Journal of Mathematical Systems, Estimation, and Control
Vol. 4, No. 3, 1994, pp. 1–25

Performance of RLS Identification Algorithms with Forgetting Factor: A Φ–Mixing Approach*

Marco Campi†

Abstract

In this paper, systems with unknown time-varying parameters and subject to stochastic disturbances are considered. The problem of tracking the parameters is tackled by resorting to a class of adaptive recursive least squares algorithms, equipped with variable forgetting factor. The basic assumption in the analysis is that the observation vector, the noise and the parameter drift are stochastic processes satisfying a ϕ–mixing condition. Furthermore, it is assumed that the observation vector satisfies an excitation condition imposed on its minimum power. It is shown that the algorithm provides estimates with bounded error whenever the so-called "covariance matrix" of the algorithm keeps bounded. Moreover, the size of such a matrix can be controlled by a suitable choice of the feasible range for the forgetting factor.

Key words: adaptive identification, time varying systems, recursive least squares, tracking properties, stochastic analysis

AMS Subject Classifications: 93E12

1 Introduction and Preliminaries

1.1 The RLS algorithm

Denoting by $\vartheta^\circ(t) \in R^{n \times m}$ a matrix of unknown parameters, consider the following system

$$y(t) = \vartheta^\circ(t)'\varphi(t) + d(t), \tag{1.1.a}$$

*Received July 1, 1991; received in final form December 5, 1991. Summary appeared in Vol. 4 Number 3, 1994.

†This paper has been supported by Centro Teoria dei Sistemi of the Consiglio Nazionale delle Ricerche and Ministero della Universita' e Ricerca Scientifica e Tecnologica.

$$\vartheta^\circ(t+1) = \vartheta^\circ(t) + \delta\vartheta^\circ(t). \tag{1.1.b}$$

In equations (1.1), $y(t) \in R^m$ is a vector of observed outputs, $\varphi(t) \in R^n$ is the measured observation vector, $d(t) \in R^m$ is the additive noise and $\delta\vartheta^\circ(t) \in R^{n\times m}$ is the parameter drift term. The initialization of equation (1.1.b), at time $t = 1$, is assumed to be L^2-bounded (i.e. $\| \vartheta^\circ(1) \|_{L^2} := E^{1/2}[\| \vartheta^\circ(1) \|^2]$ bounded).

A common technique for the estimation of the unknown parameter $\vartheta^\circ(t)$ is the Recursive Least Squares (RLS) algorithm with forgetting factor, [19], given by the equations

$$\epsilon(t) = y(t)' - \varphi(t)'\hat{\vartheta}(t-1) \tag{1.2.a}$$

$$a(t) = (1 + \varphi(t)'P(t-1)\varphi(t))^{-1} \tag{1.2.b}$$

$$K(t) = a(t)P(t-1)\varphi(t) \tag{1.2.c}$$

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + K(t)\epsilon(t) \tag{1.2.d}$$

$$P(t) = \frac{1}{\mu(t)}[P(t-1) - a(t)P(t-1)\varphi(t)\varphi(t)'P(t-1)] + Q. \tag{1.2.e}$$

The recursive equations (1.2.d) and (1.2.e) are initialized with given (deterministic) vector $\hat{\vartheta}(0)$ and matrix $P(0) = P(0)' > 0$, respectively. In equation(1.2.e), Q is a positive definite matrix and $\mu(t)$ is the forgetting factor, which is assumed to satisfy the constraint

$$0 < \mu_0 \leq \mu(t) \leq \mu_1 < 1. \tag{1.3}$$

Remark 1.1. Many recursive algorithm used in identification and control engineering, including Exponential Forgetting, Constant Trace, Prediction Error Forgetting, to quote but a few, can be seen as particular cases of equations (1.2) with $Q = 0$.

Remark 1.2. In the implementation of algorithm (1.2), the positive definite matrix Q in equation(1.2.e) guarantees that the algorithm keeps a certain degree of responsiveness in any operating condition (even when the observation vectors take quite large values).

In this paper, we will study the performance of algorithm (1.2) in a fully stochastic framework. Our main concern will be the analysis of the conditions under which the algorithm provides estimates of the time-varying parameter $\vartheta^\circ(t)$ such that the corresponding error keeps L^2-bounded.

1.2 An overview of the existing literature

A large stream of literature has been devoted to the analysis of adaptive RLS algorithms. Most papers deal with those variants characterized by an algorithm gain which progressively switches off, [2], [15], [16]. However, such algorithms asymptotically behave as the standard RLS, so that adaptivity is lost in the long run. In the truly adaptive RLS context, a stream of research focuses on the exponential convergence of the estimation to the true parameter value when this is assumed to be *time-invariant*, see e.g. [4], [5], [6], [7], [9], [14]. As a matter of fact, exponential convergence in the constant case implies a certain degree of tracking capability in the time-varying case, [1]. In [4], [5], [14], it is assumed that the data are free of disturbances and the observation vector satisfies some persistent excitation condition of deterministic type. On the contrary, a stochastic notion of persistent excitation is proposed in [6] and [7], where a fairly general variable forgetting algorithm is studied. It is proven that the algorithm enjoys the L^2-exponential convergence, [6], and the almost sure exponential convergence, [7], provided that the data are not corrupted by noise. The problem of tracking *time-varying parameters* with data measurement equations subject to stochastic disturbances is considered in [8]. However, the results of [8] are obtained under a stiff persistent excitation condition of deterministic type.

To the best knowledge of the present author, the first paper dealing with truly adaptive identification methods in a fully stochastic framework is [10]. In this paper, Gerencsér studies a general algorithm for the estimation of the parameters of linear continuous-time stochastic models. The proposed identification technique coincides with the classical RLS algorithm with constant forgetting factor in the particular case of linear regressor systems. It is shown that the probability that the tracking error is greater than a given constant δ can be made arbitrarily small provided that the time variability of the parameters is sufficiently slow and the forgetting factor is suitably chosen. Moreover, it is proven that, with probability 1, over a time interval of length T the estimation error remains less than δ in a subinterval of measure T' such that T'/T tends to 1 when T goes to infinity. Though very interesting, these results do not prevent the estimation error from becoming very large (or even unbounded) on some events in the probability space. The occurrence of a large estimation error on events with low probability leads to sudden overshoots in the estimates, a well known phenomenon in engineering applications (blow-up). In order to avoid this undesirable behaviour, one has to bound in some way the ensemble average of the estimation error. Technically speaking, in the linear regressor case this prompts the need of studying the behaviour of the so-called covariance matrix of the algorithm ($P(t)$ in our notations), what is not necessary in order to derive bounds in probability.

More recently, a stochastic analysis of kalman-filter based identification techniques appeared in [12] and [11]. In particular, in [11], a stochastic notion of persistent excitation is adopted and, by a thorough investigation, the boundedness of the tracking error in mean square and almost surely in an average sense is proven. Finite-memory RLS algorithms have been recently considered in [18], where an interesting study of the observation vector properties guaranteeing the invertibility of the so-called regression matrix is provided. These results have a special importance for rectangular-window LS techniques. However, the analysis of [18] is still based on stationarity assumptions on the observation vectors.

The reader is referred to the recent survey [17] for a comprehensive presentation of the most popular methods for the identification of time-varying systems as well as for a description of the main techniques and results in the analysis.

1.3 Assumptions and main results

Denoting by $\sigma(v)$ the σ-algebra generated by a random variable v, define $\mathcal{G}_j = \sigma(\delta\vartheta^\circ(j), \varphi(j), d(j))$ and $\mathcal{G}_o = \sigma(\vartheta^\circ(1))$. Moreover, let $\mathcal{P}_r^s = \sigma(\mathcal{G}_j \mid r \leq j \leq s), r \geq 0$. Note that the structure of equations (1.1) is such that the two sequences $\vartheta^\circ(j)$ and $y(j), j = 1, \cdots, t$, are measurable w.r. to $\mathcal{P}_o^t$. Therefore, $\mathcal{P}_o^t$ represents the σ-algebra of the past up to time t. In the sequel, it will be assumed that $\mathcal{G}_0$ is independent of $\sigma(\mathcal{G}_j \mid j \geq 1)$. We will also assume that the forgetting factor $\mu(t)$ is computed as a function of the system variables up to time t, then $\mu(t)$ is measurable w.r. to $\mathcal{P}_o^t$.

