PERGAMON

Automatica 38 (2002) 1287–1293

automatica

www.elsevier.com/locate/automatica

Asymptotically optimal smoothing of averaged LMS estimates for regression parameter tracking ☆

Alexander V. Nazin[a,1], Lennart Ljung[b,*]

[a] Institute of Control Sciences, Profsoyuznaya str., 65, 117997 Moscow, Russia
[b] Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden

Received 28 June 2001; accepted 29 January 2002

Abstract

The sequence of estimates formed by the LMS algorithm for a standard linear regression estimation problem is considered. It is known since earlier that smoothing these estimates by simple averaging will lead to, asymptotically, the recursive least-squares algorithm. In this paper, it is first shown that smoothing the LMS estimates using a matrix updating will lead to smoothed estimates with optimal tracking properties, also in case the true parameters are slowly changing as a random walk. The choice of smoothing matrix should be tailored to the properties of the random walk. Second, it is shown that the same accuracy can be obtained also for a modified algorithm, SLAMS, which is based on averages and requires much less computations. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Linear regression; LMS; Slow random walk; Parameter tracking; Smoothing; Asymptotic MSE

1. Introduction

Tracking of time varying parameters is a basic problem in many applications, and there is a considerable literature on this problem. See, among many references, e.g. Widrow and Stearns (1985), Ljung and Söderström (1983), and Ljung and Gunnarsson (1990).

One of the most common methods is the least mean-squares (LMS) algorithm (Widrow & Stearns, 1985) which is a simple gradient-based stochastic approximation (SA) algorithm. For time-invariant systems LMS does not have optimal accuracy; the accuracy could in fact be quite bad. It is well known that for such systems, the recursive least squares (RLS) algorithm is optimal, but it is on the other hand considerably more complex. A very nice observation, independently made by Polyak and Ruppert, is that this optimal accuracy can asymptotically also be obtained by a simple averaging of the LMS-estimates. See Polyak (1990), Ruppert (1988), and Polyak and Juditsky (1992), Kushner and Yang (1993) for the analysis.

In Ljung (2001) it is shown that this asymptotic convergence of the averaged LMS-algorithm to the RLS algorithm is obtained also for the tracking case, with a moving true system and constant gain algorithms. This means that, in general, the averaged algorithm will not give optimal accuracy. Optimal tracking properties then will be obtained by a Kalman-filter-based algorithm where the update direction is carefully tailored to the regressor properties, the character of the changes in the true parameter vector and the noise level.

In this paper we shall consider a more general post-processing of the LMS-estimates, obtained from a constant gain, unnormalized LMS-method. The general version of this algorithm we call SLAMS—smoothed averaged LMS (allowing a metathesis for pronouncability). It consists of first forming the standard LMS-estimates $\hat{\theta}(t)$, and then forming simple averages of these

$$\tilde{\theta}(t) = \frac{1}{m}\sum_{\tau=t-m}^{t-1}\hat{\theta}(\tau+1)$$

and finally smoothing these by a simple exponential smoother, applying a direction correction every mth sample

$$\bar{\theta}(t) = \begin{cases} \bar{\theta}(t-m) - \gamma S(\bar{\theta}(t-m) - \tilde{\theta}(t)), \\ \quad t = km, \quad k = 1,2,\ldots \\ \bar{\theta}(t-1) - \gamma(\bar{\theta}(t-1) - \hat{\theta}(t-1)), \\ \quad t \neq km, \quad k = 1,2,\ldots . \end{cases}$$

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor H. Hjalmarsson under the direction of Editor Torsten Soederstroem.

* Corresponding author. Tel.: +46-13-281310; fax: +46-13-282622.
E-mail addresses: nazine@ipu.rssi.ru (A.V. Nazin), ljung@isy.liu.se (L. Ljung).

[1] The work of the first author has been carried out while visiting Linköping University as Guest Researcher.

PII: S0005-1098(02)00028-6

This algorithm has the design variables μ (the gain of the LMS algorithm), S, m and γ. By, for example, choosing m as the dimension of θ the average number of operations per update in the SLAMS algorithm is still proportional to $\dim\theta$, just as in the simple LMS algorithm.

The main goal of this paper is to establish an asymptotic expression for the covariance matrix of the tracking error $\bar{\theta}(t)-\theta(t)$ ($\theta(t)$ being the true parameter value). We show that, by the choice of S and γ we can obtain the same asymptotic covariance as the optimal Kalman filter gives, regardless of m and μ (as long as it has a certain size relation to γ).

In Section 2 we formulate the tracking problem and state the basic assumptions. In Sections 3 we treat a special case of SLAMS, with m fixed to 1. The extension to the general algorithm is done in Section 4.

2. Problem statement and basic assumptions

Consider a discrete-time linear regression model with time-varying parameters. It means that the observed data $\{y(t),\varphi(t),t=1,\dots\}$ are generated by the linear regression structure

$$y(t)=\theta^{\mathrm{T}}(t)\varphi(t)+e(t), \qquad (1)$$

$$\theta(t)=\theta(t-1)+w(t), \qquad (2)$$

where $e(t)\in R$ and $w(t)\in R^n$ stand for observation error and parameter change, respectively. Due to the following assumptions, Eq. (2) describes evolution of slowly drifting parameter $\theta(t)\in R^n$ as a random walk.

