j=0}^{m-1}\left\|\sum_{i=m}^{k}\Psi(k+1,i+1)e_i\right\|_2 \\ &\leq \mu\sum_{j=0}^{m-1}\|\Psi(k+1,m+1)\|\cdot\|S(k,m)\|_2 \\ &\quad + \mu^2\sum_{j=0}^{m-1}\sum_{i=m+1}^{k}\|\Psi(k+1,i+1)\| \cdot\|EF_i\|\cdot\|S(k,i)\|_2 \\ &= O\left(\mu m\left(\sqrt{\mu}+\frac{\gamma}{\sqrt{\mu}}\right)\left\{(1-\alpha\mu)^{k-m}\sqrt{k-m} + \mu\sum_{i=m+1}^{k}(1-\alpha\mu)^{k-i}\sqrt{k-i}\right\}\right) \\ &\quad + O\left(\sqrt{\mu}m\left(1-\frac{\alpha}{2}\mu\right)^{k-m}\right) \\ &= O\left(\sqrt{\mu}m\varepsilon_k\left(\frac{\alpha}{2}\right)\right), \qquad \forall k \geq m. \end{aligned} \tag{A.16}$$

Note that by (A.8) and the definition of r in Condition A3), we have $1/2 = 1/q + 1/s$. Hence by (A.9) and Condition A2)

$$\begin{aligned}\mu\left\|\sum_{i=m}^{k}\Psi(k+1,i+1)\{E[F_i|\mathcal{F}_{i-m}] - EF_i\}\tilde{\theta}_{i-m}\right\|_2 &\leq \mu\sum_{i=m}^{k}\|\Psi(k+1,i+1)\|\cdot\|E[F_i|\mathcal{F}_{i-m}] - EF_i\|_q\cdot\|\tilde{\theta}_{i-m}\|_s \\ &= O\left(\phi(m,\mu)\varepsilon_k\left(\frac{\alpha}{2}\right)\right)\end{aligned} \tag{A.17}$$

for any $\mu \in (0,\mu^*]$.

Combining (A.16) with (A.17) and noting (A.15), we see that

$$\mu\left\|\sum_{i=0}^{k}\Psi(k+1,i+1)[F_i - E(F_i)]\tilde{\theta}_{i-m}\right\|_2 = O\left([\sqrt{\mu}m + \phi(m,\mu)]\varepsilon_k\left(\frac{\alpha}{2}\right)\right)$$

this in conjunction with (A.14) and (A.10) yields

$$\|\tilde{\theta}_{k+1} - \overline{\theta}_{k+1}\|_2 = O\left([\sqrt{\mu}m + \phi(m,\mu)]\varepsilon_k\left(\frac{\alpha}{2}\right)\right) = O\left(\sigma(\mu)\varepsilon_k\left(\frac{\alpha}{2}\right)\right).$$

Finally, substituting this and (A.5) into (A.3), we obtain $\forall\mu \in (0,\mu^*]$

$$\|E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^{\tau}] - \hat{\Pi}_{k+1}\| \leq O\left(\sigma(\mu)\left[\varepsilon_k\left(\frac{\alpha}{2}\right)\right]^2\right). \tag{A.18}$$

Finally, substituting (A.4) into (A.18) we see that Theorem 4.1 is true.

Appendix B
Analysis of the Riccati Equations (24) and (25)

Lemma B.1: Let Condition P1 be satisfied and $\{S_k\}$ be bounded. Then the following two assertions hold:

i) For $\{R_k\}$ recursively defined by (24) (with $R_0 = P_0 > 0$), we have for all small $\mu > 0$

$$\|R_k - \overline{P}_k\| = O(\mu), \qquad \forall k$$

where $\{\overline{P}_k\}$ is defined recursively by (45).

ii) If $\{R_k\}$ is defined by (25) (with $R_0 = P_0R^{-1}$) and $\{\overline{P}_k\}$ is defined by (54)), then for all small $\mu > 0$

$$\|R_k - \overline{P}_kR^{-1}\| = O(\mu), \qquad \forall k.$$

Proof: i) We first show that for all $m \geq 0$ and small $\mu > 0$

$$R_{mh} > 0, \quad \text{tr}(R_{mh}) < 2dh\delta^{-1} + \text{tr}(R_0) \tag{B.1}$$

where δ and h are defined in P1 and $d = \dim(\varphi_k)$.

