Recursive Least-Squares Identification Algorithms with Incomplete Excitation: Convergence Analysis and Application to Adaptive Control

SERGIO BITTANTI, PAOLO BOLZERN, AND MARCO CAMPI

Abstract—**In this note, the convergence properties of a fairly general class of adaptive recursive least-squares algorithms are studied under the assumption that the data generation mechanism is deterministic and time invariant. First, the (open-loop) identification case is considered. By a suitable notion of excitation subspace, the convergence analysis of the identification algorithm is carried out with no persistent excitation hypothesis. Precisely, it is proven that the projection of the parameter error on the excitation subspace tends to zero, while the orthogonal component of the error remains bounded. Then, the convergence of an adaptive control scheme based on the minimum variance control law is dealt with. By suitably exploiting the previously-mentioned convergence result relative to the identification case, it can be shown that, under the standard minimum-phase assumption, the tracking error tends to zero and the control variable is bounded.**

I. Introduction

Many different variants of the standard recursive least-squares (RLS) estimation algorithm have been proposed in recent years to achieve a certain degree of adaptivity. A popular technique consists of incorporating a forgetting factor in order to discount obsolete information in favor of the fresh one carried on by the most recent data. In this note, a fairly general family of RLS estimators with variable forgetting (VF) factor is analyzed. It includes the standard RLS (without forgetting factor), the exponential forgetting (EF), the prediction error forgetting (PEF), and the constant trace (CT) algorithms.

Usually, the convergence properties of identification algorithms are investigated under the assumption of persistent excitation (see, e.g., [1]–[3]). On the contrary, in this note, we consider the case when the excitation is incomplete, namely, the data do not yield sufficient information along all the directions in the parameter space. It will be shown that, under the only assumption that the "true system" is deterministic and time-invariant, the VF algorithm exhibits nice properties, whatever the information content of the data may be. More precisely, we will introduce the notion of excitation subspace as the set of directions in the parameter space along which the excitation content produced by data becomes larger and larger as time increases. It turns out that the projection of the parameter error on the excitation subspace tends to zero, while the orthogonal component of the error remains bounded.

As is well known, the situation of incomplete excitation is typical of adaptive control schemes. This is why the previously-mentioned result turns out to be very useful in the convergence analysis of adaptive controllers based on the VF identification algorithm. Along this line, we will precisely consider a minimum variance control scheme, and we will show that, under the standard minimum-phase assumption, the tracking error converges to zero whenever the reference signal is bounded. Furthermore, the control variable turns out to be bounded.

The approach provided herein looks quite general and powerful. In particular, it allows us to analyze algorithms whose covariance matrix is not necessarily bounded. Reportedly, the analysis of such a case is impervious, and the classical results of [4] are of no use, since they require an *ad hoc* constraint on the covariance matrix (see [5] and [6]).

This note is organized as follows. After some preliminaries concerning the specific algorithm to be considered (Section II), the incomplete excitation notion is introduced in Section III. In the same section, the convergence analysis of the identification algorithm is provided. Finally, in Section IV the minimum variance adaptive control scheme based on the VF identification algorithm is studied. Some final conclusions are drawn in Section V.

Manuscript received August 11, 1989; revised January 5, 1990. This work was supported by the Centro di Teoria dei Sistemi CNR and MURST.

The authors are with the Dipartimento di Elettronica del Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.

IEEE Log Number 9039309.

II. Preliminaries

Consider two discrete-time sequences $u(\cdot)$ and $y(\cdot)$ and the associated linear predictor

$$\hat{y}(t) = \varphi(t)'\vartheta(t)$$

where $\varphi(t)$ is the n-dimensional observation vector containing past values of $u(\cdot)$ and $y(\cdot)$. The parameter vector $\vartheta(t)$ can be estimated by means of the following recursive algorithm with variable forgetting:

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

$$r(t) = \varphi(t)'P(t-1)\varphi(t) \tag{1b}$$

$$K(t) = \frac{P(t-1)\varphi(t)}{1+r(t)} \tag{1c}$$

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + K(t)\epsilon(t)$$

$$P(t) = \frac{1}{\mu(t)}\left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)'P(t-1)}{1+r(t)}\right].$$

The only constraint on the forgetting factor $\mu(t)$ is that $\mu(t) \in (0, 1]$, $\forall t$. Therefore, these equations encompass many well-known algorithms. In particular

$$\mu(t) = 1 \Rightarrow \text{RLS}$$

$$\mu(t) = \bar{\mu},\ \bar{\mu} \in (0,1) \Rightarrow \text{EF}$$

$$\mu(t) = \max\left\{\mu_0, 1 - \frac{\epsilon(t)^2}{(1+r(t))\Gamma}\right\}, \qquad \mu_0 \in (0,1) \Rightarrow \text{PEF}$$

$$\mu(t) = 1 - \frac{1}{\Delta}\frac{\varphi(t)'P(t-1)P(t-1)\varphi(t)}{1+r(t)} \Rightarrow \text{CT}.$$

In PEF, Γ represents the information content of the algorithm, while constant Δ in CT is the value imposed to the trace of $P(t)$.

Actually, in order to derive results on the identification algorithms suited to be applicable to adaptive control as well, it is advisable to consider the following slight generalization of the recursive equations for $\hat{\vartheta}(\cdot)$ and $P(\cdot)$:

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

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

The scalar $a(\cdot)$ is a time-varying function to be selected in a class which will be specified in light of the recursive equation of $P(t)^{-1}$

$$P(t)^{-1} = \mu(t)P(t-1)^{-1} + \beta(t)\mu(t)\varphi(t)\varphi(t)', \tag{2a}$$

$$\beta(t) = \frac{a(t)}{1+r(t)-a(t)r(t)}. \tag{2b}$$

Equation (2a) can be interpreted as a discrete-time system with the entries of $P(t)^{-1}$ as state variables and $\beta(t)\mu(t)\varphi(t)\varphi(t)'$ as exogenous input. In the identification context, $a(t) = 1$, $\forall t$, so that $\beta(t) = 1$, $\forall t$; see (2b). Thus, if $\mu(t) \le \bar{\mu} < 1$, $\forall t$, $P^{-1} = 0$ is the globally attractive equilibrium point for the free motion of system (2a). Roughly, the input term $\mu(t)\varphi(t)\varphi(t)'$ tends then to prevent the convergence of $P(t)^{-1}$ to such an equilibrium point.

As is apparent from (2a), this qualitative behavior is retained in the case when $a(\cdot)$ is time-varying provided that $\beta(t)$ is bounded from below, i.e.,

A1: There exists $c > 0$ such that $a(t)/(1 + r(t) - a(t)r(t)) \geq c$, $\forall t$.

Note that the usual choice of $a(\cdot)$ in the identification context, i.e., $a(t) = 1$, $\forall t$, obviously satisfies this inequality.

In the sequel, the theoretical analysis of algorithm (1) will be performed under Assumption A1, so obtaining results suitable for the analysis of adaptive control schemes too.

III. Identification

Throughout this note, we will conform to the usual assumption that the data generation mechanism is described by

A2: $y(t) = \varphi(t)'\vartheta^o$.

However, the analysis will be carried out without taking any persistent excitation assumption on $\varphi(\cdot)$ for granted. With this objective in mind, the following definition is in order.

Definition: The set

$$\bar{\mathcal{E}} = \left\{ x \in \mathbb{R}^n \mid \exists L < \infty: x' \sum_{1}^{N}{}_{t} \varphi(t)\varphi(t)'x < L, \forall N \right\}$$

is named *unexcitation subspace*.