Throughout the paper $\delta\vartheta^\circ(t)$ and $d(t)$ are supposed to be conditionally bounded according to

Assumption A.1:

$$E[\| d(t) \|^2 | \mathcal{P}_0^{t-p}] \leq \Lambda_d^2, \quad p \geq 1 \text{ given constant}, \quad \forall t \geq p,$$

Assumption A.2:

$$E[\| \delta\vartheta^\circ(t) \|^2 | \mathcal{P}_0^{t-1}] \leq \Lambda_\vartheta^2, \quad \forall t \geq 1.$$

Remark 1.3. Note that assumption A.1 does not prevent $d(\cdot)$ from being a coloured noise, possibly with a nonzero expected value. As for A.2, it does not prevent $\vartheta^\circ(\cdot)$ from exhibiting trends or seasonal components.

We also introduce the following assumptions on the content of information in observations (excitation) and on the probabilistic structure of data.

Assumption A.3:

$$E[\varphi(t)\varphi(t)'] \geq aI > 0, \quad \forall t;$$

Assumption A.4:

$$E[(\varphi(t)'\varphi(t))^2] \leq b, \quad \forall t;$$

Assumption A.5:

$\{\delta\vartheta^o(t), \varphi(t), d(t)\}$ is ϕ-mixing with process dependence index $\delta \leq d$.

(see Appendix A for the notion of ϕ-mixing process and process dependence index).

Remark 1.4. Assumption A.5 requires that the correlation of the data is asymptotically vanishing. This corresponds to the natural decay of the influence of the past on the present in dissipative physical systems. Assumption A.3 is an L^2-excitation condition; as such it imposes a constraint on the power of signal $\varphi(\cdot)$ only. Obviously, this condition is much weaker than the widely adopted deterministic-type excitation condition discussed for instance in [8]. See also [3] for more discussion.

The main results of the paper can be summarized as follows.

i) The parameter error keeps bounded whenever the matrix $P(\cdot)$ is not divergent. More precisely, we will prove that the L^1-boundedness of $P(\cdot)$ implies the L^2-boundedness of the tracking error.

ii) The "amplitude" of matrix $P(\cdot)$ can be controlled by a suitable choice of the feasible range for the forgetting factor. The values taken by matrix $P(\cdot)$ depend on the quantity of information carried by data and the value of the forgetting factor. When the information content is "small", matrix $P(\cdot)$ tends to increase. This effect can be counterbalanced by taking larger values of the forgetting factor, so as to enlarge the algorithm memory length. On the other hand, it is well known that this results in a deterioration of the algorithm responsiveness. One of the achievements of this paper is the determination of a *critical value* such that, if the forgetting factor keeps greater than the critical value, then $P(\cdot)$ keeps L^1-bounded.

The paper is organized as follows. In Section 2, a stochastic condition of persistent excitation is introduced. Moreover, it is shown that the observation vectors $\varphi(\cdot)$ of system (1) satisfy this condition. The analysis of algorithm (1.2) is developed in Section 3. Some preliminary remarks are presented in Section 3.1. In Section 3.2, it is shown that $P(\cdot)$ can be kept L^1-bounded provided that the lower bound of the forgetting factor is larger than the *critical value.* The main result of the paper is given in Section 3.3, where it is proven that the L^1-boundedness of $P(\cdot)$ implies the L^2-boundedness of the tracking error. In view of the results of Section 3.2, in Section 3.3 it is concluded that the tracking error can be kept bounded by suitably tuning the lower bound of the forgetting factor.

2 Persistent Excitation

Due to the *monotonicity property* stated in Appendix A, Assumption A.5 entails that

$\{\varphi(t)\}$ is ϕ-mixing with process dependence index $\delta \leq d$.

Theorem 2.1 below investigates the properties of excitation enjoyed by the observation vector sequence $\{\varphi(t)\}$.

Theorem 2.1. *Assume that $\varphi(\cdot)$ is ϕ-mixing and satisfies Assumptions A.3 and A.4. Then, there exist an integer s and a real number $\beta > 0$ such that*

$$E\left[\lambda_{\min}\left\{\sum_{i=t+1}^{t+s}\varphi(i)\varphi(i)'\right\}\right] \geq \beta, \quad \forall t. \tag{2.1}$$

Proof: Using the inequality $\lambda_{\min}(A) \geq \lambda_{\min}(B) - \| A-B \|, A \geq 0, B \geq 0$, one obtains:

$$\lambda_{\min}\left[\sum_{i=t+1}^{t+r}\varphi(i)\varphi(i)'\right] \geq \lambda_{\min}\left[\sum_{i=t+1}^{t+r}E[\varphi(i)\varphi(i)']\right] - \| \Gamma(r,t) \| \tag{2.2}$$

where

$$\Gamma(r,t) = \sum_{i=t+1}^{t+r}[\varphi(i)\varphi(i)' - E[\varphi(i)\varphi(i)']].$$

Divide both sides of inequality (2.2) by r and apply the expectation operator. Proposition A.5 in Appendix A entails that

$$lim_{r\to\infty}\frac{1}{r}E[\| \Gamma(r,t) \|] = 0, \text{ uniformly w.r. to } t. \tag{2.3}$$

On the other hand, in view of Assumption A.3,

$$\frac{1}{r}\lambda_{\min}\left[\sum_{i=t+1}^{t+r}E[\varphi(i)\varphi(i)']\right] \geq a, \quad \forall t. \tag{2.4}$$

The statement of the theorem follows from equations (2.2)-(2.4).

Note that condition (2.1) becomes tighter as β increases and/or s decreases. The minimum integer s for which the statement of Theorem 2.1

holds true is named *order of persistent excitation* of $\varphi(\cdot)$ and will be denoted by r. It is also advisable to introduce the notion of *level of persistent excitation* of sequence $\varphi(\cdot)$ as the real number l defined by

$$l = \inf_t E\left[\lambda_{\min}\left\{\sum_{i=t+1}^{t+r} \varphi(i)\varphi(i)'\right\}\right].$$

3 Analysis of the Tracking Error

Let $\tilde{\vartheta}(t) = \hat{\vartheta}(t) - \vartheta^{\circ}(t+1)$ be the parameter tracking error. From equations (1.1), (1.2.a) and (1.2.d), it follows that

$$\tilde{\vartheta}(t) = F(t)\tilde{\vartheta}(t-1) + K(t)d(t)' - \delta\vartheta^{\circ}(t), \qquad (3.1.a)$$

where

$$F(t) = I - K(t)\varphi(t)'. \qquad (3.1.b)$$

This is a time-varying, nonlinear and stochastic system with $d(\cdot)$ and $\delta\vartheta^{\circ}(\cdot)$ as exogeneous variables. The effect of these variables on the error $\tilde{\vartheta}(\cdot)$ can be controlled by suitably selecting the algorithm gain $K(\cdot)$. Obviously, there is a chance of designing $K(\cdot)$ so that the error $\tilde{\vartheta}(\cdot)$ tends to zero only if $\delta\vartheta^{\circ}(\cdot) = 0$. This corresponds to time-invariant systems. On the contrary, when the parameters are time-varying, one expects that the parameter error will be permanently subject to fluctuations. Hence, the best one can hope for is that these fluctuations keep bounded. In this section, we will work out the conditions under which the tracking error keeps bounded in the L^2-sense.

3.1 Some preliminary remarks

The goal of the present subsection is to discuss in an informal way the connections between the boundedness of the tracking error $\tilde{\vartheta}(\cdot)$ and the boundedness of the algorithm gain $K(\cdot)$ and the covariance matrix $P(\cdot)$.

The disturbance $d(t)$ affects the tracking error $\tilde{\vartheta}(t)$ through the term $K(t)d(t)'$ appearing in equation(3.1.a). If the gain $K(t)$ takes large values, then the disturbance is amplified and determines large fluctuations in the parameter estimate. This fact is known in the engineering literature as "bursting phenomena". On the other hand, the value of the gain $K(t)$ is strictly related to that of the covariance matrix $P(t-1)$. In particular, the boundedness of $P(t-1)$ entails the boundedness of $K(t)$. Indeed, from equations (1.2.b) and (1.2.c) it is readily seen that $\| K(t) \|^2 \leq \| P(t-1) \|$.

The above discussion shows that it is most important to control the value of matrix $P(\cdot)$ in order to prevent the effect of the disturbance from beeing too large. Roughly, the boundedness of $P(\cdot)$ can be thought as a necessary condition in order that the identification algorithm behaves well. In Section 3.3, we will show the important fact that this condition is also sufficient. To be precise, we shall prove that the L^1-boundedness of $P(\cdot)$ implies the L^2-boundedness of the tracking error.