Let (B.1) hold for some $m \geq 0$, and we proceed to prove that (B.1) also holds with m replaced by $m+1$. By (24), we know that $R_k \leq (1+\mu)R_{k-1}$, and that

$$R_k = R_{k-1}^{1/2}[(1+\mu)I - \mu R_{k-1}^{1/2}S_kR_{k-1}^{1/2}]R_{k-1}^{1/2}.$$

Consequently, it is easy to see that for all suitably small $\mu > 0$

$$R_k > 0, \qquad \forall k \in [mh,(m+1)h]. \tag{B.2}$$

Now, iterating (24) we see that $R_{(m+1)h}$ can be represented by $\{R_{mh}, S_k, mh+1 \leq k \leq (m+1)h\}$ as follows

$$R_{(m+1)h} = (1+\mu h)R_{mh} - \mu\sum_{k=mh+1}^{(m+1)h}R_{mh}S_kR_{mh} + \mu^2B_m \tag{B.3}$$

where $\|B_m\| \leq B$ and B depends only on h and the upper bound of $\text{tr}(R_{mh})$ in (B.1).

Let us denote $r_m \triangleq \text{tr}(R_{mh})$. By the inequality $\text{tr}(R_{mh}^2) \geq d^{-1}r_m^2$ and Condition P1 we get from (B.3) that

$$r_{m+1} \leq (1+\mu h)r_m - \mu\delta d^{-1}r_m^2 + \mu^2Bd. \tag{B.4}$$

Note that the function $(1+\mu h)x - \mu\theta d^{-1}x^2$ is positive and increasing for $x \in (0,(1+\mu h)d/2\mu\theta)$. Thus, if μ is small enough that

$$M \triangleq 2dh\delta^{-1} + \text{tr}(R_0) < \frac{(1+\mu h)d}{2\mu\delta}.$$

Then by (B.4) and the assumption $r_m \leq M$, we have

$$\begin{aligned} r_{m+1} &\leq (1+\mu h)M - \mu\delta d^{-1}M^2 + \mu^2 Bd. \\ &= M[1+\mu h - \mu\delta d^{-1}M + \mu^2 BdM^{-1}] \\ &\leq M[1-\mu h + \mu^2 BdM^{-1}] \\ &\leq M. \qquad \text{for small } \mu > 0. \end{aligned}$$

Hence by induction, (B.1) is proved. Consequently, (B.2) holds for all $m \geq 0$.

Furthermore, by (B.1) and the inequality

$$R_k \leq (1+\mu)R_{k-1}$$

it is easy to see that $\{R_k\}$ is uniformly bounded with respect to small $\mu > 0$.

Next, by the matrix inversion formula from (24) we have

$$R_k^{-1} = \frac{R_{k-1}^{-1}}{1+\mu} + \frac{\mu}{1+\mu} S_k^{1/2} \cdot [(1+\mu)I - \mu S_k^{1/2} R_{k-1} S_k^{1/2}]^{-1} S_k^{1/2}.$$

Since $\{R_k\}$ is bounded, from this it is evident that $\{R_k^{-1}\}$ is also bounded for small $\mu > 0$ and that

$$R_k^{-1} = (1-\mu)R_{k-1}^{-1} + \mu S_k + O(\mu^2).$$

This in conjunction with (45) gives

$$\|R_k^{-1} - \overline{P}_k^{-1}\| = O(\mu).$$

Hence, by the boundedness of $\{R_k\}$ and $\{\overline{P}_k\}$ [see Lemma 5.2-i)], we get

$$\|R_k - \overline{P}_k\| = \|R_k(\overline{P}_k^{-1} - R_k^{-1})\overline{P}_k\| = O(\mu).$$

This proves the first assertion of the lemma

ii) Similarly to the above analysis for (24), by (25) it can be shown that $\{R_k\}$ is a positive definite sequence and is uniformly bounded with respect to all small $\mu > 0$.

Now, rewrite (25) as

$$R_k = R_{k-1}^{1/2}(I - \mu R_{k-1}^{1/2} S_k R_{k-1}^{1/2}) R_{k-1}^{1/2} + \mu Q R^{-1}.$$

We see that

$$\lambda_{\min}(R_k) \geq (1-\mu M)\lambda_{\min}(R_{k-1}) + \mu R^{-1}\lambda_{\min}(Q)$$

where $M \stackrel{\delta}{=} \sup_k \|R_{k-1}^{1/2} S_k R_{k-1}^{1/2}\|$. This implies that $\{R_k^{-1}\}$ is also uniformly bounded with respect to all small $\mu > 0$.

Note that (25) may also be rewritten as

$$\begin{aligned} R_k &= (I - \mu R_{k-1}S_k)R_{k-1}(I - \mu R_{k-1}S_k)^\tau \\ &\quad + \mu R_{k-1} S_k^{1/2}(I - \mu S_k^{1/2} R_{k-1} S_k^{1/2}) S_k^{1/2} R_{k-1} \\ &\quad + \mu Q R^{-1}. \end{aligned}$$

Hence by applying Theorem 3.4 in [9] we know that $\{R_{k-1}S_k\} \in \mathcal{S}$, where $\mathcal{S}$ is defined by (29). Furthermore, by Proposition 4.1 we also know that $\{R_{k-1}S_k + O(\mu)\} \in \mathcal{S}$.

