Asymptotic properties of the OBE parameter estimation algorithm under diffuse initialization

B A Skorohod

1 Sevastopol State University, 33 Universitetskaya ave., Sevastopol, 299053, Russia
E-mail: boris.skorohod@mail.ru

Abstract. The parameters estimation problem of autoregressive with exogenous inputs models is considered. Our approach is based on the study of the optimal bounding ellipsoid (OBE) recurrent algorithm properties. It is assumed that the shape matrix of the initial ellipsoid in the OBE algorithm is proportional to a large positive parameter. Asymptotic expansions of the relations that describe the OBE algorithm are obtained. The limit recursive algorithm (diffuse) not depending on a large parameter which may lead the OBE algorithm to a divergence is proposed and explored. The robust properties of the proposed algorithm are illustrated by a numerical example. By simulation, it is shown that it may provide better the tracking ability compared to the diffuse least-squares (RLS) algorithm and the diffuse RLS algorithm with the sliding window.

1. Introduction
Problems related to parameters estimation of autoregressive with exogenous inputs (ARX) models arise in various applications including in particular identification systems, signals processing and control systems. Traditional approaches to their solution depend on the used assumptions regarding the nature of the disturbances (noises). Within the framework of statistical interpretation, the recursive least-squares (RLS) algorithm is commonly used to provide the high convergence rate to optimal values of the parameters model and robustness when processing data in a changing environment [1]. However in many cases, the use of statistics is not entirely justified and may lead to erroneous decisions. In such situations, the approach based on optimal bounding ellipsoid (OBE) recurrent algorithms can be a good alternative [2-5]. Within this approach, it is assumed that noises are unknown but bounded by specified bounds. As opposed to the RLS algorithm that looks for point estimates of unknown parameters, the OBE algorithms allow us to obtain them in such a way that they are only compatible with the model assumptions. More exactly, the important advantage of the OBE algorithms compared to the RLS algorithm is that they update the estimate when the new data can improve its.

The OBE algorithms produce a sequence of ellipsoids (analogs to covariance matrices) in some time instants characterized by their centers that are used as estimates of unknown parameters and sizes. The difference between all known the OBE algorithms is determined primarily by the choice of the optimality criterion by which these ellipsoids are determined. In this paper we will interested in studying of the ODE algorithm properties developed in [5] in the absence of a priori information about unknown parameters of the ARX models. The obtained results are an extension of the approach to the construction of diffuse RLS algorithms proposed in [6] on bounding ellipsoid estimation. Taking into account that the ODE algorithm is nonlinear in contrast to the RLS algorithm and their structures differ significantly, the corresponding generalization is not obvious. In this work, it is assumed that the
shape matrix of the initial ellipsoid in the OBE algorithm is proportional to a large positive parameter. Asymptotic expansions of the relations that describe the OBE algorithm are obtained. The limit recursive algorithm (diffuse) not depending on a large parameter which may lead the OBE algorithm to a divergence is proposed and explored.

2. Problem statement

Consider the ARX model

\[
y_t = a_1 y_{t-1} + \ldots + a_m y_{t-m} + b_1 u_{t-1} + \ldots + b_k u_{t-k} + \xi_t, \quad t = 1, 2, \ldots, N,
\]

where \( y_t, u_t \in \mathbb{R}^1 \) are the measurable output and input, respectively, \( a_i, b_i, i = 1, 2, \ldots, k \) are unknown parameters and the noise \( \xi_t \) satisfies the restriction

\[
\xi_t^2 \leq \gamma^2, \quad t = 1, 2, \ldots, N
\]

with a specified constant \( \gamma^2 \). Introducing the notations

\[
\alpha = (a_1, \ldots, a_m, b_1, \ldots, b_k)^T, \quad C_t = (y_{t-1}, \ldots, y_{t-m}, u_{t-1}, \ldots, u_{t-k})^T,
\]

rewrite (1) in a more convenient for us form

\[
y_t = C_t \alpha + \xi_t, \quad t = 1, 2, \ldots, N.
\]

The OBE parameter estimation algorithm is described by the relations [5]

\[
\alpha_t = \alpha_{t-1} + K_t \delta_t, \quad \alpha_0 = 0,
\]

where

\[
K_t = \lambda_t M_t^{-1} C_t^T = \lambda_t P_t C_t^T,
\]

\[
M_t = (1 - \lambda_t) M_{t-1} + \lambda_t C_t^T C_t,
\]

\[
\delta_t = y_t - C_t \alpha_{t-1},
\]

\[
\sigma_t^2 = (1 - \lambda_t) \sigma_{t-1}^2 + \lambda_t \gamma^2 - f_t, \quad \sigma_0^2 = 1,
\]

\[
f_t = \lambda_t (1 - \lambda_t) \delta_t^2 / (1 - \lambda_t + \lambda_t G_t),
\]