Basic assumptions

A1. The sequences $\{e(t)\}$, $\{w(t)\}$ and $\{\varphi(t)\}$ are i.i.d. mutually independent sequences of random variables.

A2. The observation error $e(t)$ is unbiased and has a finite variance, that is $Ee(t)=0$ and $Ee^2(t)=\sigma_e^2\in(0,\infty)$.

A3. The parameter change $w(t)$ is an unbiased variable with positive definite covariance matrix, i.e., $Ew(t)=0$ and $Ew(t)w^{\mathrm{T}}(t)=\gamma^2R_w>0$, where a priori known $\gamma>0$ represent small parameter of the problem under consideration.

A4. The regressor covariance matrix is non-singular, i.e. $E\varphi(t)\varphi^{\mathrm{T}}(t)=Q>0$; moreover, $E|\varphi(t)|^4<\infty$.

A5. The initial parameter value $\theta(0)$ is supposed to be fixed (for the sake of simplicity).

Consider the parameter tracking problem with the performance evaluated as the asymptotic mean square error (MSE). That is, the problem is to design an estimation algorithm which on-line delivers an estimate sequence $\{\bar{\theta}(t)\}$ on the basis of past observations (1) with a minimal asymptotic error covariance matrix U:

$$U=\lim_{t\to\infty}U_t, \qquad (3)$$

where

$$U_t=E(\bar{\theta}(t)-\theta(t))(\bar{\theta}(t)-\theta(t))^{\mathrm{T}}. \qquad (4)$$

As is known (see, e.g. Nazin and Yuditskii (1991) and the lower bound below), matrix U must be proportional to the small parameter γ, that is

$$U=\gamma U_0+o(\gamma)\quad \text{as } \gamma\to+0. \qquad (5)$$

Moreover, from the lower bound for matrix U_0 proved in Nazin and Yuditskii (1991) follows, in particular, that with Gaussian random variables $e(t)$ and $w(t)$

$$U_0\geqslant U_{\mathrm{lb}}\quad \text{for any parameter estimator},$$

where U_{lb} is a symmetric solution to the equation $U_{\mathrm{lb}}QU_{\mathrm{lb}}=\sigma_e^2R_w$, that is

$$U_{\mathrm{lb}}=\sigma_eQ^{-1/2}(Q^{1/2}R_wQ^{1/2})^{1/2}Q^{-1/2}. \qquad (6)$$

By matrix inequality $A\geqslant B$ is meant that $A-B$ is a positive semidefinite matrix.

We further call the matrix U_0 (5) the limiting asymptotic error covariance matrix, since

$$U_0=\lim_{\gamma\to+0}\gamma^{-1}U. \qquad (7)$$

Below we study how the matrix U_0 depends on the parameters, both of the problem and the algorithm described in the following section. We then minimize this matrix over the design parameters in the algorithm.

3. Parameter tracking by smoothed LMS

We, here, aim at studying the following recursive constant gain SA-like procedure, which we call smoothed LMS (SLMS):

$$\hat{\theta}(t)=\hat{\theta}(t-1)+\mu\varphi(t)(y(t)-\varphi^{\mathrm{T}}(t)\hat{\theta}(t-1)), \qquad (8)$$

$$\bar{\theta}(t)=\bar{\theta}(t-1)-\gamma S(\bar{\theta}(t-1)-\hat{\theta}(t-1)). \qquad (9)$$

Here $\mu>0$ is a scalar step size while S represents an $n\times n$-matrix gain. Relation (8) is exactly the constant (scalar) gain SA-algorithm (LMS), while the recursive procedure (9) generates a sequence of smoothed-SA estimates.

Special interest might be connected to the particular case of a "scalar matrix" S when $S=\rho I_n$ with a scalar step size $\rho>0$ and identity $n\times n$-matrix I_n (see Section 3.1 below). In that case there are no matrix calculations in algorithms (8) and (9), which make it particularly simple.

Assumptions on the algorithm parameters

B1. $\mu=o(1)$ as $\gamma\to+0$.

B2. $\gamma=o(\mu)$ as $\gamma\to+0$.

B3. The matrix $(-S)$ is stable, i.e., the real part of any eigenvalue of S is positive.

Remark. Due to assumptions B1 and B2, stochastic stability of Eqs. (8) and (9) (in mean-square sense) is obviously ensured (for sufficiently small γ). This implies the existence of limit in (3).

Theorem 1. *Let the assumptions* A1–A5 *and* B1–B3 *hold, and consider the estimates $\bar{\theta}(t)$ generated by algorithms* (8) *and* (9). *Then the limiting asymptotic error covariance matrix U_0, defined by* (5), *is the solution to the equation*

$$SU_0 + U_0S^{\mathrm{T}} = R_w + \sigma_e^2 SQ^{-1}S^{\mathrm{T}}. \tag{10}$$

Remark. The relationship (10) is a Lyapunov equation with respect to U_0, see, e.g. Lancaster and Tismenetsky (1985). Hence, if $(-S)$ is stable then a unique solution $U_0 = U_0(S)$ to (10) exists which is symmetric and positive definite. Furthermore, relationship (10) has the character of an algebraic Riccati equation with respect to S. Analogously to the well-known properties of Riccati equation (see, e.g. Glad and Ljung, 2000, Lemma 5.1, p. 127) we arrive at the following corollary.