Its orthogonal complement $\mathcal{E} = \bar{\mathcal{E}}^{\perp}$ is called *excitation subspace*. □

The proof that $\bar{\mathcal{E}}$ is actually a subspace is trivial. It is also useful to introduce the Lyapunov-like function

$$V(t) = \tilde{\vartheta}(t)'P(t)^{-1}\tilde{\vartheta}(t) \tag{3}$$

where $\tilde{\vartheta}(t)$ is the parameter error, defined as

$$\tilde{\vartheta}(t) = \hat{\vartheta}(t) - \vartheta^o. \tag{4}$$

In the sequel, $\tilde{\vartheta}(t)$ will be decomposed into its projections on $\mathcal{E}$ and $\bar{\mathcal{E}}$, denoted by $\tilde{\vartheta}_E(t)$ and $\tilde{\vartheta}_U(t)$, respectively.

The analysis of the behavior of $V(\cdot)$ is the key tool to obtain the convergence result stated in Theorem 1.

Theorem 1: Let the data generation mechanism be described by Assumption A2 and consider the estimation algorithm (1) with $a(\cdot)$ subject to Assumption A1. Then, for any given $\hat{\vartheta}(0)$ and $P(0) > 0$:

1) $\|\tilde{\vartheta}(t)\| \leq h$, $\forall t$, where h is a suitable constant;

2) $\lim_{t\to\infty} \tilde{\vartheta}_E(t) = 0$.

Proof: The following recursion for $\tilde{\vartheta}(t)$ is easily derived from (1), (4):

$$\tilde{\vartheta}(t) = \tilde{\vartheta}(t-1) + a(t)\frac{P(t-1)\varphi(t)}{1 + r(t)}\epsilon(t)^2. \tag{5}$$

From (2), (3), and (5), a recursive expression for $V(t)$ can also be worked out as follows:

$$V(t) = \mu(t)V(t-1) - a(t)\mu(t)\frac{\epsilon(t)^2}{1 + r(t)}. \tag{6}$$

Since $a(t) > 0$ (in view of A1), (6) entails the following inequality:

$$V(t) \leq \mu(t)V(t-1) \tag{7}$$

so that the rate of decrease of $V(t)$ is controlled by the forgetting factor $\mu(t)$.

On the other hand, (2) and assumption A1 straightforwardly imply that

$$P(t)^{-1} \geq \mu(t)P(t-1)^{-1} + c\mu(t)\varphi(t)\varphi(t)'. \tag{8}$$

By repeatedly using (7) and (8) starting from $P(0)^{-1}$ and $V(0) = \tilde{\vartheta}(0)'P(0)^{-1}\tilde{\vartheta}(0)$, we obtain

$$\prod_{i=1}^{t}\mu(i)V(0) \geq \tilde{\vartheta}(t)'\left[P(0)^{-1}\prod_{i=1}^{t}\mu(i) + c\sum_{1}^{t}{}_{i}\left(\varphi(i)\varphi(i)'\prod_{k=i}^{t}\mu(k)\right)\right]\tilde{\vartheta}(t)$$
$$\geq \tilde{\vartheta}(t)'\left[P(0)^{-1} + cM(t)\right]\tilde{\vartheta}(t)\prod_{i=1}^{t}\mu(i)$$

where $M(t) = (\sum_{1}^{t}{}_{i})\varphi(i)\varphi(i)'$.

This implies that

$$V(0) \geq \lambda_{\min}\left[P(0)^{-1}\right]\|\tilde{\vartheta}(t)\|^2 \tag{9}$$

and

$$V(0)/c \geq \tilde{\vartheta}(t)'M(t)\tilde{\vartheta}(t). \tag{10}$$

Conclusion 1) is a direct consequence of (9).