We place emphasis on the fact that requiring the L^1-boundedness of $P(\cdot)$ is an implicit condition. Then, before turning to the main result (L^1-boundedness of $P(\cdot) \Rightarrow L^2$-boundedness of the tracking error) we are well advised to work out explicit conditions which imply the boundedness of matrix $P(\cdot)$. Among other things, this will allow us to point out a feasible range for the forgetting factor. If the forgetting factor belongs to such a range, then the boundedness of the tracking error is guaranteed.

3.2 Boundedness of the covariance matrix

Despite the complexity of recursion (1.2.e) of matrix $P(t)$ (nonlinear and stochastic), it is shown that $E[\| P(t) \|]$ satisfies a simple linear inequality. By imposing that such an inequality is contractive, a nice condition for the uniform boundedness of the L^1-norm of $P(t)$ is worked out.

Theorem 3.1 *Consider equation*(1.2.e) *initialized with* $P(0) = P(0)' > 0$. *If*

$$\exists \text{ integer } h \geq 0 \text{ such that } \mu_0^{(h+1)r} > 1 \; - \; l^2 r^{-2} b^{-1} + d^2 (hr+1)^{-2}, \quad (3.2)$$

(where μ_0 is the lower bound for the forgetting factor (equation(1.3)), r and l are the order and the level of persistent excitation of $\varphi(\cdot)$ (Section 2), and b appears in Assumption A.4), then

$$\sup_t E[\| P(t) \|] < \infty.$$

Proof: Since $\mu(\cdot)$ has to satisfy constraint (1.3) only, we will analyze equation (1.2.e) with $\mu(t)$ replaced by μ_0. Indeed, for a fixed initial condition $P(0)$, the solution of

$$\begin{aligned} P(t) &= \frac{1}{\mu_0}[P(t-1) - a(t)P(t-1)\varphi(t)\varphi(t)'P(t-1)] + Q \\ &= [\mu_0 P(t-1)^{-1} + \mu_0 \varphi(t)\varphi(t)']^{-1} + Q \end{aligned} \quad (3.3)$$

is the maximal solution of equation (1.2.e) with respect to all possible functions $\mu(\cdot)$ satisfying (1.3).

We will now apply Proposition B.2 in Appendix B with $m = 1$ to recursion (3.3). To this end, take as $\mathcal{F}_t$ the σ-algebra generated by $\varphi(t)$ and set $\xi = 1$. Then, all the assumptions of Proposition B.2 are satisfied.

Therefore, for any integer $h \geq 0$ and real numbers $\alpha > 0, \beta > 0, z \in (0,1)$, the following inequality holds:

$$E\left[\mu_0^{(h+1)r} \parallel P(t+(h+1)r) \parallel\right] \leq \nu_1 E[\parallel P(t) \parallel] + H, \qquad (3.4)$$

where the expressions of ν_1 and H are given in the statement of Proposition B.2.

We end the proof by showing that, if condition (3.2) is met with, then there exist $h \geq 0, \alpha > 0, \beta > 0$ and $z \in (0,1)$, such that (3.4) is a contractive inequality, i.e. $\mu_0^{-(h+1)r}\nu_1 < 1$. Indeed, in such a case, tacking into account that the growth of $P(\cdot)$ over the intervals $[j(h+1)r, (j+1)(h+1)r]$, $j = 0, 1, 2, \ldots$, is bounded (see equation (1.2.e)), it is easy to conclude that the L^1-norm of $P(t)$ is uniformly bounded.

From the expression of ν_1 given in Appendix B, the contractivity condition can be written as

$$\mu_0^{-(h+1)r}(1+\alpha^{-1} \parallel Q \parallel)^{(h+1)r}(rb\beta^{-2}+g(l,r^2b,z)+d^2(hr+1)^{-2}) < 1. \quad (3.5)$$

(see again Appendix B, equation (B.10), for the definition of function $g(\cdot,\cdot,\cdot)$).

Then, it is apparent that, if

$$\mu_0^{-(h+1)r}(g(l,r^2b,z)+d^2(hr+1)^{-2}) < 1, \qquad (3.6)$$

inequality (3.5) is satisfied for α and β sufficiently large. Since $g(l,r^2b,z)$ is monotonically decreasing with z and tends to $1 - l^2r^{-2}b^{-1}$ as $z \to 0$ (see expression (B.10)), (3.6) is met with by a sufficiently small z whenever (3.2) holds true.

Remark 3.1. There always exists a $\mu_0 < 1$ such that condition (3.2) is satisfied. Indeed, one can take h large enough so that the right-hand-side of the inequality in (3.2) is lower than 1. This corresponds to the fact that, even if the level of excitation of $\varphi(\cdot)$ is arbitrarily small, one can increase at will the algorithm memory length by taking μ_0 closer and closer to 1. The consequent accumulation of information prevents the L^1-divergence of $P(\cdot)$.

3.3 Boundedness of the tracking error

In this section, it will be shown that the L^1-boundedness of $P(\cdot)$ implies that the tracking error of algorithm (1.2) keeps L^2-bounded.

Since the forgetting factor $\mu(\cdot)$ may depend on $\tilde{\vartheta}(\cdot)$, the updating equation(3.1.a) for the estimation error is nonlinear. However, it is advisable to express its solution in a "linear-like form" as follows:

$$\tilde{\vartheta}(t) = \Phi(t,0)\tilde{\vartheta}(0) + \sum_{\tau=1}^{t} \Phi(t,\tau)[K(\tau)d(\tau)' - \delta\vartheta^{\circ}(\tau)], \tag{3.7}$$

where

$$\Phi(t,\tau) = \begin{cases} I, & \tau = t \\ F(t)F(t-1)\cdots F(\tau+1), & \tau < t, \end{cases}$$

plays the role of transition matrix. The dependence of $F(t)$ on the past values of $\mu(\cdot)$ is rather involved (see equations $(3.1.b), (1.2.b), (1.2.c), (1.2.e)$). However, it is possible to characterize in a simple way the influence of $\mu(\cdot)$ on the quadratic form $F(t)'P(t)^{-1}F(t)$. Indeed, from the quoted equations, one can easily obtain the inequality:

$$F(t)'P(t)^{-1}F(t) \leq \mu(t)P(t-1)^{-1}. \tag{3.8}$$

This observation motivates the forthcoming analysis where $\tilde{\vartheta}(t)$ will be considered in its weighted norm with kernel $P(t)^{-1}$, i.e. $\| \tilde{\vartheta}(t) \|_{P(t)^{-1}} = \max_{\|x\|=1}\{x'\tilde{\vartheta}(t)'P(t)^{-1}\tilde{\vartheta}(t)x\}^{1/2}$.

From equations (3.7) and (3.8) the following fundamental bound for $\| \tilde{\vartheta}(t) \|_{P(t)^{-1}}$ is readily derived

$$\| \tilde{\vartheta}(t) \|_{P(t)^{-1}} \leq \left(\prod_{i=1}^{t} \mu(i)\right)^{1/2} \| \tilde{\vartheta}(0) \|_{P(0)^{-1}}$$

$$+ \sum_{\tau=1}^{t} \left(\prod_{i=1}^{t} \mu(i)\right)^{1/2} \left[\| K(\tau)d(\tau)' \|_{P(\tau)^{-1}} + \| \delta\vartheta^{\circ}(\tau) \|_{P(\tau)^{-1}}\right]. \tag{3.9}$$

This bound will be exploited in the proof of the following fundamental

Theorem 3.2. *Assume that Assumptions A.1 - A.5 hold true. Then,*

$$\sup_t E[\| P(t) \|] < \infty \;\Rightarrow\; \sup_t \| \tilde{\vartheta}(t) \|_{L^2} < \infty. \tag{3.10}$$

Proof: Equation (3.9) can be rewritten as

$$\| \tilde{\vartheta}(t) \| \leq \left(\prod_{i=1}^{t} \mu(i)\right)^{1/2} \| P(t) \|^{1/2} \| \tilde{\vartheta}(0) \|_{P(0)^{-1}}$$

$$+ \sum_{\tau=1}^{t} \left(\prod_{i=\tau+1}^{t} \mu(i)\right)^{1/2} \| P(t) \|^{1/2} \left[\| K(\tau)d(\tau)' \|_{P(\tau)^{-1}} \right.$$

$$\left. + \| \delta\vartheta^{\circ}(\tau) \|_{P(\tau)^{-1}}\right].$$

Consequently, the L^2-norm of $\tilde{\vartheta}(t)$ can be given the following upper bound

$$\| \tilde{\vartheta}(t) \|_{L^2} \leq E^{1/2}\left[\left(\prod_{i=1}^{t}\mu(i)\right) \| P(t) \| \| \tilde{\vartheta}(0) \|^2_{P(0)^{-1}}\right]$$

$$+\sum_{\tau=1}^{t} E^{1/2}\left[\left(\prod_{i=\tau+1}^{t}\mu(i)\right) \| P(t) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}}\right]$$

$$+\sum_{\tau=1}^{t} E^{1/2}\left[\left(\prod_{i=\tau+1}^{t}\mu(i)\right) \| P(t) \| \| \delta\vartheta^{\circ}(\tau) \|^2_{P(\tau)^{-1}}\right]. \tag{3.11}$$

We will now show how to handle a single term, say

$$E^{1/2}[(\Pi_{i=\tau+1}^{t}\mu(i)) \| P(t) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}}],$$

at the right-hand-side of equation(3.11). All the other terms can be handled in the same way.