Corollary. *If the matrix gain S is subject to assumption* B3 *then the solution $U_0(S)$ to* (10) *has the following lower bound*

$$U_0(S) \geqslant U_{\min} = \sigma_e Q^{-1/2}(Q^{1/2}R_wQ^{1/2})^{1/2}Q^{-1/2} \tag{11}$$

which coincides with U_{lb} (6) *and is attained for $S = S_{\mathrm{opt}}$ with*

$$S_{\mathrm{opt}} = \sigma_e^{-2}U_{\min}Q = \sigma_e^{-1}Q^{-1/2}(Q^{1/2}R_wQ^{1/2})^{1/2}Q^{1/2}. \tag{12}$$

The corollary above is a special case of Lemma 5.1 in Glad and Ljung (2000). However, it can be easily proved independently. Indeed, from (10) and an evident matrix inequality

$$(\sigma_e^2 SQ^{-1/2} - U_0Q^{1/2})(\sigma_e^2 SQ^{-1/2} - U_0Q^{1/2})^{\mathrm{T}} \geqslant 0,$$

it directly follows that $U_0QU_0 \geqslant \sigma_e^2 R_w$ where equality is attained for

$$\sigma_e^2 SQ^{-1/2} = U_0Q^{1/2}. \tag{13}$$

Consequently (11) and (12) hold true.

Remark. Since both $U_{\min}$ and Q are positive definite matrices, then their product $U_{\min}Q$ has only eigenvalues with positive real part. Hence, the optimal matrix gain S_{opt} (12) meets the stability assumption above.

3.1. Scalar smoothing gain

Now consider the special case of "scalar matrix" gain $S = \rho I_n,\ \rho > 0$. Then Eq. (10) implies $U_0 = U_0(\rho)$ with

$$U_0(\rho) = \tfrac{1}{2}(\rho^{-1}R_w + \rho\sigma_e^2 Q^{-1}). \tag{14}$$

Hence, the optimal ρ in a sense of minimal trace $\mathrm{Tr}\, U_0$ is as follows:

$$\rho_{\mathrm{opt}} = \sigma_e^{-1}\left(\frac{\mathrm{Tr}\, R_w}{\mathrm{Tr}\, Q^{-1}}\right)^{1/2} \tag{15}$$

which ensures the following trace:

$$\mathrm{Tr}\, U_0(\rho_{\mathrm{opt}}) = \min_{\rho > 0} \mathrm{Tr}\, U_0(\rho) = \sigma_e(\mathrm{Tr}\, R_w)^{1/2}(\mathrm{Tr}\, Q^{-1})^{1/2}. \tag{16}$$

This minimum trace cannot be less than $\mathrm{Tr}\, U_{\min} = \mathrm{Tr}\, U_{\mathrm{lb}}$. For the special case of linearly dependent matrices R_w^{-1} and Q, that is

$$R_w^{-1} = \alpha Q \quad \text{for some } \alpha \in R, \tag{17}$$

the traces coincide, i.e.

$$\mathrm{Tr}\, U_0(\rho_{\mathrm{opt}}) = \mathrm{Tr}\, U_{\mathrm{lb}} \tag{18}$$

which means that $\mathrm{Tr}\, U_0(\rho)$ attains its lower bound for $\rho = \rho_{\mathrm{opt}}$ among all possible estimators (in a Gaussian case). Condition (17) is both necessary and sufficient for equality (18). That follows directly from the well-known properties of the corresponding Cauchy–Schwarz inequality for matrix traces (Lancaster & Tismenetsky, 1985), that is $(\mathrm{Tr}\, AB^{\mathrm{T}})^2 \leqslant (\mathrm{Tr}\, AA^{\mathrm{T}})(\mathrm{Tr}\, BB^{\mathrm{T}})$, with equality iff A and B are linearly dependent. This Cauchy–Schwarz inequality might be applied here for $A = Q^{-1/2}$ and $B^{\mathrm{T}} = (Q^{1/2}R_wQ^{1/2})^{1/2}Q^{-1/2}$. Finally, it follows from (15) that under condition (17) the optimal "scalar matrix" gain S_{opt} (12) is reduced to

$$S_{\mathrm{opt}} = \frac{1}{\sigma_e\sqrt{\alpha}} I_n.$$

Remark. If the properties of the regressors $\varphi(t)$ can be chosen freely then it is possible to ensure condition (17), assuming parameter variation R_w being known, by an experiment design. Such a designed experiment would thus give optimal parameter tracking with the simplest algorithm.

3.2. Proof of Theorem 1

We prove an even more general theorem having its own interest. The generalization consists in introducing non-singular matrix gain A into procedure (8), that is

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mu A\varphi(t)(y(t) - \varphi^{\mathrm{T}}(t)\hat{\theta}(t-1)). \tag{19}$$

Hence, it will be proved that the matrix U_0 (5) does not depend on A. This result explains why only a scalar step size is enough for procedure (8), and that we are not able to influence matrix U_0 by a matrix gain A in (19).

