FA1-11:15

EXPONENTIAL CONVERGENCE OF RECURSIVE LEAST SQUARES WITH EXPONENTIAL FORGETTING FACTOR*

Richard M. Johnstone[1], C. Richard Johnson Jr.[2], Robert R. Bitmead[3] and Brian D.O. Anderson[1]

[1]Department of Systems Engineering, Institute of Advanced Studies, Australian National University, Canberra, ACT 2600, Australia

[2]School of Electrical Engineering, Cornell University, Ithaca, NY 14853, USA

[3]Department of Electrical and Electronic Engineering, James Cook University, Townsville, Qld 4811, Australia

*The work of Authors[1] was supported by the Australian Research Grants Committee, while that of Author[2] was supported by NSF grant ECS-8120998.

ABSTRACT

This paper demonstrates that, provided the system input is persistently exciting, the recursive least squares estimation algorithm with exponential forgetting factor is exponentially convergent. Further, it is shown that the incorporation of the exponential forgetting factor is necessary to attain this convergence and that the persistence of excitation is virtually necessary. The result holds for stable finite-dimensional, linear, time-invariant systems but has its chief implications to the robustness of the parameter estimator when these conditions fail.

1. INTRODUCTION

This paper deals with the exponential stability of a popular adaptive estimation scheme - the recursive least squares (RLS) algorithm. In practice, the RLS algorithm is modified to incorporate an exponential forgetting factor (λ) between zero and one. The reason for including λ has most often been justified by noting that such an inclusion discounts past measurements in favour of more recent ones [1] and so offers some capacity to identify time-varying plants. Without exponential forgetting ($\lambda = 1$), and with rich inputs, the RLS algorithm effectively turns itself off, after some time. This, of course, means that the RLS algorithm progressively loses the ability to identify the time-varying plant parameters. Conversely, with λ less than 1, one maintains tracking capabilities at the cost of non-zero limiting parameter covariance.

Our main result is that, if the system input $\{u_k\}$ is persistently exciting the RLS algorithm with exponential forgetting factor is exponentially convergent, in that the parameter error vector (the difference between the actual plant and the estimation scheme parameters) approaches zero exponentially fast. Further it is shown that without this exponential forgetting factor (ie $\lambda = 1$) exponential convergence is lost. It is this property of exponential convergence which is necessary to ensure robust performance in the presence of measurement noise, time-variations of plant parameters, and under-modelling of plant order. The ease with which an exponential overbound on the parameter error is found contrasts sharply with the relative difficulty [2,9] in establishing such a convergence rate for steepest-descent type algorithms. This reflects the Newton-Raphson-like nature of RLS.

In view of the recent results of [2] the need for persistent excitation of $\{u_k\}$ should not be surprising. Indeed, if we do not have persistent excitation of $\{u_k\}$, then it is trivial to show that we cannot ensure convergence of the parameter error to zero, at all.

We also show how an estimation scheme due to Landau [3] can be modified so that an exponential forgetting factor is incorporated as in [4]. An outline of the proof of exponential convergence of the modified algorithm is given.

The approach of the present note is limited to linear, time-invariant, finite-dimensional, stable plants which are single-input, single-output (SISO) and driven by deterministic inputs. The SISO restriction can doubtless be removed with an appropriate parameterisation scheme [5]. Extensions to stochastic inputs along the lines of [6] also seem readily possible, although technically involved. Nonlinear, time-varying or infinite-dimensional plants can only be accommodated to the extent that the results given here are robust in the presence of some nonlinearity, some parameter time-variation and some modes which are neglected. See [7-9] for discussion of the effects of departure from ideality in adaptive system problems. The significance of this paper is that it demonstrates that the convergence rate advantages of RLS over steepest-descent algorithms need not be abandoned in order to gain adaptability.