As for the proof of 2), it will be shown that inequality (10) is sufficient to ensure that the parameter error tends to zero along the excitation subspace. Indeed, denoting by $G(t)$ the square root of $M(t)$, (10) can be written as

$$\left(V(0)/c\right)^{1/2} \geq \|G(t)\tilde{\vartheta}_E(t) + G(t)\tilde{\vartheta}_U(t)\|.$$

Using the triangular inequality, it follows that

$$\left(V(0)/c\right)^{1/2} \geq \left\|\frac{G(t)\tilde{\vartheta}_E(t)}{\|\tilde{\vartheta}_E(t)\|}\|\tilde{\vartheta}_E(t)\|\right\| - \left\|\frac{G(t)\tilde{\vartheta}_U(t)}{\|\tilde{\vartheta}_U(t)\|}\|\tilde{\vartheta}_U(t)\|\right\|. \tag{11}$$

From part 1), one can conclude that $\|\tilde{\vartheta}_U(t)\| \leq h$; moreover, since $\tilde{\vartheta}_U(t) \in \bar{\mathcal{E}}$, $G(t)\tilde{\vartheta}_U(t)/\|\tilde{\vartheta}_U(t)\|$ is obviously bounded. Hence, the second term on the right-hand side of (11) is bounded. Therefore, the first term on the right-hand side must be bounded as well. On the other hand, by means of the lemma given in the Appendix, one can conclude that $G(t)\tilde{\vartheta}_E(t)/\|\tilde{\vartheta}_E(t)\|$ diverges, so that $\tilde{\vartheta}_E(t)$ must converge to zero.

IV. Adaptive Control

In this section, it is assumed that the system under control is given by A2 with $\varphi(t)$ defined as

$$\varphi(t) = \left[u(t-d)\cdots u(t-d-n_u+1)y(t-d)\cdots y(t-d-n_y+1)\right]'. \tag{12}$$

As is well known, any system described by an ARX model can be reduced to such a form.

We will also assume that:

A3: d is the "true" input-output delay, i.e., the first element of vector ϑ^o is nonzero.

Then, denoting by $y^o(t)$ the reference signal, the minimum variance adaptive control law is obtained by solving

$$\varphi(t)'\hat{\vartheta}(t-d) = y^o(t) \tag{13}$$

with respect to $u(t-d)$. This can actually be done provided that the first element of the parameter estimate is nonzero. The coefficient $a(t)$ is incorporated in (1) just to this purpose. Precisely, $a(t)$ is generally set to 1. If the first element of $\hat{\vartheta}(t-d)$, as computed from (1d), turns out to be zero, then the updating step is repeated by choosing any value of $a(t)$, $a(t) \neq 1$, subject to Condition A1.

As usual in the minimum variance control context, we will also assume the following.

A4: The reference signal $y^o(\cdot)$ is bounded, i.e., $|y^o(t)| < \gamma$, $\forall t$.

A5: The system under control is minimum phase.

With reference to the notion of excitation subspace, introduced in Section II, define $\varphi_E(t)$ $[\varphi_U(t)]$ as the projection of $\varphi(t)$ on $\mathcal{E}$ $[\bar{\mathcal{E}}]$. From A2, (4), and (13), it is then easy to derive the following relationship between the reference signal $y^o(t)$ and the output $y(t)$ of the closed-loop system as follows:

$$y(t) = y^o(t) - \varphi_E(t)'\tilde{\vartheta}_E(t-d) - \varphi_U(t)'\tilde{\vartheta}_U(t-d). \tag{14}$$

Equation (14) points out how the parameter error $\tilde{\vartheta}(t-d)$ reflects into the output tracking error $e(t) = y(t) - y^o(t)$. As seen in Theorem 1, the only component of $\tilde{\vartheta}(t-d)$ which may not converge to zero is the one belonging to the unexcitation subspace. On the other hand, in view of the definition of $\bar{\mathcal{E}}$, the component $\varphi_U(t)$ of the observation vector tends to zero. Therefore, from (14) it is apparent that the effect of $\tilde{\vartheta}_U(t-d)$ on $e(t)$ becomes asymptotically negligible. This implies that, from the tracking error point of view, what really matters is the behavior of the control system on the excitation subspace only. As pointed out by Theorem 1, on this subspace the identification algorithm well performs in the long run, namely $\tilde{\vartheta}_E(t-d) \to 0$. Then, from (14), one can conclude that the effect of $\tilde{\vartheta}_E(t-d)$ on $e(t)$ becomes negligible too, provided that $\varphi_E(t)$ remains bounded. This question is addressed in the following theorem.