Proof. Let the estimates $\hat{\theta}(t)$ be generated by the more general procedure (19), instead of (8). Let the matrix gain A be non-singular and assume that $(-AQ)$ is stable. Denote the related estimation error covariance matrix and its limit by

$$V_t = E(\hat{\theta}(t) - \theta(t))(\hat{\theta}(t) - \theta(t))^{\mathrm{T}}, \quad V = \lim_{t\to\infty} V_t. \tag{20}$$

The limit equation

$$AQV + VQA^{\mathrm{T}} + O(\mu)V = \mu\sigma_e^2 AQA^{\mathrm{T}} + \frac{\gamma^2}{\mu}R_w(1 + O(\mu)) \quad \text{as } \gamma \to +0 \tag{21}$$

follows directly from well-known previous results (see, e.g. Ljung and Gunnarsson (1990)). Therefore, due to assumptions B1, B2 [2]

$$\|V\| = O(\mu) \quad \text{as } \gamma \to +0 \tag{22}$$

which, together with (21), imply the Lyapunov equation

$$AQV + VQA^{\mathrm{T}} = \mu\sigma_e^2 AQA^{\mathrm{T}} + O(\mu^2) + o(\gamma) \quad \text{as } \gamma \to +0. \tag{23}$$

Furthermore, for the cross covariance matrix

$$R_t = E(\hat{\theta}(t) - \theta(t))(\bar{\theta}(t) - \theta(t))^{\mathrm{T}} \tag{24}$$

we obtain from (19), (9) and (1), (2) that

$$R_t = (I_n - \mu AQ)[R_{t-1}(I_n - \gamma S)^{\mathrm{T}} + \gamma V_{t-1}S^{\mathrm{T}} + \gamma^2 R_w]. \tag{25}$$

In order to evaluate the limit $R_\infty = \lim_{t\to\infty} R_t$ we now let $t \to \infty$ and, taking assumptions B1–B3 into account, obtain

$$R_\infty = \gamma\mu^{-1}Q^{-1}A^{-1}VS^{\mathrm{T}} + o(\gamma). \tag{26}$$

In a similar manner, we evaluate U_t (defined by (4)) and U (defined by (3)):

$$U_t = (I_n - \gamma S)U_{t-1}(I_n - \gamma S)^{\mathrm{T}} + \gamma^2 R_w + \gamma^2 SV_{t-1}S^{\mathrm{T}} + \gamma(I_n - \gamma S)R_{t-1}^{\mathrm{T}}S^{\mathrm{T}} + \gamma SR_{t-1}(I_n - \gamma S)^{\mathrm{T}}$$

and, taking (22), (26) as well as B1, B2 into account, we find that

$$SU + US^{\mathrm{T}} = \gamma R_w + R_\infty^{\mathrm{T}} S^{\mathrm{T}} + SR_\infty + o(\gamma). \tag{27}$$

Note that (27) is a Lyapunov equation with respect to U entering linearly. Hence, due to (22) and (26), $\|U\| = O(\gamma)$ as $\gamma \to +0$, and substituting (26) into (27) we obtain

$$\begin{aligned} SU + US^{\mathrm{T}} &= \gamma R_w + \gamma\mu^{-1}(SVA^{-\mathrm{T}}Q^{-1}S^{\mathrm{T}} + SQ^{-1}A^{-1}VS^{\mathrm{T}}) + o(\gamma) \\ &= \gamma R_w + \gamma\mu^{-1}S(AQ)^{-1}(AQV + VQA^{\mathrm{T}})(AQ)^{-\mathrm{T}}S^{\mathrm{T}} + o(\gamma). \end{aligned}$$

Finally, using (23), we arrive at

$$\begin{aligned} SU + US^{\mathrm{T}} &= \gamma R_w + \gamma\sigma_e^2 SQ^{-1}A^{-1}(AQA^{\mathrm{T}} + O(\mu))A^{-\mathrm{T}}Q^{-1}S^{\mathrm{T}} + o(\gamma) \\ &= \gamma R_w + \gamma\sigma_e^2 SQ^{-1}S^{\mathrm{T}} + o(\gamma). \end{aligned} \tag{28}$$

Therefore, the limit matrix U_0 defined by (5) meets Eq. (10) and does not depend on A. Theorem 1 is proved. □

[2] Here and further on we use matrix norm $\|A\| = (\operatorname{Tr} AA^{\mathrm{T}})^{1/2}$ which corresponds to inner product $\langle A, B\rangle = \operatorname{Tr} AB^{\mathrm{T}}$.

4. SLAMS: a more general algorithm

Let us consider the following modification of the parameter tracking algorithm (8) and (9). It contains a natural number m as a parameter.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mu\varphi(t)(y(t) - \varphi^{\mathrm{T}}(t)\hat{\theta}(t-1)), \tag{29a}$$

$$\tilde{\theta}(t) = \frac{1}{m}\sum_{\tau=t-m}^{t-1}\hat{\theta}(\tau), \tag{29b}$$

$$\bar{\theta}(t) = \begin{cases} \bar{\theta}(t-m) - \gamma S(\bar{\theta}(t-m) - \tilde{\theta}(t)), & t = km, \quad k = 1,2,\ldots \\ \bar{\theta}(t-1) - \gamma(\bar{\theta}(t-1) - \hat{\theta}(t-1)), & t \neq km, \quad k = 1,2,\ldots . \end{cases} \tag{29c}$$

Since it is a smoothing algorithm based on the averaged estimates from the LMS procedure, we call it SLAMS. Evidently, this algorithm coincides with (8) and (9), when $m = 1$. However, when $m > 1$, procedure (29a)–(29c) takes less arithmetic calculations per time unit (in a multi-variate case) than (8) and (9). Moreover, it turns out that procedure (29a)–(29c) can ensure the same asymptotic MSE as (8), (9).