2. THE RLS ALGORITHM WITH EXPONENTIAL FORGETTING

The RLS algorithm can be described by the following equations

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{P_{k-2}\psi_{k-1}[y_k - \psi_{k-1}^T\hat{\theta}_{k-1}]}{\lambda + \psi_{k-1}^T P_{k-2}\psi_{k-1}} \quad (2.1)$$

and

$$P_{k-1} = \frac{1}{\lambda}\left[I - \frac{P_{k-2}\psi_{k-1}\psi_{k-1}^T}{\lambda + \psi_{k-1}^T P_{k-2}\psi_{k-1}} \right] P_{k-2} \quad (2.2)$$

or

$$P_{k-1}^{-1} = \lambda P_{k-2}^{-1} + \psi_{k-1}\psi_{k-1}^T \quad (2.3)$$

where $\lambda \varepsilon [0,1]$, P_{-1} is a positive definite matrix, and $\hat{\theta}_0$ is arbitrary. The parameter estimate vector $\hat{\theta}$ can be shown [1] to converge to the value θ which minimizes the quadratic criterion

$$J_k(\theta) = \sum_{i=1}^{k} \lambda^{k-i} [y_i - \psi_{i-1}^T \theta]^2 \tag{2.4}$$

where y is the output of the plant being identified and ψ is the measurable information vector from which y is presumed to be formed.

If we now define a parameter error vector as

$$\tilde{\theta}_k = \theta - \hat{\theta}_k \tag{2.5}$$

and restrict our attention to the "homogeneous" case [6], where y_k is given exactly by $\psi_{k-1}^T \theta$, then (2.3) takes on the form

$$\tilde{\theta}_k = (I - \frac{P_{k-2}\psi_{k-1}\psi_{k-1}^T}{\lambda + \psi_{k-1}^T P_{k-2}\psi_{k-1}})\tilde{\theta}_{k-1} \tag{2.6}$$

which is a homogeneous, time-varying, linear difference equation.

With $\lambda = 1$ in the above equations we have the standard RLS algorithm, while choosing $\lambda < 1$ can be seen via (2.4) to be effecting an exponential weighing on the past data. Similarly, (2.3) is an asymptotically stable recursion only for $\lambda < 1$ so that if ψ is persistently exciting P^{-1} will remain bounded.

The question of the origin of the ψ measurement vector will for the moment be left in abeyance, to be addressed later in the discussion of the persistence of excitation condition and its interpretation in, for example, ARMA modelling.

3. MAIN RESULTS

As a preliminary to the main result we define persistence of excitation and then prove that if $\{\psi_k\}$ is persistently exciting then P_{k-1}^{-1} is uniformly bounded above and uniformly positive definite (ie bounded away from zero). This latter property can be seen to follow from (2.3) because, with $\lambda < 1$, P_k^{-1} is the output of an exponentially stable system driven by a persistent input.

Definition

We say that the vector sequence $\{\psi_k\}$ is persistently exciting if for some constant integer S and all j there exist positive constants α and β such that

$$0 < \alpha I \leq \sum_{i=j}^{j+S} \psi_i \psi_i^T \leq \beta I < \infty \tag{3.1}$$

Lemma 1: Suppose that $\{\psi_k\}$ is persistently exciting then for all $k \geq S$

$$\frac{\beta}{1-\lambda^{s+1}} I + O[\lambda^k] \geq P_{k-1}^{-1} \geq \frac{\alpha(\frac{1}{\lambda} - 1)}{(\frac{1}{\lambda})^{s+1} - 1} I > 0 \tag{3.2}$$

Proof:

From (2.3) and (3.1) we obtain that

$$[P_{j-1}^{-1} + \dots + P_{j+s-1}^{-1}] \geq \sum_{k=j}^{j+s} \psi_{k-1} \psi_{k-1}^T \geq \alpha I \tag{3.3}$$