Consider the function

$$\nu_1(h,\alpha,\beta,z) = (1+\alpha^{-1} \| Q \|)^{(h+1)r}(rb\beta^{-2} + g(l,r^2b,z) + d^2(hr+1)^{-2}),$$

introduced in Proposition B.2 in Appendix B, where h is a nonnegative integer, $\alpha > 0, \beta > 0, z \in (0,1)$ are real numbers and $g(\cdot,\cdot,\cdot)$ is the function defined by equation (B.10) in Appendix B. Fix $(\bar{h},\bar{\alpha},\bar{\beta},\bar{z})$ such that $\nu_1(\bar{h},\bar{\alpha},\bar{\beta},\bar{z}) < 1$ (since $g(l,r^2b,z) < 1$, this is possible). Write

$$t = \tau + m_{t-\tau}(\bar{h}+1)r + s_{t-\tau},$$

where $s_{t-\tau} \in [0,(\bar{h}+1)r-1]$. Then, in view of equation (1.2.e),

$$E^{1/2}\left[\left(\prod_{i=\tau+1}^{t}\mu(i)\right) \| P(t) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}}\right]$$

$$\leq E^{1/2}\left[\left(\prod_{i=\tau+1}^{\tau+m_{t-\tau}(\bar{h}+1)r}\mu(i)\right) \| P(\tau+m_{t-\tau}(\bar{h}+1)r) \| \right.$$

$$\left. \times \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}}\right]$$

$$\leq E^{1/2}\left[\left(\prod_{i=\tau+1}^{\tau+m_{t-\tau}(\bar{h}+1)r}\mu(i)\right) (\bar{h}+1)r \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}}\right], \tag{3.12}$$

where, for $m_{t-\tau}=0$, $\left(\prod_{i=\tau+1}^{\tau}\mu(i)\right)$ has to be interpreted as 1.

We will now apply Proposition B.2 in Appendix B to the first term at the right-hand-side of equation (3.12). To this purpose, define $\mathcal{F}_1 = \mathcal{P}_0^1$ and $\mathcal{F}_i = \mathcal{P}_i^i, i \geq 2$. Since $\vartheta^o(1)$ is assumed to be independent of $\mathcal{P}_1^\infty$, in view of Proposition A.1 in Appendix A, one can conclude that the σ-algebra $\{\mathcal{F}_t \mid t \geq 1\}$ is ϕ-mixing with the same dependence index $\delta \leq d$ of $\{\mathcal{P}_t^t \mid t \geq 1\}$. Moreover, it is straightforward to see that assumtions i) - iv) of Proposition B.2 are satisfied and that $\xi = \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}}$ is measurable w.r. to $\{\mathcal{F}_i \mid i \leq \tau\}$. Then, from Proposition B.2 with $h = \bar{h}, \alpha = \bar{\alpha}, \beta = \bar{\beta}$, and $z = \bar{z}$, we have

$$
\begin{aligned}
E^{1/2} & \left[\left(\prod_{i=\tau+1}^{\tau+m_{t-\tau}(\bar{h}+1)r} \mu(i) \right) \| P(\tau + m_{t-\tau}(\bar{h}+1)r) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \\
& \leq \left\{ \nu^{m_{t-\tau}} E \left[\| P(\tau) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right. \\
& \quad \left. + \, m_{t-\tau}\nu^{m_{t-\tau}-1} H \, E \left[\| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right\}^{1/2} \\
& \leq \left\{ \nu^{m_{t-\tau}} E \left[\| P(\tau) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right\}^{1/2} \\
& \quad + \left\{ m_{t-\tau}\nu^{m_{t-\tau}-1} H \, E \left[\| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right\}^{1/2},
\end{aligned} \tag{3.13}
$$

where $\nu = \max(\nu_1(\bar{h}, \bar{\alpha}, \bar{\beta}, \bar{z}), \nu_2(\bar{h}, \bar{\alpha}, \bar{\beta}, \bar{z}))$ (ν_1 and ν_2 are defined in Proposition B.2). By substituting (3.13) into (3.12), one finally obtains

$$
\begin{aligned}
& E^{1/2} \left[\left(\prod_{i=\tau+1}^{t} \mu(i) \right) \| P(t) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \\
& \leq \nu^{m_{t-\tau}/2} \left\{ E^{1/2} \left[\| P(\tau) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right. \\
& \left. + \left(m_{t-\tau}^{1/2} (H/\nu)^{1/2} + [(\bar{h}+1)r \| Q \|]^{1/2} \right) E^{1/2} \left[\| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right\}.
\end{aligned}
$$

By handling in an analogous way all the terms in equation (3.11), one gets the following bound for $\| \tilde{\vartheta}(t) \|_{L^2}$

$$
\begin{aligned}
\| \tilde{\vartheta}(t) \|_{L^2} \leq \, & \nu^{m_t/2} \left\{ E^{1/2} \left[\| P(0) \| \| \tilde{\vartheta}(0) \|^2_{P(0)^{-1}} \right] \right. \\
& \left. + (m_t^{1/2} c_1 + c_2) E^{1/2} \left[\| \tilde{\vartheta}(0) \|^2_{P(0)^{-1}} \right] \right\} \\
& + \sum_{\tau=1}^{t} \nu^{m_{t-\tau}/2} \left\{ E^{1/2} \left[\| P(\tau) \| \| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right. \\
& \left. + (m_{t-\tau}^{1/2} c_1 + c_2) E^{1/2} \left[\| K(\tau)d(\tau)' \|^2_{P(\tau)^{-1}} \right] \right\}
\end{aligned}
$$

$$+\sum_{\tau=1}^{t} \nu^{m_{t-\tau}/2} \left\{ E^{1/2} \left[\| P(\tau) \| \| \delta\vartheta^{\circ}(\tau) \|_{P(\tau)^{-1}}^{2} \right] \right.$$

$$\left. +(m_{t-\tau}^{1/2} c_1 + c_2) E^{1/2} \left[\| \delta\vartheta^{\circ}(\tau) \|_{P(\tau)^{-1}}^{2} \right] \right\} \tag{3.14}$$

where

$$c_1 = (H/\nu)^{1/2}, \qquad c_2 = [(h+1)r \| Q \|]^{1/2}.$$

Our next step consists in deriving suitable upper bounds for $\| K(\tau)d(\tau)' \|_{P(\tau)^{-1}}^2$ and $\| \delta\vartheta^{\circ}(\tau) \|_{P(\tau)^{-1}}^2$. To this purpose, note first that $P(\tau) \geq Q$, see equation (1.2.e). Consequently,

$$\| \delta\vartheta^{\circ}(\tau) \|_{P(\tau)^{-1}}^{2} \leq \| Q^{-1} \| \| \delta\vartheta^{\circ}(\tau) \|^{2} . \tag{3.15}$$

Turn now to $\| K(\tau)d(\tau)' \|_{P(\tau)^{-1}}^2 = \| d(\tau)K(\tau)P(\tau)^{-1}K(\tau)d(\tau)' \|$, and replace $K(\tau)$ with its expression (1.2.c). Then, one obtains

$$\| K(\tau)d(\tau)' \|_{P(\tau)^{-1}}^{2} \leq \frac{\varphi(\tau)'P(\tau-1)\varphi(\tau)}{1+\varphi(\tau)'P(\tau-1)\varphi(\tau)} \| d(\tau)d(\tau)' \|$$

$$\leq \| d(\tau) \|^{2} . \tag{3.16}$$

Note that the right-hand-sides of inequalities (3.15) and (3.16) are measurable w.r. to the σ-algebra $\mathcal{P}_{\tau}^{\tau}$ associated with time τ only.

