15th Triennial World Congress, Barcelona, Spain

ASYMPTOTICALLY OPTIMAL SMOOTHING OF AVERAGED LMS FOR REGRESSION PARAMETER TRACKING

Alexander V. Nazin [**] **and Lennart Ljung** [*]

[*] *Division of Automatic Control, Linköping University, SE-58183 Linköping, Sweden. E-mail:* ljung@isy.liu.se
[**] *Institute of Control Sciences, Profsoyuznaya str., 65, 117997 Moscow, Russia. E-mail:* nazine@ipu.rssi.ru

Abstract: The sequence of estimates formed by the LMS algorithm for a standard linear regression estimation problem is considered. 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 the 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. *Copyright © 2002 IFAC*

Keywords: Regression; Parameter estimation; Random walk; Recursive algorithms; Tracking; Smoothing; Mean-square error; Asymptotic properties

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; Ljung and Gunnarsson, 1990).

One of the most common methods is the least mean squares algorithm, LMS, (Widrow and 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 (1990) and Ruppert (1988), is that this optimal accuracy can asymptotically also be obtained by a simple averaging of the LMS-estimates. See (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, a more general post-processing of the LMS-estimates is considered, obtained from a constant gain, unnormalized LMS-method. The general version of this algorithm is called 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

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

and finally smoothing these by a simple exponential smoother, applying a direction correction every m:th sample:

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

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 θ, 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). It is shown that by the choice of S and γ, the same asymptotic covariance can be obtained 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, the parameter tracking problem is formulated and the basic assumptions are stated. A special case of SLAMS is treated in Sections 3, 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

$$y(t) = \theta^T(t)\varphi(t) + e(t) \quad (3)$$

$$\theta(t) = \theta(t-1) + w(t) \quad (4)$$

where $e(t) \in R$ and $w(t) \in R^n$ stand for observation error and parameter change respectively. Due to the following assumptions, the equation (4) 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 unbiased variable with positive definite covariance matrix, i.e. $Ew(t) = 0$ and $Ew(t)w^T(t) = \gamma^2 R_w > 0$, where a priori known $\gamma > 0$ represent small parameter of the problem under consideration.
- A4 The regressor covariance matrix $E\varphi(t)\varphi^T(t) = Q$ is non-singular; 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 (3) with a minimal asymptotic error covariance matrix U:

$$U = \lim_{t\to\infty} U_t \quad (5)$$

where

$$U_t = E(\bar{\theta}(t) - \theta(t))(\bar{\theta}(t) - \theta(t))^T \quad (6)$$

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

$$U = \gamma U_0 + o(\gamma) \quad \text{as} \quad \gamma \to +0 \quad (7)$$

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 $w(t)$

$$U_0 \geq U_{lb} \qquad \text{for any parameter estimator}, \quad (8)$$

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

$$U_{lb} = \sigma_e Q^{-1/2}(Q^{1/2}R_w Q^{1/2})^{1/2}Q^{-1/2} \quad (9)$$

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

Let us further call the matrix U_0 in (7) the limiting asymptotic error covariance matrix, since

$$U_0 = \lim_{\gamma\to+0} \gamma^{-1} U \quad (10)$$

It is studied below how the matrix U_0 depends on the parameters both of the problem and the algorithm described in the following section. This matrix is then minimized over the design parameters in the algorithm.

3. PARAMETER TRACKING BY SMOOTHED LMS

Now, the aim is to study the following recursive constant gain SA-like procedure, which is called SLMS, (Smoothed LMS):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mu\varphi(t)(y(t) - \varphi^T(t)\hat{\theta}(t-1)) \quad (11)$$

$$\bar{\theta}(t) = \bar{\theta}(t-1) - \gamma S(\bar{\theta}(t-1) - \hat{\theta}(t-1)) \quad (12)$$

Here $\mu > 0$ is a scalar step size while S represents an $n\times n$ - matrix gain. The relation (11) is exactly the constant (scalar) gain SA-algorithm (LMS), while recursive procedure (12) 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 subsection 3.1 below). In that case there are no matrix calculations in the algorithm (11), (12), which makes 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–B2, stochastic stability of equations (11), (12) (in mean-square sense) is obviously ensured (for sufficiently small γ). This implies the existence of limit in (5).