Theorem 2: Consider the system defined by Assumption A2 and (12), together with the controller implicitly given by (13), where the estimate $\hat{\vartheta}(\cdot)$ is supplied by algorithm (1) subject to A1. Under Assumptions A3–A5, the tracking error $e(t) = y(t) - y^o(t)$ tends to zero and the control variable $u(\cdot)$ is bounded.

Proof: Thanks to Assumptions A3 and A5, [4, Lemma 3.2] can be applied to yield

$$\|\varphi_E(t)\| \leq \|\varphi(t)\| \leq c_1 + c_2 \max_{1 \leq \tau \leq t} |y(\tau)|. \tag{15}$$

From (14) and (15), taking into account Assumption A4, it follows that

$$|y(t)| \leq \gamma + \left[c_1 + c_2 \max_{1 \leq \tau \leq t} |y(\tau)|\right] \|\tilde{\vartheta}_E(t-d)\| + \|\varphi_U(t)\| \, \|\tilde{\vartheta}_U(t-d)\|. \tag{16}$$

It will now be shown that $y(\cdot)$ is bounded. Indeed, in the opposite, there should exist a sequence of time points t_i such that

$$|y(t_i)| \geqq |y(t)|, \quad \forall t < t_i, \text{and } |y(t_i)| \to \infty.$$

Then, consider inequality (16) at time points t_i, and observe that $\max_{1 \leq \tau \leq t_i} |y(\tau)| = |y(t_i)|$. When $t_i \to \infty$, the third term on the right-hand side of (16) tends to zero since $\tilde{\vartheta}_U(t_i - d)$ is bounded (Theorem 1) and $\varphi_U(t_i) \to 0$ (definition of unexcitation subspace). As for the second term, $\tilde{\vartheta}_E(t_i - d) \to 0$ (Theorem 1, again). Therefore, there exists a $\bar{t}_i$ such that

$$|y(t_i)| \leq \alpha(t) |y(t_i)| + \text{const.}, \quad \forall t_i \geq \bar{t}_i$$

where $\alpha(t) \to \infty$. This is an obvious contradiction, since $|y(t_i)| \to \infty$. Therefore, $y(\cdot)$ is bounded.

Then, (15) implies that $\varphi_E(\cdot)$ is bounded as well, so that in view of the discussion following relation (14), the tracking error $y(t) - y^o(t)$ tends to zero. Finally, from (15) again, it is obvious that the control variable $u(\cdot)$ is bounded.

Remark: It is interesting to compare the approach taken herein to the classical one proposed in [4]. The rationale of [4] represented a breakthrough for the convergence analysis of LMS-type algorithms. Subsequently, its extension to the RLS case was the subject of many papers. However, the line of reasoning in [4] can be adopted for RLS algorithms at the price of equipping them with an alien control mechanism to keep the covariance matrix bounded (see [5], [6]). The approach taken herein looks much more natural for the convergence analysis of RLS-type algorithms. Indeed, by no additional constraints on the covariance matrix, it can be shown that convergence holds true. Moreover, the analysis enables one to put sharply into focus the features of the identification algorithm relevent to the closed-loop analysis. On the other hand, our approach cannot be straightforwardly applied to the LMS case, since it can be shown that property 2) of Theorem 1 is no more true in that case.

V. Concluding Remarks

A fairly general class of adaptive RLS algorithms has been considered in this note. First, by introducing the notion of excitation subspace, the convergence analysis of the (open-loop) identification algorithms is carried out. The results so obtained lend themselves to be easily applied to the control context, so leading to a convergence result for a wide class of minimum variance adaptive control schemes based on RLS.

Appendix

Lemma: Consider an increasing sequence $B(\cdot)$ of real positive $n \times n$ matrices and let $\mathscr{S} = \{v \in \mathbb{R}^n | \exists L < \infty: v'B(t)v < L, \forall t\}$. Then, $\forall x(\cdot)$ such that $x(t) \in \mathscr{S}^\perp$, $\|x(t)\| = 1$, $\forall t$, it follows that $\lim_{t \to \infty} x(t)'B(t)x(t) = \infty$.