We are now in the position to bound term by term the right-hand-side of equation (3.14). Letting $\sup_t E[\| P(t) \|] = b_p$, from (3.16) it turns out that:

$$E^{1/2} \left[\| P(\tau) \| \| K(\tau)d(\tau) \|_{P(\tau)^{-1}}^{2} \right]$$

$$+(m_{t-\tau}^{1/2} c_1 + c_2) E^{1/2} \left[\| K(\tau)d(\tau)' \|_{P(\tau)^{-1}}^{2} \right]$$

$$\leq \Lambda_d \left[\mu_0^{-p/2} (b_p + p \| Q \|)^{1/2} + m_{t-\tau}^{1/2} c_1 + c_2 \right], \quad \forall \tau \geq p. \tag{3.17}$$

Indeed, in view of (3.16) and the inequality $P(t) \leq P(t-1)/\mu(t) + Q$ (see equation (1.2.e)), the following inequality holds:

$$E^{1/2} \left[\| P(\tau) \| \| k(\tau)d(\tau) \|_{P(\tau)^{-1}}^{2} \right]$$

$$\leq E^{1/2} \left[\mu_0^{-p} (\| P(\tau-p) \| + p \| Q \|) \| d(\tau) \|^{2} \right] .$$

Consequently, in view of Assumption A.1,

$$E^{1/2} \left[\| P(\tau) \| \| k(\tau)d(\tau) \|_{P(\tau)^{-1}}^{2} \right]$$

$$\leq E^{1/2} \left[\mu_0^{-p} (\| P(\tau-p) \| + p \| Q \|) \Lambda_d^{2} \right] .$$

This, together with (3.16), leads to (3.17).

Analogously, by resorting to inequality (3.15) and Assumption A.2, one obtains

$$E^{1/2}\left[\| P(\tau) \|\| \delta\vartheta^{\circ}(\tau) \|^2_{P(\tau)^{-1}}\right] + (m_{t-\tau}^{1/2}c_1 + c_2)E^{1/2}\left[\| \delta\vartheta^{\circ}(\tau) \|^2_{P(\tau)^{-1}}\right]$$

$$\leq \Lambda_{\vartheta} \| Q^{-1} \|^{1/2}\left[\mu_0^{-1/2}(b_p + \| Q \|)^{1/2} + m_{t-\tau}^{1/2}c_1 + c_2\right], \quad \forall \tau \geq 1. \quad (3.18)$$

Considering that $m_{t-\tau}$ is linearly increasing whereas $\nu^{m_{t-\tau}}$ is exponentially decreasing with $t-\tau$, from (3.14), (3.17) and (3.18), the thesis follows.

From theorems 3.1 and 3.2, the following conclusion can be drawn:

Theorem 3.3. *Assume that Assumptions A.1 - A.5 hold true and denote by r and l the order and the level of persistent excitation of the observation vector sequence $\varphi(\cdot)$. Then, provided that*

$$\exists \text{ integer } h \geq 0 \text{ such that } \mu_0^{(h+1)r} > 1 \; - \; l^2 r^{-2} b^{-1} + d^2(hr+1)^{-2},$$

one has $\sup_t \| \tilde{\vartheta}(t) \|_{L^2} < \infty$.

Remark 3.2. From Theorem 3.3 and Remark 3.1, it follows that, under Assumtions A.1 - A.5, there always exists a choice of the forgetting factor such that the tracking error keeps bounded.

4 Conclusions

In this paper, the tracking capability of recursive least squares identification algorithms with forgetting factor is analyzed in a fully stochastic framework. The main achievements are:

- The boundedness of matrix $P(\cdot)$ is a sufficient condition for the boundedness of the parameter tracking error.
- A feasible range for the forgetting factor has been worked out. If the forgetting factor belongs to such a range at any time-point, then the boundedness of the parameter tracking error is guaranteed.

This is believed to be the first paper where results of this type are established. Further work is expected in order to derive:

- quantitative bounds for the parameter tracking error, when condition (3.2) is satisfied;
- optimal tuning-rules for the selection of the forgetting factor in the feasible range.

Acknowledgements. The author would like to acknowledge his gratitude to Prof. Sergio Bittanti for useful discussions and for his insightful comments and suggestions. Moreover, the thanks of the author go to an unknown reviewer for bringing to his attention interesting works about state variable identification.

Appendix A: Φ–mixing Stochastic Processes

Given a probability space $(\Omega, \mathcal{F}, p)$, let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two sub σ-algebras of $\mathcal{F}$.

Definition A.1 *(Dependence coefficient between two σ-algebras)*
The dependence coefficient of $\mathcal{F}_1$ with respect to $\mathcal{F}_2$ is defined as

$$\phi(\mathcal{F}_1, \mathcal{F}_2) = \sup_{A \in \mathcal{F}_2} \left\{ \text{ess}\sup_{\omega \in \Omega} |p(A/\mathcal{F}_1) - p(A)| \right\}.$$

Note that $\phi(\mathcal{F}_1, \mathcal{F}_2) \in [0,1], \forall \mathcal{F}_1, \mathcal{F}_2$. Furthermore, given two σ-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ contained in $\mathcal{F}$, such that $\mathcal{G}_1 \supseteq \mathcal{F}_1$ and $\mathcal{G}_2 \supseteq \mathcal{F}_2$, then $\phi(\mathcal{F}_1, \mathcal{F}_2) \leq \phi(\mathcal{G}_1, \mathcal{G}_2)$(*monotonicity property*). Given another σ-algebra $\mathcal{F}_0 \subseteq \mathcal{F}$, consider the σ-algebra $\sigma(\mathcal{F}_0, \mathcal{F}_1)$. From the monotonicity property, it follows that $\phi(\sigma(\mathcal{F}_0, \mathcal{F}_1), \mathcal{F}_2) \geq \phi(\mathcal{F}_1, \mathcal{F}_2)$. However, if $\mathcal{F}_0$ is independent of $\sigma(\mathcal{F}_1, \mathcal{F}_2)$, the following statement holds true.

Proposition A.1. *Suppose that $\mathcal{F}_o$ is independent of $\sigma(\mathcal{F}_1, \mathcal{F}_2)$. Then*

$$\phi(\sigma(\mathcal{F}_0, \mathcal{F}_1), \mathcal{F}_2) = \phi(\mathcal{F}_1, \mathcal{F}_2).$$

This result can be proven by observing that an integrable random variable ξ, measurable w.r. to $\mathcal{F}_2$, is such that $E[\xi \mid \mathcal{F}_1] = E[\xi \mid \sigma(\mathcal{F}_0, \mathcal{F}_1)]$ whenever $\mathcal{F}_0$ is independent of $\sigma(\mathcal{F}_1, \mathcal{F}_2)$.

The following propositions point out the role played by the dependence coefficient in order to extend results classically stated under the independence assumption.

Proposition A.2. *Let A and B be two subsets of Ω belonging to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. Then,*

$$|p(A \cap B) - p(A)p(B)| \ \leq \ \phi(\mathcal{F}_1, \mathcal{F}_2)p(A).$$

Proposition A.3. *Let the random variables ξ_i be $\mathcal{F}_i$ measurable, i=1,2, and $E[|\xi_1|^p] < \infty, E[|\xi_2|^q] < \infty$ with $1/p + 1/q = 1$. Then,*

$$|E[\xi_1 \xi_2] - E[\xi_1]E[\xi_2]| \ \leq \ 2\phi^{1/p}(\mathcal{F}_1, \mathcal{F}_2) E^{1/p}[|\xi_1|^p] \, E^{1/q}[|\xi_2|^q].$$

Proposition A.4. *Let $\xi \geq 0$ be a random variable measurable w.r. to the σ-algebra $\mathcal{F}_1$ and denote by B an element of the σ-algebra $\mathcal{F}_2$. Then*

$$\int_B \xi dp \;\leq\; (p(B) + \phi(\mathcal{F}_1, \mathcal{F}_2))E[\xi].$$

Propositions A.2 and A.3 are proven in [13]; as for Proposition A.4, its statement is an obvious consequence of Proposition A.2 when ξ is a simple function. Then, the general statement can be obtained by a limiting process.

From Definition A.1 and Proposition A.2, it follows that $\phi(\mathcal{F}_1, \mathcal{F}_2) = 0$ iff $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent.