Theorem 1. Let the assumptions A1–A5 and B1–B3 hold, and consider the estimates $\overline{\theta}(t)$ generated by the algorithm (11), (12). Then the limiting asymptotic error covariance matrix U_0, defined by (7), is the solution to the equation

$$SU_0 + U_0 S^T = R_w + \sigma_e^2 S Q^{-1} S^T \tag{13}$$

Remark. The relationship (13) 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 (13) exists which is symmetric and positive definite. Furthermore, the relationship (13) might be considered as an algebraic Riccati equation with respect to S. Due to well known properties of Riccati equation the following consequence holds true.

Corollary. If the matrix gain S is subject to assumption B3 then the solution $U_0(S)$ to (13) has the following lower bound

$$U_0(S) \geq U_{\min} = \sigma_e Q^{-1/2}(Q^{1/2} R_w Q^{1/2})^{1/2} Q^{-1/2} \tag{14}$$

which coincides with U_{lb} (9) and is attained for $S = S_{opt}$ with

$$\begin{aligned} S_{opt} &= \sigma_e^{-2} U_{\min} Q \\ &= \sigma_e^{-1} Q^{-1/2}(Q^{1/2} R_w Q^{1/2})^{1/2} Q^{1/2} \end{aligned} \tag{15}$$

The corollary above is a special case of Lemma 5.1 in (Ljung and Glad, 2000). However, it can be easily proved independently. Indeed, from (13) and an evident matrix inequality

$$(\sigma_e^2 S Q^{-1/2} - U_0 Q^{1/2})(\sigma_e^2 S Q^{-1/2} - U_0 Q^{1/2})^T \geq 0 \tag{16}$$

it directly follows that $U_0 Q U_0 \geq \sigma_e^2 R_w$ where equality is attained for

$$\sigma_e^2 S Q^{-1/2} = U_0 Q^{1/2} \tag{17}$$

Consequently (14) and (15) hold true.

Remark. Since both $U_{\min}$ and Q are positive definite matrices, then its product $U_{\min} Q$ has only real positive eigenvalues. Hence, the optimal matrix gain S_{opt} (15) 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 equation (13) implies $U_0 = U_0(\rho)$ with

$$U_0(\rho) = \frac{1}{2}\left(\rho^{-1} R_w + \rho \sigma_e^2 Q^{-1}\right) \tag{18}$$

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

$$\rho_{opt} = \sigma_e^{-1}\left(\frac{\mathrm{Tr}\,R_w}{\mathrm{Tr}\,Q^{-1}}\right)^{1/2} \tag{19}$$

which ensures the following trace

$$\begin{aligned} \mathrm{Tr}\,U_0(\rho_{opt}) &= \min_{\rho>0} \mathrm{Tr}\,U_0(\rho) \\ &= \sigma_e \left(\mathrm{Tr}\,R_w\right)^{1/2}\left(\mathrm{Tr}\,Q^{-1}\right)^{1/2} \end{aligned} \tag{20}$$

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

$$R_w^{-1} = \alpha Q \quad \text{for some } \alpha \in R, \tag{21}$$

the traces coincide, i.e.

$$\mathrm{Tr}\,U_0(\rho_{opt}) = \mathrm{Tr}\,U_{lb} \tag{22}$$

which means that $\mathrm{Tr}\,U_0(\rho)$ attains its lower bound for $\rho = \rho_{opt}$ among all possible estimators (in a Gaussian case). The condition (21) is both necessary and sufficient for the equality (22). That follows directly from the well known properties of the corresponding Cauchy–Schwarz inequality for matrix traces (Lancaster and Tismenetsky, 1985), that is $\left(\mathrm{Tr}\,AB^T\right)^2 \leq \left(\mathrm{Tr}\,AA^T\right)\left(\mathrm{Tr}\,BB^T\right)$, with equality iff A and B are linearly dependent. This Cauchy–Schwarz inequality might be applied here for $A = Q^{-1/2}$ and $B^T = (Q^{1/2} R_w Q^{1/2})^{1/2} Q^{-1/2}$. Finally, it follows from (19) that under condition (21) the optimal “scalar matrix” gain S_{opt} (15) is reduced to

$$S_{opt} = \frac{1}{\sigma_e \sqrt{\alpha}} I_n \tag{23}$$

Remark. If the properties of the regressors $\varphi(t)$ can be chosen freely then it is possible to ensure the condition (21), 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*