Theorem 2. *Assume that assumptions* A1–A5 *and* B1–B3 *hold, and consider the estimates $\bar{\theta}(t)$ generated by algorithm* (29a)–(29c). *Then for any fixed natural number m, the asymptotic error covariance matrix*

$$U^{(m)} = \lim_{t\to\infty} E(\bar{\theta}(t) - \theta(t))(\bar{\theta}(t) - \theta(t))^{\mathrm{T}} \tag{30}$$

is the solution to the equation

$$SU^{(m)} + U^{(m)}S^{\mathrm{T}} = \gamma\left(mR_w + \frac{\sigma_e^2}{m}SQ^{-1}S^{\mathrm{T}}\right) + o(\gamma) \quad \text{as } \gamma \to +0. \tag{31}$$

Hence, the lower bound $U_{\mathrm{lb}}^{(m)}$ to the limiting asymptotic error covariance matrix

$$U_0^{(m)} = \lim_{\gamma\to+0}\gamma^{-1}U^{(m)} \tag{32}$$

coincides with $U_{\mathrm{lb}} = U_{\min}$ (see (6) *and* (11)), *that is*

$$U_0^{(m)}(S) \geqslant U_{\mathrm{lb}}^{(m)} = \sigma_e Q^{-1/2}(Q^{1/2}R_wQ^{1/2})^{1/2}Q^{-1/2}. \tag{33}$$

Therefore, the lower bound to $U_0^{(m)}$ does not depend on m and is attained for $S = S_{\mathrm{opt}}^{(m)}$ with

$$S_{\mathrm{opt}}^{(m)} = m\sigma_e^{-2}U_{\mathrm{lb}}^{(m)}Q = m\sigma_e^{-1}Q^{-1/2}(Q^{1/2}R_wQ^{1/2})^{1/2}Q^{1/2}. \tag{34}$$

The proof of Theorem 2 is analogous to that of Theorem 1. Note that a comparison of Eq. (31) with (28) shows that the modification of the tracking algorithm suggested above corresponds to a simultaneous m times increase in the drift covariance matrix R_w and m times decrease in the variance of observation error σ_e^2 in the right-hand side of Lyapunov equations (10) and (28). Since the optimal gain matrix (34) has balanced the influence of correspondent summands in the right-hand side of (31), this explains that the lower bounds (11) and (33) coincide.

4.1. Proof of Theorem 2

Introduce the estimation errors

$$\hat{\delta}(t) = \hat{\theta}(t) - \theta(t), \quad \tilde{\delta}(t) = \tilde{\theta}(t) - \theta(t),$$
$$\bar{\delta}(t) = \bar{\theta}(t) - \theta(t).$$

As to $\hat{\delta}(t)$, we can use the relations from the proof of Theorem 1. Particularly, relations (20) and (22) imply $E\|\hat{\delta}(t)\|^2 = O(\mu)$. Consequently

$$E\|\tilde{\delta}(t)\|^2 = O(\mu).$$

Consider a subsequence of time instants $t=km$, $k=1,2,\ldots$ and first prove the theorem for the partial limit

$$U^{(m)} = \lim_{k\to\infty} E(\bar{\theta}(km) - \theta(km))(\bar{\theta}(km) - \theta(km))^{\mathrm{T}}. \tag{35}$$

The estimation error $\bar{\delta}(t)$ for $t=km$ is recursively represented as

$$\bar{\delta}(t) = (I_n - \gamma S)\bar{\delta}(t-m) + \gamma S\tilde{\delta}(t) - (I_n - \gamma S)\sum_{\tau=t-m}^{t-1} w(\tau). \tag{36}$$

Note that the first and the last summands in the r.h.s. of (36) are uncorrelated. Furthermore, the correlation between the second and the last summands is evaluated as $O(\gamma^2\mu^{1/2}) = o(\gamma^2)$, since by Cauchy–Schwarz inequality

$$\|E\tilde{\delta}(t)w^{\mathrm{T}}(\tau)\| \leqslant (E\|\tilde{\delta}(t)\|^2)^{1/2}(E\|w(\tau)\|^2)^{1/2} \underset{k\to\infty}{\to} O(\mu^{1/2}\gamma) = o(\gamma).$$

Hence, the covariance matrix (35) meets the equation

$$SU^{(m)} + U^{(m)}S^{\mathrm{T}} + O(\gamma)U^{(m)} = \gamma m R_w + (\tilde{R}_m^{\mathrm{T}}S^{\mathrm{T}} + S\tilde{R}_m)(1 + O(\gamma)) + o(\gamma), \tag{37}$$

where

$$\tilde{R}_m = \lim_{k\to\infty} E\tilde{\delta}(km)\bar{\delta}^{\mathrm{T}}(km - m). \tag{38}$$

We move trivial and rather bulky calculations to the Appendix A which prove that

$$\tilde{R}_m = \lim_{k\to\infty} E\tilde{\delta}(km)\bar{\delta}^{\mathrm{T}}(km) + o(\gamma) = \frac{\gamma}{2m}\sigma_e^2 Q^{-1}S^{\mathrm{T}} + o(\gamma). \tag{39}$$

Substituting (39) into (37), we arrive at (31) for the partial limit considered.