Turn now to consider a sequence of σ-algebras $\{\mathcal{F}_t \subseteq \mathcal{F} \mid t \geq 1\}$ and denote by $\mathcal{F}_r^s$ the σ-algebra generated by $\mathcal{F}_r, \cdots, \mathcal{F}_s$. We also let $\mathcal{F}_r^\infty$ be the σ-algebra generated by $\{\mathcal{F}_j, j \geq r\}$.

Definition A.2. (*Dependence function of* a *σ-algebra sequence)*
The dependence function $\rho(\cdot)$ of a σ-algebra sequence $\{\mathcal{F}_t \mid t \geq 1\}$ is defined by

$$\rho(m) = \sup_i \phi(\mathcal{F}_1^i, \mathcal{F}_{i+m}^\infty), \quad m \geq 0.$$

The value $\rho(m)$ can be roughly interpreted as a correlation coefficient between the past and the m-steps forward future. If $\{\mathcal{F}_t \mid t \geq 1\}$ is an independent sequence, then $\rho(m) = 0, \forall m \geq 1$. In general, in view of the *monotonicity property*, $\rho(\cdot)$ is monotonically decreasing.

Definition A.3. (*ϕ-mixing sequence of σ-algebras and dependence index of* a *σ-algebra sequence)*
The sequence $\{\mathcal{F}_t \mid t \geq 1\}$ is said to be ϕ-mixing if $\sum_{m=1}^\infty \rho(m)^{1/2} < \infty$. The quantity $\delta = \sum_{m=1}^\infty \rho(m)^{1/2}$ is then said to be the dependence index of the σ -algebra sequence.

The notion of ϕ-mixing sequence of σ-algebras can be applied when $\{\mathcal{F}_t \mid t \geq 1\}$ is generated by a stochastic process $\{\xi(t) \mid t \geq 1\}$. In this case, the dependence function and the dependence index of the σ-algebra sequence are also called *process dependence function* and *process dependence index*, respectively.

Many results valid for independent processes still hold true under the ϕ-mixing condition. This is the case of the L^2-law of large numbers as indicated in

Proposition A.5. *Consider a sequence of vector random variables $\{\xi(t)\}$ such that*

i) *There exists b such that $E[(\xi(t)'\xi(t))^2] \leq b, \quad \forall t$;*

ii)$\{\xi(t)\}$ *is ϕ-mixing.*
Then,

$$\lim_{r\to\infty}\frac{1}{r^2}E\left[\|\sum_{i=t+1}^{t+r}\{\xi(i)\xi(i)'-E[\xi(i)\xi(i)']\}\|^2\right] \;=\; 0,$$

uniformly w.r to t.

Proof: Consider the j,k-element of matrix $\xi(i)\xi(i)'-E[\xi(i)\xi(i)']$:

$$\zeta_{j,k}(i)=\xi_j(i)\xi_k(i)-E[\xi_j(i)\xi_k(i)],$$

where $\xi_l(i)$ is the l-th component of $\xi(i)$.

Since

$$\begin{aligned} E[(\xi(i)'\xi(i))^2] &\geq E[(\xi_j(i)\xi_k(i))^2] \\ &\geq E[\zeta_{j,k}(i)^2], \quad \forall i, \forall j,k, \end{aligned}$$

from condition i), it follows that:

$$E[\zeta_{j,k}(i)^2]\leq b, \quad \forall i, \quad \forall j,k. \tag{A.1}$$

Consider now two time points τ and τ'; by applying Proposition A.3 to $\zeta_{j,k}(\tau)$ and $\zeta_{j,k}(\tau')$, one obtains:

$$E[\zeta_{j,k}(\tau)\zeta_{j,k}(\tau')]\leq 2\rho(|\,\tau-\tau'\,|)^{1/2}E^{1/2}[\zeta_{j,k}(\tau)^2]E^{1/2}[\zeta_{j,k}(\tau')^2], \qquad \forall j,k, \tag{A.2}$$

where $\rho(\cdot)$ is the process dependence function of $\xi(\cdot)$.

Letting δ be the dependence index of process $\xi(\cdot)$, from (A.1) and (A.2), it follows that:

$$\begin{aligned} \frac{1}{r^2}E\left[\left(\sum_{i=t+1}^{t+r}\zeta_{j,k}(i)\right)^2\right] &= \frac{1}{r^2}\sum_{\tau,\tau'=t+1}^{t+r}E[\zeta_{j,k}(\tau)\zeta_{j,k}(\tau')] \\ &\leq \frac{2b}{r^2}\sum_{\tau,\tau'=t+1}^{t+r}\rho(|\,\tau-\tau'\,|)^{1/2} \\ &\leq 4b(1+\delta)/r, \qquad \forall j,k \end{aligned} \tag{A.3}$$

On the other hand, denoting by n the dimension of vector $\xi(\cdot)$,

$$\frac{1}{r^2}E\left[\|\sum_{i=t+1}^{t+r}\{\xi(i)\xi(i)'-E[\xi(i)\xi(i)']\}\|^2\right]$$

$$\leq \frac{1}{r^2} E \left[\left\{ \sum_{j,k=1}^{n} \mid \sum_{i=t+1}^{t+r} \zeta_{j,k}(i) \mid \right\}^2 \right]$$

$$\leq n^4 \max_{j,k} \frac{1}{r^2} E \left[\left\{ \sum_{i=t+1}^{t+r} \zeta_{j,k}(i) \right\}^2 \right], \quad \forall t.$$

Hence, from (A.3) the thesis follows.

Remark A.1. From the above proof, one can conclude that the rate of convergence to zero of $\frac{1}{r^2}E[\| \sum_{i=t+1}^{t+r} \{\xi(i)\xi(i)' - E[\xi(i)\xi(i)']\} \|^2]$ depends on the dimension n of vector $\xi(i)$, the dependence index δ of process $\xi(\cdot)$ and parameter b only.

For further discussion on the ϕ-mixing notion, see [13].

Appendix B: Some Technical Results on the Covariance Matrix Recursion

Consider the recursive equation

$$P(t) = \frac{1}{\mu(t)}[P(t-1) - a(t)P(t-1)\varphi(t)\varphi(t)'P(t-1)] + Q. \qquad (B.1.a)$$

$$a(t) = (1 + \varphi(t)'P(t-1)\varphi(t))^{-1} \qquad (B.1.b)$$

with initial condition

$$P(0) = P(0)' > 0 \qquad (B.2)$$

and $\mu(\cdot)$ and Q such that

$$0 < \mu_0 \leq \mu(t) \leq \mu_1 < 1, \qquad (B.3.a)$$

$$Q > 0. \qquad (B.3.b)$$

In this appendix, some important results concerning $\| P(t) \|$ are derived. The first one follows from the structure of the recursive equation and does not require any assumption on the exogenous sequences $\mu(\cdot)$ and $\varphi(\cdot)$.

Proposition B.1. *Consider three time points t_1, t_2, and t_3 such that $1 \leq t_1 \leq t_2 < t_3$. Given any pair of positive real numbers $\alpha > 0$ and $\beta > 0$, we have*

$$\lambda_{\min}[P(t_3)^{-1}]$$

$$\geq \left(\prod_{i=t_1+1}^{t_3}\right) \min\left\{f_1(\alpha, t_3-t_1),\ \max\left\{f_2(\alpha,t_3-t_1)\lambda_{\min}[P(t_1)^{-1}],\right.\right.$$

$$\left.\left. f_3(\beta, t_3-t_2) I_{A(\beta,t_2,t_3)}\lambda_{min}\left[\sum_{t_2+1}^{t_3}\varphi(i)\varphi(i)'\right]\right\}\right\}$$

where

$$f_1(\alpha, t_3-t_1) = (\alpha + (t_3-t_1) \parallel Q \parallel)^{-1}$$

$$f_2(\alpha, t_3-t_1) = (1+\alpha^{-1} \parallel Q \parallel)^{-(t_3-t_1)}$$

$$f_3(\beta, t_3-t_2) = (1+ \parallel Q \parallel \parallel Q^{-1}+\beta I \parallel)^{-(t_3-t_2)}$$

$$A(\beta, t_2, t_3) = \bigcap_{i=t_2+1}^{t_3} \{\parallel \varphi(i) \parallel^2 \leq \beta\},$$

and I_A denotes the indicator of set A.