Also, from (2.3) it is obvious that

$$P_{j-1}^{-1} \geq \lambda P_{j-2}^{-1} \tag{3.4}$$

so that for $j \geq 0$,

$$(\frac{1}{\lambda^s} + \frac{1}{\lambda^{s-1}} + \dots + 1) P_{j+s-1}^{-1} \geq \alpha I \tag{3.5}$$

and consequently for all $k \geq s$

$$P_{k-1}^{-1} \geq \frac{\alpha(\frac{1}{\lambda} - 1)}{\frac{1}{\lambda^{s+1}} - 1} I > 0 \tag{3.6}$$

The upper bound follows from the solution of

$$(P_{j-1}^{-1} + \dots + P_{j+s-1}^{-1}) \leq \lambda(P_{j-2}^{-1} + \dots + P_{j+s-2}^{-1}) + \sum_{k=j}^{j+s} \psi_{k-1}\psi_{k-1}^T \tag{3.7}$$

Using (3.1) we conclude that

$$(P_{j-1}^{-1} + \dots + P_{j+s-1}^{-1}) \leq \frac{1 - \lambda^{j-s}}{1 - \lambda} \beta I + \lambda^{j-s}[P_{s-1}^{-1} + \dots + P_{-1}^{-1}] \tag{3.8}$$

or, because $\lambda P_{j-1}^{-1} \leq P_j^{-1}$

$$P_{j-1}^{-1} \leq \frac{1 - \lambda^{j-s}}{1 - \lambda^{s+1}} \beta I + O[\lambda^j] \leq \frac{\beta}{1 - \lambda^{s+1}} I + O[\lambda^i j] \tag{3.9}$$

We can now state our main result.

Theorem 1: If the measurement vector sequence $\{\psi_k\}$ is persistently exciting, then the RLS estimation scheme with exponential forgetting factor is exponentially stable.

Proof: Consider equations (2.2), (2.3) and (2.5). We choose a Lyapunov function candidate as

$$V_k = \tilde{\theta}_k^T P_{k-1}^{-1} \tilde{\theta}_k \tag{3.10}$$

Then straightforward algebra using (2.3) and (2.1) yields

$$\begin{aligned} V_k - V_{k-1} &= \tilde{\theta}_k^T P_{k-1}^{-1} \tilde{\theta}_k - \tilde{\theta}_{k-1}^T P_{k-2}^{-1} \tilde{\theta}_{k-1} \\ &= \tilde{\theta}_{k-1}^T [(\lambda-1)P_{k-2}^{-1} - \frac{\lambda\psi_{k-1}\psi_{k-1}^T}{\lambda + \psi_{k-1}^T P_{k-2}\psi_{k-1}}] \tilde{\theta}_{k-1} \\ &\leq (\lambda-1)\tilde{\theta}_{k-1}^T P_{k-2}^{-1} \tilde{\theta}_{k-1} \\ &= (\lambda-1)V_{k-1} \end{aligned} \tag{3.11}$$

and therefore

$$V_k \leq \lambda V_{k-1} \leq \lambda^k V_0 = \lambda^k \tilde{\theta}_0^T P_{-1}^{-1} \tilde{\theta}_0 \tag{3.12}$$

Finally, (2.3), (3.10) and Lemma 1 (particularly the lower bound of (3.3)) implies for $k \geq s$

$$\|\tilde{\theta}_k\|^2 \leq \frac{(\frac{1}{\lambda})^{s+1} - 1}{\alpha(\frac{1}{\lambda} - 1)} \lambda^k \lambda_{max}(P_{-1}^{-1}) \|\tilde{\theta}_0\|^2 \tag{3.13}$$

Q.E.D.

The main result of Theorem 1 depends critically on the persistent excitation of $\{\psi_k\}$. Lemma 1 showed that P_k and P_k^{-1} were bounded with persistently exciting $\{\psi_k\}$ and if P_k is unbounded it is not possible to conclude from the fact that $\tilde{\theta}_k^T P_{k-1}^{-1} \tilde{\theta}_k \to 0$ that $\tilde{\theta}_k \to 0$.