Now, consider subsequence of time instants $t = km + 1$, $k = 1,2,\ldots$; then we have for the estimation error $\bar{\delta}(t)$ the simplier recursive-like equation

$$\bar{\delta}(t) = (1-\gamma)\bar{\delta}(t-1) + \gamma\hat{\delta}(t-1) - w(t), \tag{40}$$

from which follows that

$$\lim_{k\to\infty} E\bar{\delta}(km+1)\bar{\delta}^{\mathrm{T}}(km+1) = U^{(m)} + O(\gamma^2) \tag{41}$$

with $U^{(m)}$ as defined by (35); hence, the partial limit (41) also meet Eq. (31).

Further continuing this finite induction, we obtain

$$\lim_{k\to\infty} E\bar{\delta}(km+s)\bar{\delta}^{\mathrm{T}}(km+s) = U^{(m)} + O(\gamma^2), \quad s = 1,2,\ldots,m-1$$

with $U^{(m)}$ as defined by (35). This proves Eq. (31) for any partial limits of the matrix sequence $\{U(t)\}$.

The rest of the theorem is proved in completely the same manner as that in the proof of Theorem 1. This complete the proof. □

5. Conclusion

From the obtained results it follows that the optimal limiting asymptotic error covariance matrix $U_{\min}$ (11) for the SLAMS algorithm (29) coincides with the lower bound U_{lb} (6). Thus, Theorem 2 and the lower bound (6) imply that under Gaussian distributions of $e(t)$ and $w(t)$ the algorithm (29) with optimal matrix gain $S = S_{\mathrm{opt}}$ (12) delivers asymptotically optimal estimates among all the possible estimators.

An interesting theoretical aspect of this is that it is possible to achieve asymptotically optimal accuracy with an algorithm that is considerably simpler that the optimal Kalman-filter-based algorithm. This might also prove useful in certain practical applications.

It might be seen as a paradox that the result is independent of the integer m, which also governs the algorithm complexity. One should bear in mind that the result is asymptotic in $\gamma \to +0$. For fixed, non-zero $\gamma > 0$, there will be an upper limit of m for which the limit expression is a good approximation of the true covariance matrix.

Acknowledgements

This work was supported by the Swedish Research Council under the contract on System Modeling. The authors would also like to thank Prof. J.P. LeBlanc (Luleå) for his creative acronyms.

Appendix A

Proof of (39). First, note that we may use relations (20)–(22) from the proof of Theorem 1, since procedure (19) coincides with that of (29a) under $A = I_n$. Thus, putting $A = I_n$ and introducing the tracking error for $\hat{\theta}(t)$ as $\hat{\delta}(t) = \hat{\theta}(t) - \theta(t)$ we obtain the solution to (23)

$$V = \lim_{t\to\infty} E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(t) = \frac{\mu}{2}\sigma_e^2 I_n(1 + O(\mu)) + o(\gamma) \tag{A.1}$$

and consequently

$$\lim_{t\to\infty} \|E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(t)\| = O(\mu) \quad \text{as } \gamma \to +0. \tag{A.2}$$

Note that (A.2) can be extended to

$$\lim_{t\to\infty} \|E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(t+\ell)\| = O(\mu) \quad \text{as } \gamma \to +0 \quad \forall \ell = 1,\dots,m. \tag{A.3}$$

Indeed, for $\ell = 1$ we obtain the equation

$$E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(t+1) = E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(t)(I_n - \mu Q) \tag{A.4}$$

which follows from the straightforward recursive equation

$$\hat{\delta}(t) = (I_n - \mu\varphi(t)\varphi^{\mathrm{T}}(t))(\hat{\delta}(t-1) - w(t)) + \mu\varphi(t)e(t). \tag{A.5}$$

Hence, by induction relation (A.3) follows from (A.4) and (A.2).