Proof: From (B.1) and (B.2), it follows that $P(t) > Q > 0, \forall$t. Therefore $P(t)$ is nonsingular and recursion (B.1a) can also be written as

$$P(t) = [\mu(t)P(t-1)^{-1} + \mu(t)\varphi(t)\varphi(t)']^{-1} + Q, \qquad (B.4)$$

which entails the following inequality

$$\parallel P(t) \parallel \leq \mu(t)^{-1} \parallel P(t-1) \parallel + \parallel Q \parallel, \quad \forall t. \qquad (B.5)$$

It is advisable to distinguish two complementary events:

$$B = \{\parallel P(t) \parallel > \alpha, \quad \forall t \in [t_1, t_3-1]\},$$

$$\bar{B} = \{\exists \bar{t} \in [t_1, t_3-1] : \quad \parallel P(\bar{t}) \parallel \leq \alpha\}$$

The following inequalities hold:

$$\lambda_{\min}[P(t_3)^{-1}] \geq \left(\prod_{i=t_1+1}^{t_3}\right)\left(1+\alpha^{-1} \parallel Q \parallel\right)^{t_1-t_3} \lambda_{\min}[P(t_1)^{-1}], \quad on B, \qquad (B.6)$$

$$\lambda_{\min}[P(t_3)^{-1}] \geq \left(\prod_{i=t_1+1}^{t_3}\right)(\alpha + (t_3-t_1) \parallel Q \parallel)^{-1}, \quad on \bar{B}, \qquad (B.7)$$

Inequality (B.7) is a straightforward consequence of inequality (B.5). Inequality (B.6) follows from the inequality

$$\parallel P(t) \parallel \leq \mu(t)^{-1} \parallel P(t-1) \parallel + \alpha^{-1} \parallel Q \parallel \parallel P(t-1) \parallel, \quad \forall t \in [t_1+1, t_3],$$

which can also be derived from (B.5).

Moreover, we have

$$\lambda_{\min}[P(t_3)^{-1}]$$

$$\geq \left(\prod_{i=t_1+1}^{t_3}\right) \left(1+ \| Q \| \| Q^{-1} + \beta I \|\right)^{t_2-t_3} \lambda_{min}\left[\sum_{i=t_2+1}^{t_3} \varphi(i)\varphi(i)'\right],$$

$$on A(\beta, t_2, t_3). \qquad (B.8)$$

Indeed, since $P(t) > Q$, the inequality $\mu(t)P(t-1)^{-1} + \mu(t)\varphi(t)\varphi(t)' \leq Q^{-1} + \beta I, \forall t \in [t_2+1, t_3]$ holds true on $A(\beta, t_2, t_3)$. Then, (B.4) leads to the inequality

$$P(t) \leq (1+ \| Q \| \| Q^{-1} + \beta I \|) \left[\mu(t)P(t-1)^{-1} + \mu(t)\varphi(t)\varphi(t)'\right]^{-1},$$

$$\forall t \in [t_2+1, t_3] \qquad (B.9)$$

Inequality (B.8) is a straightforward consequence of (B.9).

The thesis of the Proposition follows from (B.6) - (B.8).

The second result, stated below as Proposition B.2, deals with the time-evolution of $P(\cdot)$ when the exogenous variables $\varphi(\cdot)$ and $\mu(\cdot)$ are described as particular stochastic processes. Preliminarily, we provide two technical lemmas.

Lemma B.1. *Consider two random variables $\psi \geq 0$ and $\zeta > 0$ measurable w.r. to the σ-algebra $\mathcal{G}$ and a random variable $\eta \geq 0$ measurable w.r. to the σ-algebra $\mathcal{H}$. Given a real number $c > 0$, let $A = \{\eta > c\}$. Then*

$$E\left[\frac{\psi}{\max\{\zeta, \eta\}}\right] \leq E\left[\frac{\psi}{c}\right] + (p\bar{A} + \phi(\mathcal{G}, \mathcal{H}))E\left[\frac{\psi}{\zeta}\right],$$

where $\phi(\mathcal{G}, \mathcal{H})$ is the dependence coefficient of the two σ-algebras.

Proof: Since

$$E\left[\frac{\psi}{\max\{\zeta, \eta\}}\right] \leq \int_A \frac{\psi}{c} dp + \int_{\bar{A}} \frac{\psi}{\zeta} dp,$$

the thesis follows from Proposition A.4.

Lemma B.2. *Consider a random variable $\eta \geq 0$ such that*

i) $E[\eta] \geq \gamma > 0$,

ii) $E[\eta^2] \leq \sigma < \infty$.

Given any real $z \in (0,1)$, the probability of $A_z = \{\omega : \eta \leq z\gamma\}$ is bounded from above by the quantity

$$g(\gamma, \sigma, z) = (2\gamma^2 z^2)^{-1}\{2z\gamma^2 - \sigma + [4\gamma^2\sigma z(z-1) + \sigma^2]^{1/2}\} < 1. \quad (B.10)$$

Proof: Observe that

$$\int_{\bar{A}_z} \eta^2 dp \ \geq \ \frac{1}{p(\bar{A}_z)}\left\{\int_{\bar{A}_z} \eta dp\right\}^2. \quad (B.11)$$

Moreover, in view of assumption i),

$$\gamma \leq z\gamma p(A_z) + \int_{\bar{A}_z} \eta dp. \quad (B.12)$$

Then, from ii), (B.11) and (B.12)

$$\sigma \geq E[\eta^2] \geq \int_{\bar{A}_z} \eta^2 dp \geq \frac{1}{p(\bar{A}_z)}\left\{\int_{\bar{A}_z} \eta dp\right\}^2 \geq \frac{1}{p(\bar{A}_z)}(1 - zp(A_z))^2\gamma^2.$$

The statement easily follows from this inequality.

Proposition B.2. *Let $\{\mathcal{F}_t \mid t \geq 1\}$ be a ϕ-mixing sequence of σ-algebras with dependence index $\delta \leq d$ and assume that*

i) *$\mu(t)$ is measurable w.r. to $\sigma(\mathcal{F}_i \mid i \leq t)$,* $\forall t$

ii) *$\varphi(t)$ is measurable w.r. to $\mathcal{F}_t$,* $\forall t$

iii) $E[\varphi(t)\varphi(t)'] \geq aI > 0, \quad \forall t$

iv) $E[(\varphi(t)'\varphi(t))^2] \leq b < \infty, \quad \forall t.$

Given an integer $h \geq 0$, three real numbers $\alpha > 0, \beta > 0$, and $z \in (0,1)$ and a random variable $\xi \geq 0$ measurable w.r. to $\sigma(\mathcal{F}_i \mid i \leq \bar{t})$, the following inequality holds true

$$E\left[\left(\prod_{i=\bar{t}+1}^{\bar{t}+m(h+1)r} \mu(i)\right) \xi \parallel P(\bar{t} + m(h+1)r) \parallel\right]$$

$$\leq \nu_1^m E[\xi \parallel P(\bar{t}) \parallel] + \sum_{i=1}^{m} \nu_1^{m-i}\nu_2^{i-1} H E[\xi], \quad \forall m \geq 1,$$

with

$$\nu_1 = (1 + \alpha^{-1} \parallel Q \parallel)^{(h+1)r}(rb\beta^{-2} + g(l, r^2 b, z) + d^2(hr+1)^{-2}),$$

$$\nu_2 = \mu_1^{(h+1)r},$$

$$H = (1+ \| Q \| \| Q^{-1} + \beta I \|)^r z^{-1} l^{-1} + (\alpha + (h+1)r \| Q \|),$$

where r and l are the order and the level of persistent excitation of $\varphi(\cdot)$ and $g(\cdot,\cdot,\cdot)$ is defined in (B.10). (See the discussion in Section 2 for the notion of order and level of persistent excitation of process $\varphi(\cdot)$).