One can also easily show that <u>if ψ_k is bounded, persistency of excitation of ψ_k is necessary for P_k to be bounded</u>. Suppose that $0 < A < \| P_k^{-1} \| \leq B < \infty$ for all k. Choose s so that $\lambda^s B < \frac{A}{2}$. Then from (2.3)

$$P_{k-1}^{-1} = \lambda^s P_{k-s-1}^{-1} + \sum_{i=0}^{s-1} \lambda^i \psi_{k-1-i} \psi_{k-1-i}^T \tag{3.14}$$

Applying the bounds shows that

$$\sum_{i=0}^{s-1} \lambda^i \psi_{k-1-i} \psi_{k-1-i}^T > \frac{A}{2} I$$

or

$$\sum_{i=0}^{s-1} \psi_{k-d-i} \psi_{k-d-i}^T > \frac{A}{2} \lambda^{-(s-1)} I \tag{3.15}$$

so that ψ_k is persistently exciting.

ARMA Modelling

Given that Theorem 1 guarantees exponential convergence of RLS with $\lambda < 1$ subject to sufficient excitation of the measurement vector ψ_k, we note that this property of ψ_k can be achieved in the important ARMA modelling situation by requiring stability of the plant, minimality of the parameterization, and persistence of excitation of the input sequence $\{u_k\}$ only. In particular, if the plant satisfies

$$y_k + \alpha_1 y_{k-1} + \ldots + \alpha_n y_{k-n} = \beta_d u_{k-d} + \ldots + \beta_m u_{k-m} \tag{3.16}$$

and is stable, and further

$$\alpha(z) = z^n + \alpha_1 z^{n-1} + \ldots + \alpha_n \tag{3.17}$$

$$\beta(z) = \beta_d z^{n-d} + \beta_{d+1} z^{n-d-1} + \ldots + \beta_m z^{n-m} \tag{3.18}$$

are relatively prime, then Theorem 2.2 of [2] ensures persistence of excitation of $\psi_k = [y_k \ldots y_{k-n+1}\ u_{k-d} \ldots u_{k-m+1}]^T$ if $[u_{k-d} \ldots u_{k-m-n}]$ is persistently exciting.

Nonexponential Convergence of RLS with $\lambda = 1$

To demonstrate that RLS with $\lambda = 1$ is not exponentially convergent in spite of sufficiently exciting ψ_k we consider the following scalar example.

Let θ and $\hat{\theta}$ be scalars and the persistently exciting scalar $\psi_k = 1$ for all k. Then choosing $P_{-1} = c$ we have from (2.3)

$$P_k = \frac{1}{c+k+1}$$

and using (2.1),

$$\tilde{\theta}_{k+1} = [1 - \frac{1}{c+k+1}] \tilde{\theta}_k = \prod_{i=0}^{k} [1 - \frac{1}{c+i+1}] \tilde{\theta}_o = \frac{c}{c+k+1} \tilde{\theta}_o$$

Consequently $\tilde{\theta}$ decay as $\frac{1}{K}$ which cannot be over bounded by an exponential and so does not converge exponentially fast.

These problems with slow convergence rate and non-persistently exciting inputs are also known to occur in Kalman filtering without exponential data weighting. A discussion along these lines concerned with data saturation and filter divergence is given in [10].