Furthermore, from (A.3) and (29b) the result

$$\lim_{t\to\infty} \|E\tilde{\delta}(t)\tilde{\delta}^{\mathrm{T}}(t)\| = O(\mu) \quad \text{as } \mu \to +0, \ \gamma/\mu \to +0 \tag{A.6}$$

follows directly for the second tracking error (i.e., the one for $\tilde{\theta}(t)$)

$$\tilde{\delta}(t) = \tilde{\theta}(t) - \theta(t) = \frac{1}{m}\sum_{\tau=t-m}^{t-1}\left(\hat{\delta}(\tau) - \sum_{s=\tau}^{t-1} w(s+1)\right) \tag{A.7}$$

since for each s and τ under consideration due to (A.2) by Cauchy–Schwarz inequality

$$\|E\hat{\delta}(\tau)w^{\mathrm{T}}(s+1)\| \leqslant (E\|\hat{\delta}(\tau)\|^2)^{1/2}(E\|w(s+1)\|^2)^{1/2} \underset{t\to\infty}{\to} O(\mu^{1/2}\gamma) = o(\mu).$$

Now we are ready to directly evaluate (38), that is

$$\tilde{R}_m = \lim_{k\to\infty} E\tilde{\delta}(km)\tilde{\delta}^{\mathrm{T}}(km - m). \tag{A.8}$$

Taking (A.7) into account, we obtain for $t = km$

$$\begin{aligned} &E\tilde{\delta}(t)\tilde{\delta}^{\mathrm{T}}(t-m) \\ &= \frac{1}{m}\sum_{\tau=t-m}^{t-1} E\left(\hat{\delta}(\tau) - \sum_{s=\tau}^{t-1} w(s+1)\right)\tilde{\delta}^{\mathrm{T}}(t-m) \\ &\underset{t\to\infty}{\to} (1 + O(\mu))\lim_{k\to\infty} E\hat{\delta}(km)\tilde{\delta}^{\mathrm{T}}(km) \end{aligned} \tag{A.9}$$

since due to (A.5) we obtain for t, τ and s under consideration

$$\begin{aligned} E\hat{\delta}(\tau)\tilde{\delta}^{\mathrm{T}}(t-m) &= (I_n - \mu Q)E\hat{\delta}(\tau-1)\tilde{\delta}^{\mathrm{T}}(t-m) = \dots \\ &= (I_n - \mu Q)^{\tau-t+m}E\hat{\delta}(t-m)\tilde{\delta}^{\mathrm{T}}(t-m) \\ &\underset{t\to\infty}{\to} (1 + O(\mu))\lim_{k\to\infty} E\hat{\delta}(km)\tilde{\delta}^{\mathrm{T}}(km) \end{aligned}$$

and since $Ew(s+1)\tilde{\delta}^{\mathrm{T}}(t-m) = 0$. Thus, we have to evaluate

$$\hat{R}_m = \lim_{k\to\infty} E\hat{\delta}(km)\tilde{\delta}^{\mathrm{T}}(km). \tag{A.10}$$

Denote

$$\tilde{w}(t) = \sum_{\tau=t-m}^{t-1} w(\tau). \tag{A.11}$$

Due to (36) and (A.5), we obtain for $t = km$

$$\begin{aligned} &E\hat{\delta}(t)\tilde{\delta}^{\mathrm{T}}(t) \\ &= E\hat{\delta}(t)[(I_n - \gamma S)(\tilde{\delta}(t-m) - \tilde{w}(t)) + \gamma S\tilde{\delta}(t)]^{\mathrm{T}} \end{aligned} \tag{A.12}$$

$$\begin{aligned} &= E\hat{\delta}(t)(\tilde{\delta}(t-m) - \tilde{w}(t))^{\mathrm{T}}(I_n - \gamma S)^{\mathrm{T}} \\ &\quad + \gamma E\hat{\delta}(t)\tilde{\delta}^{\mathrm{T}}(t)S^{\mathrm{T}}. \end{aligned} \tag{A.13}$$

Consequently applying (A.5) to the first expectation term of (A.13) we obtain

$$\begin{aligned} E\hat{\delta}(t)\tilde{\delta}^{\mathrm{T}}(t-m) &= (I_n - \mu Q)E\hat{\delta}(t-1)\tilde{\delta}^{\mathrm{T}}(t-m) = \dots \\ &= (I_n - \mu Q)^m E\hat{\delta}(t-m)\tilde{\delta}^{\mathrm{T}}(t-m) \end{aligned}$$

and simlarly, due to (A.5) and (A.11),

$$E\hat{\delta}(t)\tilde{w}^{\mathrm{T}}(t) = \sum_{\tau=t-m}^{t-1} E\hat{\delta}(t)w^{\mathrm{T}}(\tau) \tag{A.14}$$

$$\begin{aligned} &= \sum_{\tau=t-m}^{t-1} (I_n - \mu Q)^{t-\tau}E\hat{\delta}(\tau)w^{\mathrm{T}}(\tau) \\ &= -\gamma^2\sum_{\tau=t-m}^{t-1}(I_n - \mu Q)^{t-\tau+1}R_w \\ &= O(\gamma^2). \end{aligned} \tag{A.15}$$

At last, for the second expectation term in (A.13) which turns out to be of the main order $O(\mu)$ we obtain (reminding that we consider $t = km$)

$$E\hat{\delta}(t)\tilde{\delta}^{\mathrm{T}}(t) = \frac{1}{m}\sum_{\tau=t-m}^{t-1} E\hat{\delta}(t)\left(\hat{\delta}(\tau) - \sum_{s=\tau}^{t-1} w(s+1)\right)^{\mathrm{T}} \tag{A.16}$$

$$\begin{aligned} &= \frac{1}{m}\sum_{\tau=t-m}^{t-1} E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(\tau) \\ &\quad - \frac{1}{m}\sum_{\tau=t-m}^{t-1}\sum_{s=\tau}^{t-1} E\hat{\delta}(t)w^{\mathrm{T}}(s+1). \end{aligned} \tag{A.17}$$