\[
\lambda_t = \begin{cases} 
(1 - \beta_t) / 2, & G_t = 1, \sigma_{t-1}^2 + \delta_t^2 \leq \gamma^2, \\
(1 - (G_t / (1 + \beta_t (G_t - 1))))^{1/2} / (1 - G_t) & G_t = 1, \sigma_{t-1}^2 + \delta_t^2 > \gamma^2,
\end{cases}
\]

\[
\beta_t = (\gamma^2 - \sigma_{t-1}^2) / \delta_t^2,
\]

\[
P_t = (P_{t-1} - \lambda_t P_{t-1} C_t^T C_t P_{t-1} / (1 - \lambda_t + \lambda_t G_t)) / (1 - \lambda_t), \quad P_0 = \bar{P},
\]

\[
G_t = C_t P_{t-1} C_t^T.
\]

To evaluate \( \alpha \) with help of this algorithm you need to set the initial condition for the matrix \( P_t \). Taking into account that there is no a priori information about \( \alpha \), the standard approach is to use soft initialization assuming \( P_0 = \mu I \) where a large parameter \( \mu > 0 \) is selected by a user and characterizing the degree of the initial uncertainty of \( \alpha \). The limiting cases \( \mu \to \infty \) will be called the diffuse initialization.

It is required to study the OBE algorithm described by the relations (4) – (12) as \( \mu \to \infty \). More exactly, we will consider properties of \( \alpha_t, P_t, K_t, \sigma_t^2 \) as \( \mu \to \infty \) and develop a limit algorithm based on them which we call the diffuse version of the OBE (DOBE algorithm).

3. Main results

The proof of Theorem 1 is based on the following auxiliary statement.

**Lemma 1.** Let \( \Omega_t \) be an \( n \times n \) matrix defined by the expression

\[
\Omega_t = I / \mu + \sum_{k=1}^{l} F_{i,k}^T F_{i,k} = I / \mu + \Lambda_t,
\]

\( i = 1, 2, \ldots, l \).
where $F_{i,k}$ is any $m \times n$ matrix, $k = 1,2,\ldots,t$, $t = 1,2,\ldots,N$. Then:

1. The matrix function $\Omega_{i}^{-1}$ possesses the uniform asymptotic expansion with respect to $t$ on any bounded set $T = \{1,2,\ldots,N\}$

\[
\Omega_{i}^{-1} = (I - \Lambda_{i}^{*})\mu + \Lambda_{i}^{*} + O(1/\mu) \quad \text{as} \quad \mu \to \infty. \quad (14)
\]

2. The following equalities are valid

\[
(I - \Lambda_{i}^{*})F_{i,k}^{T} = 0, \quad k = 1,2,\ldots,t, \quad t = 1,2,\ldots,N. \quad (15)
\]

**Theorem 1.** Let $0 < \lambda_{i} < 1$. Then $P_{t}$ and $K_{t}$ possess the uniform asymptotic expansions with respect to $t$ on any bounded set $T = \{1,2,\ldots,N\}$ and $\xi_{t}^{2} \leq \gamma^{2}, t \in T$

\[
P_{t} = g_{t}^{-1}(I - W_{t}^{-} W_{t}^{+})\mu + W_{t}^{+} + O(1/\mu) \quad \text{as} \quad \mu \to \infty, \quad (16)
\]

\[
K_{t} = \lambda_{t} W_{t}^{+} C_{t}^{T} + O(1/\mu) = \bar{K}_{t} + O(1/\mu), \quad \text{as} \quad \mu \to \infty, \quad (17)
\]

where

\[
W_{t} = (1 - \lambda_{t})W_{t-1} + \lambda_{t}C_{t}^{T}C_{t}, \quad W_{0} = 0, \quad (18)
\]

\[
g_{t} = (1 - \lambda_{t})(1 - \lambda_{t})\ldots(1 - \bar{\lambda}_{t}). \quad (19)
\]

**Consequence 1.** Neglecting the terms of the order of smallness $O(1/\mu)$ in (17) gives

\[
\bar{\alpha}_{t} = (1 - \lambda_{t})\bar{\alpha}_{t-1} + \bar{K}_{t}\bar{\delta}_{t}, \quad \bar{\alpha}_{0} = 0, \quad (20)
\]

where

\[
\bar{\delta}_{t} = y_{t} - C_{t}\alpha_{t-1}. \quad (21)
\]

Denote $e_{t} = \alpha_{t} - \bar{\alpha}_{t}$. The function $\|e_{t}\|$ possesses the uniform asymptotic expansion with respect to $t$ on any bounded set $T = \{1,2,\ldots,N\}$ and $\xi_{t}^{2} \leq \gamma^{2}, t \in T$

\[
\|e_{t}\| = O(1/\mu) \quad \text{as} \quad \mu \to \infty. \quad (22)
\]