Proof: The proof will be given by contradiction. Suppose that there exists a sequence $w(\cdot)$ with $w(t) \in \mathscr{S}^\perp$, $\|w(t)\| = 1$, $\forall t$ such that $w(t)'B(t)w(t)$ does not diverge. Then, there exist a constant k and a subsequence $\{w(t_i), i = 1, 2, \cdots\}$ such that

$$w(t_i)'B(t_i)w(t_i) < k, \quad \forall i. \tag{17}$$

Consider now a subsequence $\{w(\tau_i), i = 1, 2, \cdots\}$ of sequence $\{w(t_i), i = 1, 2, \cdots\}$ such that

$$\lim_{i \to \infty} w(\tau_i) = \overline{w}.$$

Since $\overline{w} \in \mathscr{S}^\perp$, $\overline{w}'B(t)\overline{w}$ diverges, so that there exists a time point $\bar{t}$ such that

$$\overline{w}'B(\bar{t})\overline{w} \geq 2k. \tag{18}$$

Then, in view of the monotonicity of $B(\cdot)$, for the time points $t = \{\tau_i, i = 1, 2, \cdots\} \cap (t \geq \bar{t})$, one obtains

$$w(t)'B(t)w(t) \geq w(t)'B(\bar{t})w(t), \quad \forall t.$$

The left-hand side function of t is bounded from above by (17), while the right-hand side function is asymptotically bounded from below by (18). These bounds are obviously incompatible with each other. □

Interestingly enough, the conclusion stated in the aforementioned lemma does not hold true anymore if the assumption of monotonicity of $B(\cdot)$ is dropped out.

References

[1] R. M. Johnstone, C. R. Johnson, R. R. Bitmead, and B. D. O. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor," *Syst. Contr. Lett.*, vol. 2, pp. 77–82, 1982.

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

[3] R. M. Canetti and M. D. España, "Variable forgetting factor least-squares identification algorithm: Convergence analysis," in *Proc. 8th IFAC/IFORS Symp. Ident. Syst. Parameter Estimation*, Vol. 1, Beijing, China, 1988, pp. 488–491.

[4] G. C. Goodwin, P. J. Ramadge, and P. E. Caines, "Discrete-time multivariable adaptive control," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 3, pp. 449–453, 1980.

[5] A. O. Cordero and Q. Mayne, "Deterministic convergence of a self-tuning regulator with variable forgetting factor," *IEE Proc.*, part D, vol. 128, pp. 19–23, 1981.

[6] G. C. Goodwin and H. G. Elliot, "Deterministic convergence of a self-tuning regulator with covariance resetting," *IEE Proc.*, vol. 120, pp. 6–8, 1983.

Dynamical Discontinuous Feedback Control of Nonlinear Systems

HEBERTT SIRA-RAMIREZ AND PABLO LISCHINSKY-ARENAS

***Abstract*—In this note, a technique is presented for the analysis of discontinuous dynamical feedback regulation of nonlinear systems. A pulse width modulation feedback interconnection scheme, with general duty ratio function, is shown to be easily analyzable in terms of an average model which captures the essential features of the discontinuously feedback controlled system.**

Manuscript received August 18, 1989; revised February 16, 1990. This work was supported by the Consejo de Desarrollo Cientifico, Humanístico y Tecnológico of the Universidad de Los Andes under Grant I-325-90.

H. Sira-Ramirez is with the Departamento Sistemas de Control, Escuela de Ingeniería de Sistemas, Universidad de Los Andes, Mérida, Venezuela.

P. Lischinsky-Arenas is with the Departamento de Computación, Escuela de Ingeniería de Sistemas, Universidad de Los Andes, Mérida, Venezuela.

IEEE Log Number 9039310.

0018-9286/90/1200-1373$01.00