Proof: Let

$$\tau(n) = \bar{t} + n(h+1)r,$$

$$A(n) = \bigcap_{i=\tau(n-1)+hr+1}^{\tau(n)} \{\| \varphi(i) \|^2 \leq \beta\},$$

$$\Phi(n) = \sum_{i=\tau(n-1)+hr+1}^{\tau(n)} \varphi(i)\varphi(i)',$$

$$M(n) = \begin{cases} \prod_{i=\tau(0)+1}^{\tau(n)} \mu(i), n \geq 1 \\ 1, n = 0. \end{cases}$$

Resorting to Proposition B.1 with $t_1 = \tau(m-1), t_2 = \tau(m-1) + hr, t_3 = \tau(m)$, after simple calculations one obtains

$$E[M(m)\xi \| P(\tau(m)) \|]$$

$$\leq \; E[M(m-1)\xi f_1(\alpha,(h+1)r)^{-1}] + E[M(m-1)\xi \chi(m)^{-1}], \qquad (B.13)$$

where

$$\chi(m)$$

$$= \max\left\{ f_2(\alpha,(h+1)r)\lambda_{\min}[P(\tau(m-1))^{-1}], \quad I_{A(m)} f_3(\beta,r)\lambda_{\min}[\Phi(m)] \right\}.$$

In order to derive an upper bound for the second term at the right-hand-side of (B.13), apply Lemma B.1 with

$$\mathcal{G} = \{\mathcal{F}_i \mid i \leq \tau(m-1)\},$$

$$\mathcal{H} = \{\mathcal{F}_i \mid i \geq \tau(m-1) + hr + 1\},$$

$$\psi = M(m-1)\xi,$$

$$\zeta = f_2(\alpha,(h+1)r)\lambda_{\min}[P(\tau(m-1))^{-1}],$$

$$\eta = I_{A(m)} f_3(\beta,r)\lambda_{\min}[\Phi(m)],$$

$$c = zlf_3(\beta,r).$$

It turns out that:

$$E[M(m-1)\xi\chi(m)^{-1}] \leq E[M(m-1)\xi(zlf_3(\beta,r))^{-1}]$$

$$+ \left(p\{I_{A(m)}\lambda_{\min}[\Phi(m)] \leq zl\} + \rho(hr+1)\right)$$

$$\times E\left[M(m-1)\xi f_2(\alpha,(h+1)r)^{-1} \parallel P(\tau(m-1)) \parallel\right], \qquad (B.14)$$

where $\rho(\cdot)$ is the dependence function of $\{\mathcal{F}_t \mid t \geq 1\}$.

The probability $p\{I_{A(m)}\lambda_{\min}[\Phi(m)] \leq zl\}$ can also be bounded from above. With this objective in mind, notice first that

$$E\left[(\lambda_{\min}[\Phi(m)])^2\right] \leq E\left[\left(\sum_{i=\tau(m-1)+hr+1}^{\tau(m)} \parallel \varphi(i) \parallel^2\right)^2\right]$$

$$\leq r^2 \left(\max_{i\in[\tau(m-1)+hr+1,\tau(m)]} E[\parallel \varphi(i) \parallel^4]\right).$$

Hence, from assumption iv),

$$E\left[\lambda_{\min}[\Phi(m)])^2\right] \leq r^2 b. \qquad (B.15)$$

In view of assumptions ii) - iv), $\varphi(\cdot)$ satisfies equation (2.1) with $s = r$ and $\beta = l$ (see Theorem 2.1 and the discussion which follows this theorem). Consequently,

$$E(\lambda_{\min}[\Phi(m)]) \;\geq\; l. \qquad (B.16)$$

In view of Lemma B.2, (B.15) and (B.16) entail that:

$$p\{\lambda_{\min}[\Phi(m)] \leq zl\} \;\leq\; g(l, r^2 b, z). \qquad (B.17)$$

Consider now the complement of A(m):

$$\bar{A}(m) = \bigcup_{i=\tau(m-1)+hr+1}^{\tau(m)} \{\parallel \varphi(i) \parallel^2 > \beta\}.$$

Thanks to assumption iv), the Chebyshev inequality entails that $p\{\parallel \varphi(i) \parallel^2 > \beta\} \leq b\beta^{-2}$. Therefore:

$$p(\bar{A}(m)) \;\leq\; rb\beta^{-2}. \qquad (B.18)$$

Finally, inequalities (B.17) and (B.18) can be used to derive the required upper bound

$$p\{I_{A(m)}\lambda_{\min}[\Phi(m)] \leq zl\} \;\leq\; rb\beta^{-2} + g(l, r^2 b, z). \qquad (B.19)$$

Substituting (B.19) into (B.14), and further substituting into (B.13), one obtains

$$E[M(m)\xi \parallel P(\tau(m)) \parallel]$$

$$\leq\; \left((zlf_3(\beta, r))^{-1} + f_1(\alpha,(h+1)r)^{-1}\right) E[M(m-1)\xi]$$

$$+ \ \big(rb\beta^{-2} + g(l, r^2b, z) + \rho(hr+1)\big)\, f_2(\alpha, (h+1)r)^{-1}$$
$$\times E\left[M(m-1)\xi \parallel P(\tau(m-1)) \parallel\right]. \qquad (B.20)$$

Recall now that the dependence function $\rho(\cdot)$ enjoys the fundamental monotonicity property. Therefore, being the dependence index of $\{\mathcal{F}_t \mid t \geq 1\}$ bounded from above by d, $\rho(hr+1) \leq d^2(hr+1)^{-2}$. In conclusion, from (B.20) and the definitions of ν_1, ν_2 and H given in the statement of the proposition, one obtains

$$E[M(m)\xi \parallel P(\tau(m)) \parallel] \leq \nu_1 E\left[M(m-1)\xi \parallel P(\tau(m-1)) \parallel\right] + H\nu_2^{m-1}E[\xi].$$

From this recursive inequality the thesis is easily obtained.

References

[1] B.D.O. Anderson and R.M. Johnstone. Adaptive systems and time varying plants, *Int. J. Control* **37** (1983), 367-377.

[2] A. Benveniste. Introduction a la methode de l'equation differentielle moyenne pour l'etude des algorithmes recursifs—exemples, in *Outils et Models Mathematiques pour L'automatique, L'analyse de Systemes et le Traitement du signal.* Editions du CNRS, 1981.

[3] R.R. Bitmead and B.D.O. Anderson. Performance of adaptive estimation algorithms in dependent random environments, *IEEE Trans. Autom. Control* **25** (1980), 788-794.

[4] S. Bittanti, P. Bolzern, and M. Campi. Adaptive identification via prediction-error directional-forgetting factor: Convergence analysis, *Int. J. Control* **50** (1989), 2407-2421.

[5] S. Bittanti, P. Bolzern, and M. Campi. Convergence and exponential convergence of identification algorithms with directional forgetting factor, *Automatica* **26(5)** (1990), 929-932.

[6] S. Bittanti and M. Campi. Adaptive RLS algorithms under stochastic excitation—L^2 convergence analysis, *IEEE Trans. Autom. Control* **36(8)** (1991), 963-967.

[7] S. Bittanti and M. Campi. Adaptive RLS algorithms under stochastic excitation—strong consistency analysis, *Systems & Control Letters* **17** (1991), 3-8.

[8] R.M. Canetti and M.D. Espana. Convergence analysis of the least-squares identification algorithm with a variable forgetting factor for time-varying linear systems, *Automatica* **4** (1989), 609-612.

[9] J.J. Fuchs. Recursive least-squares algorithms revisited, *IEE Proc.* **128(2)** (1981), 74-76.

[10] L. Gerencser. Parameter tracking of time-varying continuous-time linear stochastic systems, in *Modelling, Identification and Robust Control* (C.I. Byrnes and A. Lindquist, eds.) Elsevier Science Publishers B.V., North Holland, 1986.

[11] L. Guo. Estimating time-varying parameters by the Kalman filter based algorithm: Stability and convergence, *IEEE Trans. Autom. Control* **35** (1990), 141-147.

[12] L. Guo, L. Xia and J.B. Moore. Tracking randomly varying parameters, *Proc. 27th IEEE Conference on Decision and Control* Austin, 1988, 1514-1519.

[13] M. Iosifescu and R. Theodorescu. *Random Processes and Learning.* New York: Springer Verlag, 1969.

[14] R.M. Johnstone, C.R. Johnson, R.R. Bitmead, and B.D.O. Anderson. Exponential convergence of recursive least squares with exponential forgetting factor, *Systems & Control Letters* **2** (1982), 77-82.

[15] L. Ljung. Analysis of stochastic gradient algorithms for linear regression problems, *IEEE Trans. Inform. Theory* **30** (1984), 151-160.

[16] L. Ljung. Analysis of recursive stochastic algorithms, *IEEE Trans. Autom. Control* **22** (1977), 551-575.

[17] L. Ljung and S. Gunnarson. Adaptation and tracking in system identification - A survey, *Automatica* **26** (1990), 7-21.

[18] M. Niedzwiecki and L. Guo. Nonasymptotic results for finite-memory WLS filters, *IEEE Trans. Autom. Control* **36(2)** (1991), 198-206.

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

Dipartimento di Elettronica per l'Automazione, Universita' di Brescia, via Branze, 25123 Brescia, Italy.

Communicated by Anders Lindquist