Below is the proof of even more general theorem having its own interest. The generalization consists in introducing non-singular matrix gain A into procedure (11), that is

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mu A \varphi(t)(y(t) - \varphi^T(t)\hat{\theta}(t-1)) \tag{24}$$

Hence, it will be proved that the matrix U_0 (7) does not depend on A. This result explains why only a scalar step size is enough for the procedure (11), and that matrix U_0 can not be influenced by a matrix gain A in (24).

Proof. Let the estimates $\hat{\theta}(t)$ be generated by the more general procedure (24), instead of (11). 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

$$\begin{aligned} V_t &= E(\hat{\theta}(t)-\theta(t))(\hat{\theta}(t)-\theta(t))^T \\ V &= \lim_{t\to\infty} V_t \end{aligned} \tag{25}$$

The limit equation (and asymptotics as $\gamma \to +0$)

$$\begin{aligned} &AQV + VQA^T + O(\mu)V \\ &= \mu\sigma_e^2 AQA^T + \frac{\gamma^2}{\mu} R_w(1+O(\mu)) \end{aligned} \tag{26}$$

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

$$||V|| = O(\mu) \quad \text{as } \gamma \to +0 \tag{27}$$

which, together with (26), imply Lyapunov equation (as $\gamma \to +0$)

$$AQV + VQA^T = \mu\sigma_e^2 AQA^T + O(\mu^2) + o(\gamma) \tag{28}$$

Here and further on matrix norm $||A|| = \left(\mathrm{Tr} AA^T\right)^{1/2}$ is used which corresponds to the inner product $\langle A, B\rangle = \mathrm{Tr} AB^T$. Furthermore, for the cross covariance matrix

$$R_t = E(\hat{\theta}(t)-\theta(t))(\bar{\theta}(t)-\theta(t))^T \tag{29}$$

it follows from (24), (12) and (3), (4) that

$$\begin{aligned} R_t = (I_n - \mu AQ)\,&[R_{t-1}(I_n-\gamma S)^T \\ &+ \gamma V_{t-1}S^T + \gamma^2 R_w] \end{aligned} \tag{30}$$

In order to evaluate the limit $R_\infty = \lim_{t\to\infty} R_t$ by letting $t \to \infty$, assumptions B1, B2 are taken into account, from what follows

$$R_\infty = \gamma\mu^{-1}Q^{-1}A^{-1}VS^T + o(\gamma) \tag{31}$$

In a similar manner, evaluation for U_t (defined by (6)) and U (defined by (5)) follows:

$$\begin{aligned} U_t = &(I_n-\gamma S)U_{t-1}(I_n-\gamma S)^T + \gamma^2 R_w \\ &+ \gamma^2 SV_{t-1}S^T + \gamma(I_n-\gamma S)R_{t-1}^T S^T \\ &+ \gamma SR_{t-1}(I_n-\gamma S)^T \end{aligned} \tag{32}$$

and from (27), (31) as well as B1, B2 it follows that

$$SU + US^T = \gamma R_w + R_\infty^T S^T + SR_\infty + o(\gamma) \tag{33}$$

Note that (33) is a Lyapunov equation with respect to U entering linearly. Hence, due to (27) and (31), $||U|| = O(\gamma)$ as $\gamma \to +0$, and substitution (31) into (33) gives

$$\begin{aligned} SU + US^T &= \gamma R_w + \gamma\mu^{-1}\left(SVA^{-T}Q^{-1}S^T + \right. \\ &\quad \left. + SQ^{-1}A^{-1}VS^T\right) + o(\gamma) \\ &= \gamma R_w + \gamma\mu^{-1}S(AQ)^{-1}(AQV + \\ &\quad + VQA^T)(AQ)^{-T}S^T + o(\gamma) \end{aligned} \tag{34}$$