**Consequence 2.** The function $f_{t}$ possesses the uniform asymptotic expansion with respect to $t$ on any bounded set $T = \{1,2,\ldots,N\}$ and $\xi_{t}^{2} \leq \gamma^{2}, t \in T$

\[
f_{t} = \begin{cases} 
O(1/\mu) & \text{as} \quad \mu \to \infty, 0 \leq t \leq T^{*} \\
\lambda_{t}(1 - \lambda_{t})\bar{\delta}_{t}^{2}/(1 - \lambda_{t} + \bar{\lambda}_{t}\bar{G}_{t}) + O(1/\mu) & \text{as} \quad \mu \to \infty, t \geq T^{*} + 1
\end{cases}. \quad (23)
\]

where

\[
T^{*} = \min_{t\in T} \{t : W_{t} > 0, t = 1,2,\ldots,N\} = \\
\min_{t\in T} \{t : \sum_{k=1}^{t} q_{i,k}C_{k}^{T}C_{t} > 0, t = 1,2,\ldots,N\} < N, \quad (24)
\]

\[
\bar{G}_{t} = C_{t}W_{t}^{-1}C_{t}^{T}. \quad (24)
\]

Neglecting the terms of the order of smallness $O(1/\mu)$ in (8) gives for $0 \leq t \leq T^{*}$

\[
\bar{\alpha}_{t}^{2} = (1 - \lambda_{t})\bar{\alpha}_{t-1}^{2} + \lambda_{t}\gamma^{2} + f_{t}, \quad \bar{\alpha}_{0}^{2} = 1, \quad (25)
\]

\[
\bar{f}_{t} = \lambda_{t}(1 - \lambda_{t})\bar{\delta}_{t}^{2}/(1 - \lambda_{t} + \bar{\lambda}_{t}\bar{G}_{t}). \quad (26)
\]

**Consequence 3.** It follows from the expansion (16) that the term in $P_{t}$ which is proportional to a large parameter $\mu$ (the diffuse component) vanishing when $t \geq T^{*}$

\[
P_{t}^{\text{diff}} = g_{t}^{-1}(I - W_{t}^{-} W_{t}^{+})\mu
\]

At the same time, the expansion (17) implies that $K_{t}$ does not contain any diffuse component.

**Consequence 4.** Let us show that numerical implementation of the OBE algorithm can result divergence for large values of $\mu$. Indeed, let $\partial W_{t}^{*}$ be a calculations error of $W_{t}^{*}$. Then it follows from (5), (16), (17)
\[ K_t = \lambda_t P_t C_t^T = \lambda_t \{ g_t^{-1} [I - W_t (W_t^+ + \partial W_t^+) \mu + W_t^+ + O(1/\mu)] C_t^T = \lambda_t [-g_t^{-1} W_t \partial W_t^+] \mu + O(1) \} C_t^T \] as \( \mu \to \infty \).

Consider the system for the state transition matrix of the homogeneous part of (20)
\[ X_{i,s} = (I - \bar{K}_t C_t) X_{i-1,s} = (I - \lambda_t W_t^+ C_t^T C_t) X_{i-1,s} = A_t X_{i-1,s}, \quad X_{1,s} = I. \] (28)

The following auxiliary statement is used in the proof of Theorem 2.

**Lemma 2.** The representations
\[ X_{i,s} = \tilde{g}_{i,s} W_{i-1} W_s, \quad t \geq T^*, s < t, \] (29)
\[ X_{i,0} = I - W_t W_t^+, \quad t \in T, \] (30)
\[ X_{i,s} = \lambda_t \tilde{g}_{i,s} W_t^+ C_t^T, \quad t \geq T^*, s < t \] (31)
are valid, where \( \tilde{g}_{i,s} = (1 - \lambda_t) (1 - \lambda_{t-1}) \ldots (1 - \lambda_{s+1}) \).

**Consequence 5.** In the absence of perturbations, the estimation error of the DOBE turns to zero in a finite time as for the diffuse RLS algorithm [6]. So if \( \xi_t = 0, \quad T = \{1, 2, \ldots, N\}, \quad \lambda^* = N, \quad \text{then} \quad e_t = \alpha_t - \alpha = 0 \quad \text{for} \quad \lambda^* \leq t \leq N. \)

**Theorem 2.** Let \( \lambda_t, \quad t = 1, 2, \ldots, N^* \) be any sequence of numbers satisfying the condition \( 0 < \lambda \leq 1. \)

Then the estimate \( \tilde{\alpha}_t \) from (20), (21) does not depend on \( \lambda \), for \( 0 \leq t \leq T^* \).