4. THE "CLASS B" ALGORITHM OF LANDAU WITH EXPONENTIAL WEIGHTING

In [3] Landau details a "Class B" algorithm which is very much like the RLS algorithm of this paper. The equations which describe this algorithm are:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{P_{k-1}\psi_{k-1}\nu_k^o}{1+\psi_{k-1}^T P_{k-1}\psi_{k-1}} \tag{4.1}$$

$$P_k = P_{k-1} - \frac{1}{\lambda} \frac{P_{k-1}\psi_{k-1}\psi_{k-1}^T P_{k-1}}{1+\frac{1}{\lambda}\psi_{k-1}^T P_{k-1}\psi_{k-1}} \quad \lambda > 0.5;\ P_o > 0 \tag{4.2}$$

and

$$\nu_k^o = y_k - \psi_{k-1}^T \hat{\theta}_{k-1} + \sum_{i=1}^{n} d_i \varepsilon_{k-i} \tag{4.3}$$

$$\varepsilon_k = y_k - \psi_{k-1}^T \hat{\theta}_k \tag{4.4}$$

where

$$\psi_{k-1}^T = [z_{k-1} \ldots z_{k-n}\ u_{k-1} \ldots u_{k-m}] \tag{4.5}$$

and z_k is the output of the "parallel" estimation model (see [3], pp. 164-167 and [2], Section 3).

Provided that

$$h(z^{-1}) = \frac{1 + \sum_{i=1}^{n} d_i z^{-1}}{1 + \sum_{i=1}^{n} \alpha_i z^{-i}} - \frac{1}{2\lambda} \tag{4.6}$$

is a strictly positive real transfer function, the alorithm converges in the sense that the parameter error goes to zero. The extra complexity of the algorithm is offset by a tendency for measurement noise to cause less damage than for an ordinary least squares algorithm.

The similarity of (4.1)-(4.5) to (2.1)-(2.3) is obvious. In fact the only striking differences are that the Landau algorithm assumes unity time-delay in the plant (i.e. d = 1); there is an additional term appearing in (4.3) which does not appear in (2.1); "parallel" estimation model outputs, z_{k-i}, appear in (4.3) whereas ARMA plant outputs y_{k-i}, appear in the definition of ψ_{k-d} in Section 3 of this paper and, finally, (4.2) does not incorporate an exponential forgetting factor.

Landau has proved the asymptotic stability of the class B algorithm using hyperstability arguments. Unfortunately one cannot prove exponential stability of the algorithm as it stands. This is essentially because the matrix sequence $\{P_k^{-1}\}$ cannot be guaranteed to be bounded. In particular if $\{\psi_{k-1}\}$ is persistently exciting then use of the matrix Inversion Lemma and (4.2) shows that $P_k^{-1} \to \infty$. This problem may be circumvented if we use a slight modification of the class B algorithm [4], which amounts to introduction of exponential forgetting.

Use

$$P_k = \frac{1}{\alpha} [I - \frac{1}{\lambda} \frac{P_{k-1}\psi_{k-1}\psi_{k-1}^T}{\alpha+\frac{1}{\lambda}\psi_{k-1}^T P_{k-1}\psi_{k-1}}] P_{k-1} \qquad 0 < \alpha < 1 \tag{4.7}$$

in place of (4.2). By using similar arguments to those in [2] (see Section 3 of [2]) it is then possible to prove the following result. Due to space limitations only an outline of the proof is given.

Theorem 2: Suppose that

(a) the plant is described by (3.16) (with $d = 1$) and $\alpha(z^{-1})$ is stable,

(b) $\alpha(z^{-1})$ and $\beta(z^{-1})$ (with $d = 1$), of (3.17), (3.18), are relatively prime,

(c) $\{u_k\}$ is a persistently exciting input sequence

(d) $h(z^{-1})$ in (4.6) is strictly positive real.

then the modified class B algorithm of (4.1), (4.7) and (4.3)-(4.5) is exponentially stable.

Outline of Proof:

(a) A Lyapunov function candidate is chosen as

$$V[e_k, \tilde{\theta}_k] = e_k^T e_k + \tilde{\theta}_k^T \frac{1}{\alpha} P_{k-1}^{-1} \tilde{\theta}_k \qquad (4.8)$$

where

$$e_k^T = [y_{k-1} - z_{k-1} \dots y_{k-n} - z_{k-n}]$$

(b) Following [2] we set-up state variable equations equivalent to (4.1), (4.7) and (4.3)-(4.4).