Here, for each $\tau, s = t - m, \ldots, t - 1$, we obtain from (A.1) and (A.4)

$$E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(\tau) = (I_n - \mu Q)^{t-\tau} E\hat{\delta}(\tau)\hat{\delta}^{\mathrm{T}}(\tau) \underset{k\to\infty}{\rightarrow} \frac{\mu}{2}\sigma_e^2 I_n(1 + O(\mu)) + o(\gamma)$$

and (similarly to (A.14) and (A.15)) $\|E\hat{\delta}(t)w^{\mathrm{T}}(s+1)\| \underset{k\to\infty}{\rightarrow} O(\gamma^2)$. Hence, from (A.16) and (A.17) follow

$$E\hat{\delta}(t)\hat{\delta}^{\mathrm{T}}(t) \underset{k\to\infty}{\rightarrow} \frac{\mu}{2}\sigma_e^2 I_n + O(\mu^2) + o(\gamma). \tag{A.18}$$

Substituting (A.18) into (A.12) and (A.13), then tending $k \to \infty$, and taking denotion (A.10) into account, we arrive at the following equation:

$$\hat{R}_m = (I_m - \mu Q)^m \hat{R}_m + \gamma\left(\frac{\mu}{2}\sigma_e^2 I_n + O(\mu^2) + o(\gamma)\right) S^{\mathrm{T}} + O(\gamma^2). \tag{A.19}$$

Since $(I_m - \mu Q)^m = I_m - \mu m Q + O(\mu^2)$ as $\mu \to +0$, then the asymptotic solution to Eq. (A.19) is as follows:

$$\hat{R}_m = \frac{\gamma}{2m}\sigma_e^2 Q^{-1} S^{\mathrm{T}} + o(\gamma). \tag{A.20}$$

Thus, combining (38), (A.9), (A.10) and (A.20) we arrive at (39). □

References

Glad, T., & Ljung, L. (2000). *Control theory. Multivariable and nonlinear methods*. London, New York: Taylor & Francis.

Kushner, H. J., & Yang, J. (1993). Stochastic approximation with averaging of the iterates: Optimal asymptotic rate of convergence for general processes. *SIAM Journal of Control and Optimization*, *31*(4), 1045–1062.

Lancaster, P., & Tismenetsky, M. (1985). *The theory of matrices* (2nd ed.). Boston, San Diego: Academic Press.

Ljung, L. (2001). Recursive least-squares and accelerated convergence in stochastic approximation schemes. *International Journal of Adaptive Control and Signal Processing*, *15*(2), 169–178.

Ljung, L., & Gunnarsson, S. (1990). Adaptive tracking in system identification—a survey. *Automatica*, *26*(1), 7–22.

Ljung, L., & Söderström, T. (1983). *Theory and practice of recursive identification*. Cambridge, MA: MIT Press.

Nazin, A. V., & Yuditskii, A. B. (1991). Optimal and robust estimation of slowly drifting parameters in linear-regression. *Automation and Remote Control*, *52*(6, Part 1), 798–807.

Polyak, B. T. (1990). New method of stochastic approximation type. *Automation and Remote Control*, *51*(7, Part 2), 937–946.

Polyak, B. T., & Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, *30*(4), 838–855.

Ruppert, D. (1988). *Efficient estimations from a slowly convergent Robbins–Monro process*. Technical Report No. 781, Cornell University.

Widrow, B., & Stearns, S. (1985). *Adaptive signal processing*. Englewood-Cliffs, NJ: Prentice-Hall.

Alexander Nazin received the M.Sc. degree from the Moscow Physical-Technical Institute (MPhTI), Moscow, USSR, in 1974, and the Ph.D. degree from the Institute of Control Sciences, Russian Academy of Sciences, Moscow, in 1979, both in technical cybernetics. In 1995, he received the Sc.D. degree from the Institute of Control Sciences, in physics and mathematics. Since 1977, he is with the Institute of Control Sciences, currently holding position of Leading Scientific Researcher. Since 1982, he held a part-time positions of Assistant, Associate, and Full Professor at the Department of Theoretical Mechanics, MPhTI. Since 1997, he also held a part-time position of Full Professor at the Department of Probability Theory, Moscow Aviation Institute. His research interests are in the areas of estimation, identification, signal processing, stochastic optimization, and adaptive stochastic control. He has published one book and more than 60 papers.

Lennart Ljung received his Ph.D. in Automatic Control from Lund Institute of Technology in 1974. Since 1976 he is Professor of the chair of Automatic Control in Linkoping, Sweden, and is currently Director of the Competence Center "Information Systems for Industrial Control and Supervision" (ISIS). He has held visiting positions at Stanford and MIT and has written several books on System Identification and Estimation. He is an IEEE Fellow and an IFAC Advisor as well as a member of the Royal Swedish Academy of Sciences (KVA), a member of the Royal Swedish Academy of Engineering Sciences (IVA), and an Honorary Member of the Hungarian Academy of Engineering. He has received honorary doctorates from the Baltic State Technical University in St. Petersburg, and from Uppsala University. In 2002 he received the Quazza Medal from IFAC.