Finally, it follows from (28) that

$$\begin{aligned} SU + US^T &= \gamma R_w + \gamma\sigma_e^2 SQ^{-1}A^{-1}(AQA^T + \\ &\quad + O(\mu))A^{-T}Q^{-1}S^T + o(\gamma) \\ &= \gamma R_w + \gamma\sigma_e^2 SQ^{-1}S^T + o(\gamma) \end{aligned} \tag{35}$$

Therefore, the limit matrix U_0 defined by (7) meet the equation (13) and does not depend on A. Theorem 1 is proved. □

4. SLAMS: A MORE GENERAL ALGORITHM

Let us consider the following modification of the parameter tracking algorithm (11), (12). It contains a natural number m as a parameter.

$$\begin{aligned} \hat{\theta}(t) = &\hat{\theta}(t-1) + \\ &+ \mu\varphi(t)(y(t) - \varphi^T(t)\hat{\theta}(t-1)) \end{aligned} \tag{36a}$$

$$\tilde{\theta}(t) = \frac{1}{m}\sum_{\tau=t-m}^{t-1}\hat{\theta}(\tau) \tag{36b}$$

$$\bar{\theta}(t) = \begin{cases} \bar{\theta}(t-m) - \gamma S(\bar{\theta}(t-m) - \tilde{\theta}(t)), \\ \qquad t = km,\ k = 1,2,\dots \\ \bar{\theta}(t-1) - \gamma(\bar{\theta}(t-1) - \hat{\theta}(t-1)), \\ \qquad t \neq km,\ k = 1,2,\dots \end{cases} \tag{36c}$$

Since it is a Smoothing algorithm based on the Averaged estimates from the LMS procedure, it is called SLAMS. Evidently, this algorithm coincides with (11), (12), when $m = 1$. However, when $m > 1$, the procedure (36a) – (36c) takes less arithmetic calculations per time unit (in a multi-variate case) than (11), (12). Moreover, it turns out that the procedure (36a) – (36c) can ensure the same asymptotic MSE as (11), (12).

Theorem 2. Assume that the assumptions A1–A5 and B1–B3 hold, and consider the estimates $\bar{\theta}(t)$ generated by the algorithm (36a) – (36c). Then for any fixed natural number m the asymptotic error covariance matrix

$$U^{(m)} = \lim_{t\to\infty} E(\bar{\theta}(t)-\theta(t))(\bar{\theta}(t)-\theta(t))^T \tag{37}$$

is the solution to the equation

$$\begin{aligned} SU^{(m)} + U^{(m)}S^T \;=\; &\gamma\left(mR_w + \frac{\sigma_e^2}{m}SQ^{-1}S^T\right) \\ &+ o(\gamma) \quad \text{as } \gamma \to +0 \end{aligned} \tag{38}$$

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

$$U_0^{(m)} = \lim_{\gamma\to+0}\gamma^{-1}U^{(m)} \tag{39}$$

coincides with $U_{lb} = U_{\min}$ (see (9), (14)), that is

$$U_0^{(m)}(S) \geq U_{lb}^{(m)} \\ = \sigma_e Q^{-1/2}(Q^{1/2}R_w Q^{1/2})^{1/2}Q^{-1/2} \tag{40}$$

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

$$S_{opt}^{(m)} = m\sigma_e^{-2}U_{lb}^{(m)}Q \\ = m\sigma_e^{-1}Q^{-1/2}(Q^{1/2}R_w Q^{1/2})^{1/2}Q^{1/2} \tag{41}$$

The proof of Theorem 2 is analogous to that of Theorem 1. Note that a comparison of the equation (38) with (35) 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 (13), (35). Since the optimal gain matrix (41) is balanced the influence of correspondent summands in the right hand side (38), this explains that the lower bounds (14) and (40) coincide.