(c) Making use of the Kalman-Yakubovic lemma equations [2] we obtain an expression for $V[e_{k+1}, \tilde{\theta}_{k+1}] - V[e_k, \tilde{\theta}_k]$. This quantity is monotone decreasing unless $e_k = 0$ (and, in case P_{k-1}^{-1} is bounded, $\tilde{\theta}_k = 0$).

(d) We use arguments like those in the proof of Lemma 1 to show that if $\{u_k\}$ is persistently exciting then $\chi_{k-1}^T = [y_{k-1} \dots y_{k-n} u_{k-n} \dots u_{k-m}]$ is persistently exciting.

(e) From the asymptotic stability results of [4] we conclude that $z_k \to y_k$. Lemma 3.2 of [2] then allows us to conclude that $\{\psi_{k-1}\}$ is also persistently exciting.

(f) The Matrix Inversion Lemma and (4.7) gives

$$P_k^{-1} = \alpha P_{k-1}^{-1} + \frac{1}{\lambda} \psi_{k-1} \psi_{k-1}^T \qquad (4.9)$$

We conclude, then, using arguments as in the proof of Lemma1 that if $\{\psi_{k-1}\}$ is persistently exciting then $\{P_k^{-1}\}$ is a bounded matrix sequence.

(g) Finally, (c), (f) and the form of (4.8) allows us to claim that the modified class B algorithm of Landau is exponentially stable.

5. CONCLUSIONS

The main result demonstrates that the RLS estimation scheme with exponential forgetting factor is *exponentially stable* provided that ψ_k is persistently exciting. With bounded input signals the converse is also true. This algorithm when used with ARMA models is sometimes called a "series-parallel" algorithm [3]; we have also outlined an exponential stability result for the corresponding "parallel" RLS scheme, with exponential forgetting.

These results are potentially very important in the cases of time-varying systems or under-modelling [7-9], and imply robustness of the algorithms in the presence of measurement noise. Also, the results complement those known for the stochastic approximation-type algorithm [2].

REFERENCES

[1] G.C. Goodwin and R.L. Payne, Dynamic System Identification: Experiment Design and Analysis, Academic Press, 1977.

[2] B.D.O. Anderson and C.R. Johnson Jr., Exponential convergence of adaptive identification and control algorithms, *Automatica* (1981) to appear. Also IFAC 8th Triennial World Congress, Kyoto, Japan, August 1981.

[3] I.D. Landau, *Adaptive Control* (Dekker, New York, 1979).

[4] I.D. Landau and H.M. Silveira, A stability theorem with applications to adaptive control, *Proc. 17th IEEE CDC Conf,* (1979) 1074-1079.

[5] R. Schumann, Identification and adaptive control of multivariable stochastic linear systems, *5th IFAC Symposium on Identification and Parameter Estimation* (1979).

[6] R.R. Bitmead and B.D.O. Anderson, Lyapunov techniques for the exponential stability of linear difference equations with random co-efficients, *IEEE Trans. on Auto. Contr.* AC-25 (1980) 782-787.

[7] B.D.O. Anderson and R.M. Johnstone, Robust Lyapunov stability results and adaptive systems, *IEEE CDC Conf.,* (1981) to appear.

[8] B.D.O. Anderson and C.R. Johnson, Jr., On reduced order adaptive output error identification and adaptive IIR filtering, *IEEE Trans. on Auto. Contr,* AC-27 (1982), to appear.

[9] A. Weiss and D. Mitra, Digital adaptive filters: Conditions for convergence, rates of convergence, effects of noise and errors arising from the implementation, *IEEE Trans. on Info. Theory* IT-25, (1979) 637-652.

[10] B.D.O. Anderson and J.B. Moore, Optimal Filtering, Prentice-Hall, 1979.