Next, by the matrix inversion formula it can be verified that

$$R_k = (R_{k-1}^{-1} + \mu S_k)^{-1} + \mu R^{-1} Q + O(\mu^2).$$

Similarly to the proof of Lemma 5.5, from this and (54) we have

$$\begin{aligned} R_k - \overline{P}_k R^{-1} &= [I - \mu R_{k-1}S_k + O(\mu^2)] \\ &\quad \times [R_{k-1} - \overline{P}_{k-1}R^{-1}](I - \mu \overline{F}_k^\tau) + O(\mu^2) \end{aligned} \tag{B.5}$$

where $\{\overline{F}_k\}$ is defined as in Lemma 5.5. Since both $\{R_{k-1}S_k + O(\mu)\}$ and $\{\overline{F}_k\}$ belong to $\mathcal{S}$, by iterating the above equation it is easy to see that the desired result holds.

REFERENCES

[1] A. Benveniste, "Design of adaptive algorithms for the tracking of time-varying systems," *Int. J. Adaptive Contr. Sig. Proces.*, vol. 1, pp. 3–29, 1987.

[2] A. Benveniste, M. Metivier, and P. Priouret, *Algorithmes Adaptifs Et Approximations Stochastiques*, Masson, 1987.

[3] S. Bittanti and M. Campi, "A fake information matrix approach to the analysis of finite memory RLS identification techniques," in *Proc. IEEE Conf. Decis. Contr.*, San Antonio, TX, 1993, pp. 2026–2031.

[4] W. Gardner, "Nonstationary learning characteristics of the LMS algorithms: A general study, analysis and critique," *IEEE Trans. Circuits Syst.*, vol. CAS-34, no. 10, pp. 1199–1207, 1987.

[5] L. Gerencser, "Almost sure exponential stability of random linear differential equations," *Stochastics and Stochastics Reports*, vol. 36, pp. 91–107, 1991.

[6] ——, "The law of the cubic root," in *Selected Papers from the Ninth IFAC/IFORS Symposium on System Identification*, C. Banyasz and L. Keviczky, Eds., Budapest, Hungary, July 8–12, 1991.

[7] L. Guo, "Estimating time-varying parameters by Kalman filter based algorithms: Stability and convergence," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 141–147, 1990.

[8] ——, "Stability of recursive stochastic tracking algorithms," *SIAM J. Control and Optimization*, vol. 32, pp. 1195–1225, 1994.

[9] L. Guo and L. Ljung, "Exponential stability of general tracking algorithms," *IEEE Trans. Autom. Control*, 1995, this issue.

[10] L. Guo, L. Ljung, and P. Priouret, "Performance analysis of the forgetting factor RLS algorithm," *Int. J. Adaptive Contr. Sig. Proces.*, vol. 7, pp. 525–537, 1993.

[11] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Applications.* New York: Academic, 1980.

[12] L. Ljung and S. Gunnarsson, "Adaptive tracking in system identification—A survey, *Automatica*, vol. 26, no. 1, pp. 7–22, 1990.

[13] L. Ljung and P. Priouret, "Remarks on the mean square tracking error," *Int. J. Adaptive Contr.*, vol. 5, no. 6, pp. 395–403, 1991.

[14] ——, "A result of the mean square error obtained using general tracking algorithms," *Int. J. Adaptive Contr.*, vol. 5, no. 4, pp. 231–250, 1991.

[15] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification.* Cambridge, MA: MIT Press, 1983.

[16] O. Macchi and E. Eweda, "Second-order convergence analysis of stochastic adaptive linear filtering," *IEEE Trans. Automat. Contr.*, vol. AC-28, no. 1, pp. 76–85, 1983.

[17] M. Niedzwiecki and L. Guo, "Nonasymptotic results for finite memory WLS filters," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 198–206, 1991.

[18] R. Ravikanth and S. P. Meyn, "Performance evaluation in identification and adaptive control of time varying systems," in *Proc. 33rd IEEE Conf Decis. Contr.*, Lake Buena Vista, FL, 1994.

[19] R. J. Serfling, "Contributions to central limit theory for dependent variables," *Ann. Math. Statistics*, vol. 39, no. 4, pp. 1158–1175, 1968.

[20] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance.* Englewood Cliffs, NJ: Prentice-Hall, 1994.

[21] B. Widrow and M. E. Hoff, Jr, "Adaptive switching circuits," *IRE WESCON Convention Record*, part 4, pp. 96–104, 1960.

[22] B. Widrow and S. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

Lei Guo, for a photograph and biography, see p. 1387 of this TRANSACTIONS.

Lennart Ljung, for a photograph and biography, see p. 1387 of this TRANSACTIONS.