4.1 *Proof of Theorem 2*

Introduce the estimation errors

$$\hat{\delta}(t) = \hat{\theta}(t) - \theta(t) \tag{42}$$
$$\widetilde{\delta}(t) = \widetilde{\theta}(t) - \theta(t) \tag{43}$$
$$\bar{\delta}(t) = \bar{\theta}(t) - \theta(t) \tag{44}$$

As to $\hat{\delta}(t)$, the relations from the proof of Theorem 1 can be used. Particularly, relations (25), (27) imply $E\|\hat{\delta}(t)\|^2 = O(\mu)$. Consequently

$$E\|\widetilde{\delta}(t)\|^2 = O(\mu) \tag{45}$$

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

$$U^{(m)} = \lim_{k\to\infty} E(\bar{\theta}(km) - \theta(km))(\bar{\theta}(km) - \theta(km))^T \tag{46}$$

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

$$\bar{\delta}(t) = (I_n - \gamma S)\bar{\delta}(t-m) + \gamma S\widetilde{\delta}(t) \\ - (I_n - \gamma S)\sum_{\tau=t-m}^{t-1} w(\tau) \tag{47}$$

Note that the first and the last summands in the r.h.s. of (47) 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

$$\|E\widetilde{\delta}(t)w^T(\tau)\| \leq (E\|\widetilde{\delta}(t)\|^2)^{1/2}(E\|w(\tau)\|^2)^{1/2} \\ \underset{k\to\infty}{\longrightarrow} O(\mu^{1/2}\gamma) = o(\gamma) \tag{48}$$

Hence, the covariance matrix (46) meets the equation

$$SU^{(m)} + U^{(m)}S^T + O(\gamma)U^{(m)} \\ = \gamma m R_w + (\widetilde{R}_m^T S^T + S\widetilde{R}_m)(1 + O(\gamma)) + o(\gamma) \tag{49}$$

where

$$\widetilde{R}_m = \lim_{k\to\infty} E\widetilde{\delta}(km)\bar{\delta}^T(km - m) \tag{50}$$

Some calculations (see (Nazin and Ljung, to appear)) prove that

$$\widetilde{R}_m = \lim_{k\to\infty} E\widetilde{\delta}(km)\bar{\delta}^T(km) + o(\gamma) \\ = \frac{\gamma}{2m}\sigma_e^2 Q^{-1}S^T + o(\gamma) \tag{51}$$

Substitution (51) into (49) leads to (38) for the partial limit considered.

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

$$\bar{\delta}(t) = (1-\gamma)\bar{\delta}(t-1) + \gamma\hat{\delta}(t-1) - w(t)\,, \tag{52}$$

from which follows that

$$\lim_{k\to\infty} E\bar{\delta}(km+1)\bar{\delta}^T(km+1) = U^{(m)} + O(\gamma^2) \tag{53}$$

with $U^{(m)}$ as defined by (46); hence, the partial limit (53) also meet the equation (38).

Consequently, by this finite induction,

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

where $U^{(m)}$ is defined by (46). This proves the equation (38) 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 completes the proof. □

5. CONCLUSION

From the obtained results it follows that the optimal limiting asymptotic error covariance matrix $U_{\min}$ (14) for the SLAMS algorithm (36) coincides with the lower bound U_{lb} (9). Thus, Theorem 1 and the lower bound (9) imply that under Gaussian distributions of $e(t)$ and $w(t)$ the algorithm (11), (12) with optimal matrix gain $S = S_{opt}$ (15) delivers asymptotically optimal estimates among all possible estimators.

An interesting theoretic 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.

6. REFERENCES

Glad, T. and L. Ljung (2000). *Control Theory. Multivariable and Nonlinear Methods*. Taylor & Francis, London and New York.

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

Lancaster, P. and M. Tismenetsky (1985). *The Theory of Matrices, 2nd ed*. Academic Press, Inc., Boston, San Diego.

Ljung, L. (2001). Recursive least-squares and accelerated convergence in stochastic approximation schemes. *Int. J. Adapt. Control Signal Process*, **15**, 169–178.

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

Ljung, L. and T. Söderström (1983). *Theory and Practice of Recursive Identification*. MIT press, Cambridge, Mass.

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

Nazin, A.V. and L. Ljung (2002). Asymptotically optimal smoothing of averaged LMS estimates for regression parameter tracking. *Automatica*. Accepted for publication.

Polyak, B. T. and A. B. Juditsky (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, **30**, 838–855.

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

Ruppert, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process. Techn. Rep. No. 781, Cornell Univ.

Widrow, B. and S. Stearns (1985). *Adaptive Signal Processing*. Prentice-Hall, Englewood-Cliffs.