Summing up, let us write the limit relations for quantities \( \alpha_t, \quad P_t, \quad K_t, \quad \sigma_t^2 \) that describe the DOBE algorithm:
\[ \tilde{\alpha}_t = (1 - \lambda_t) \tilde{\alpha}_{t-1} + \bar{K}_t \tilde{\sigma}_t, \quad \tilde{\alpha}_0 = 0, \] (32)
\[ \tilde{\sigma}_t = y_{i} - C_t \tilde{\alpha}_{t-1}, \] (33)
\[ \bar{K}_t = \lambda_t W_t^+ C_t^T, \] (34)
\[ W_t = (1 - \lambda_t) W_{t-1} + \lambda_t C_t^T C_t, \quad W_0 = 0, \] (35)
\[ \tilde{\sigma}_t^2 = (1 - \lambda_t) \tilde{\sigma}_{t-1}^2 + \lambda_t \gamma^2 - \tilde{f}_t, \quad \tilde{\sigma}_0^2 = 1, \] (36)
\[ \tilde{f}_t = (1 - \lambda_t) \tilde{\sigma}_{t-1}^2 (1 - \lambda_t + \lambda_t \tilde{G}_t), \] (37)
\[ \tilde{G}_t = C_t^T W_t^+ C_t^T, \] (38)

where \( \lambda_t \) is defined by expressions (9), (10) up to the point of replacing \( \sigma_t^2, \tilde{\sigma}_t, G_t \) with \( \tilde{\sigma}_t^2, \tilde{\sigma}_t, \tilde{G}_t \), respectively.

**4. Simulation**
To illustrate the capability of the proposed DOBE algorithm we use the ARX model of the form [5]
\[ y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + a_4 y_{t-4} + \xi_t, \] (39)
where \( a_1 = 0.34, \quad a_2 = -0.28, \quad a_3 = 0.46, \quad a_4 = 0.14, \quad \xi_t \) is the uncorrelated uniformly distributed random process with zero means and known covariance \( \mathbb{E}(\xi_t \xi_t^T) = 1. \) Suppose that each parameter undergoes a ten-percent step change in magnitude at every 200 sampling points. Figure 1 shows the trajectories of actual parameters, their estimates by the DOBE algorithm, the diffuse RLS (DRLS) algorithm and the finite impulse response (DFIRS) algorithms. It is seen that the DOBE algorithm may provide better the tracking ability compared to the DRLS and the DFIRS algorithms. The update sample number of the DOBE was 1585. The used DRLS algorithm is described by the relations [6]
\[ \alpha_t^d = \alpha_{t-1}^d + K_t^d (y_t - C_t \alpha_{t-1}^d), \quad \alpha_0^d = 0, \] (40)
where
\[ K_t^d = R^+_t C_t^T, \] (41)
\[ R_t = \lambda R_{t-1} + C_t^T C_t, \quad 0 < \lambda \leq 1, \quad R_0 = 0. \]  
\hfill (42)

During the simulation process, we set \( \lambda = 1 \). The DFIRLS algorithm use relations (40), (41), (42) in sliding window mode with the size 250 samples. Note that the implementation of the DFIRLS algorithm requires significantly more computing resources compared to the DOBE algorithm.

---

**Figure 1.** Tracking of the parameters \( a_1, a_2, a_3, a_4 \)

---

5. Conclusion

The parameters estimation problem of the ARX models with help of the OBE algorithm in the absence of a priori information about their parameters was considered. It is assumed that the shape matrix of the initial ellipsoid in the OBE algorithm is proportional to a large positive parameter. Asymptotic expansions of the algorithm are obtained. The limit recursive algorithm not depending on a large parameter which may lead the OBE algorithm to a divergence is developed and explored. By simulation, it is shown that the proposed DOBE algorithm may provide better the tracking ability compared to the diffuse RLS algorithm and the diffuse RLS algorithm with the sliding window.

6. References

[1] Haykin S 1991 *Adaptive Filter Theory* (Prentice-Hall, Englewood Cliffs NJ)

[2] Norton J P (Ed.) 1994 Special issue on bounded-error estimation. Part I *International Journal of Adaptive Control and Signal Processing* **8**(1)

[3] Kurzhanski A B 1996 Ellipsoidal calculus for estimation and control *Basel: Birkhauser*

[4] Deller J R, Jr Huang Y F 2002 Set-membership identification and filtering for signal processing applications *Circuits Systems and Signal Processing* volume **21** 69–82

[5] Dasgupta S, Huang Y F. 1987 Asymptotically convergent modified recursive least-squares with data dependent updating and forgetting factor for systems with bounded noise [J] *IEEE Trans.on Information Theory* **33**(3) 383–391

[6] Skorohod B 2017 *Diffuse Algorithms for Neural and Neuro-Fuzzy Networks: With Applications in Control Engineering and Signal Processing* (Elsevier)