Proceeding of the 2004 American Control Conference
Boston, Massachusetts June 30 - July 2, 2004

Extended Recursive Least Squares Algorithm for Nonlinear Stochastic Systems

Huixin Chen*

School of Computing & Technology
University of Sunderland
Sunderland SR6 0DD, UK

Abstract

The strong consistency of parameter estimation has always been one of the main problems in system identification theory especially for the nonlinear systems. Although there are several approaches and algorithms set up for the nonlinear stochastical system, the strong consistency of the parameter estimation has not been discussed and proved for most of them. In this paper, for a class of discrete-time nonlinear stochastic systems, an Extended Recursive Least Squares (ERLS) algorithm is developed.

It has been proved that the proposed ERLS algorithm has strong consistency without the strict restrictions on the system. i.e. (1) The persistent excitation condition has been replaced by a less restrictive condition. (2) The system noises don't have to be white noise. (3) The variance functions of the system noises don't have to be bounded. The convergence rate of the parameter estimation can also be worked out. The results presented in this paper can be applied in more general classes of nonlinear systems such as some of the variable parameter and state dependent parameter nonlinear stochastic systems as well.

1 Introduction

System identification using linear model structures has been extensively developed and the theories such as model order selection, consistency and optimal input selection are mature and have been well discussed in several landmark literatures (Young 1968, Solo 1979, 1980, Chen 1981a, b, Chen 1982, Ljung and Söderström 1983, Chen and Guo 1985, Söderström and Stoica, 1989, Ljung 1999).

However, most of the real system are nonlinear systems. In recent years, there has been much study of the identification problem of nonlinear systems. Although several significant efforts have been put during the past decades to develop techniques and theories for nonlinear system identification theory, the strong consistency issues haven't been discussed and resolved as they have been done in linear systems.

Chen *et al* (1996) has studied the problem for a class of discrete-time bilinear time-invariant stochatic system with coloured noise, the strong consistency of the parameter estimates of extended least squares identification and its convergence rate has been studied for various conditions.

The field of non-linear system identification has had a rapid development in the last decade. Hu *et al* (2001) proposed a class of quasi-ARMAX models for nonlinear systems, which have been successfully applied to prediction, fault detection and adaptive control of nonlinear systems. Van Pelt and Bernstein (2001), considered the identification of Hammerstein/nonlinear feedback models by approximating internal nonlinearities using piecewise linear static maps. They claimed the identification method simultaneously identifies the linear dynamic and static nonlinear blocks without requiring prior assumptions on the form of the static nonlinearity. Coca and Billings (2001) developed a new methodology for identifying nonlinear NARMAX models, from noise corrupted data, is introduced based semi-orthogonal wavelet multiresolution approximations. Young *et al* (2001) outlined how improved estimates of time variable parameters in modles of stochastic dynamic systems can be obtained using recursive filtering and fixed interval smoothing (FIS) techniques, with the associated hyper-parameters optimized by maximum likeihood based on prediction error decomposition. Westwick and Kearney (2001) proposed the use of separable least squares algorithm for the identification of Hammerstein cascades and analyzing stretch reflex electromyogram data from two experimental subjects. Le

*Tel. +44 0191 515 3043 Fax. + 44 0191 515 2781
Email huixin.chen@sunderland.ac.uk

Caillec and Garello (2001) studied the identification of a special class of nonlinear systems, Quadratic AutoRegressive Moving Average systems, QARMA.
The strong consistency of the parameter estimate means the estimated parameters in the algorithm converge to the real parameters of the system with probability one (or almost sure). The strong consistency in the parameter estimate is very important in the mathematical modeling since we need to know if the estimated model is appropriate to the real system. We need to see if the proposed algorithm is convergent. Even if it does converge, does it converge to the real parameters of the model. What we have done in this paper is to develop an algorithm, which can guarantee the system parameters converges to the real parameter and the convergence rate can be calculated as well under the assumptions of the system.
This paper is organised as follows: In section 2, a class of nonlinear stochastic system is introduced. Some notations, assumptions and detailed ERLS algorithm are given in the section 3. The main system and convergence analysis for the nonlinear ERLS are presented in the section 4. Some simulation results are presented in the section 5. Some conclusion and further comments are made in the section 6.

2 A class of nonlinear stochastic Systems

$$A(z)y_t = B(z)f_t + C(z)w_t, t = 1, 2, 3, ..., \tag{1}$$

where z is the unit shift backshift operator, $A(z), B(z)$ and $C(z)$ are the polynomials:

$$\begin{aligned} A(z) &= I_n - A_1 z - \ldots - A_{n_y} z^{n_y}; \\ B(z) &= B_1 z + \ldots + B_{n_f} z^{n_f}; \\ C(z) &= I_n + C_1 z + \ldots + C_{n_w} z^{n_w}; \end{aligned}$$

y_t, u_t and w_t are the n-dimensional output vectors, m-dimentional input vectors and n-dimensional noise vetors of the system respectively and $y_t \equiv w_t \equiv 0; u_t \equiv 0 \quad \forall t \le 0$; f_t is a m_f-dimensional known bounded function of the system input, output with noises, i.e. $f_t = f(u_t, ..., u_{t-q}, y_t, ..., y_{t-p})$.
where p and q are integers; I_n is an $n \times n$ unit matrix, A_i , B_j and C_k are $n \times n$, $n \times m_f$ and $n \times n$ matrices respectively, $i = 1, ..., n_y, j = 1, \ldots, n_f$ and $k = 1, \ldots, n_w$;
The model (1) we introduced in this paper is a quite general nonlinear system model. The model structure we presented in the form of (1) theoretically can as an approximated model for most of the nonlinear relationship systems. In practical, first we can estimate an unknown system by using of all previous mentioned algorithms or bigger dimension parameter structure form of (1) as a primary model, then can use the primary models' identification information to do more precise model structure estimation and validation for the model (1). So it is very important we introduce the model (1) as our key model to discuss algorithms.
We introduce the following notations:

$$\begin{aligned} \theta &= [A_1, A_2, \ldots, A_{n_y}, B_1, \ldots, B_{n_f}, \\ &\quad C_1, C_2, \ldots, C_{n_w}]^T; \\ \varphi_t^o &= [y_t^T, y_{t-1}^T, \ldots, y_{t-n_y}^T, f_t^T, \ldots, , f_{t-n_f}^T, \\ &\quad w_{t-1}^T, \ldots, w_{t-n_w}^T]^T \end{aligned} \tag{2}$$

and the system (1) can also be written as the following form:

$$y_t = \theta^T \varphi_t^o + w_t \tag{3}$$

3 Notations, Assumptions and ERLS Algorithm

The extended least squares method is applied to estimate the unknown parameter θ of (2), and let θ_t represent the estimate of θ at time t. The recursive algorithms is the same as that presented by Chen *at al* (1986).

$$K_t = R_{t-1}\varphi_t/(1 + \varphi_t^T R_{t-1}\varphi_t) \tag{4}$$
$$R_t = R_{t-1} - K_t \varphi_t^T R_{t-1} \tag{5}$$
$$\theta_t = \theta_{t-1} + K_t(y_t^T - \varphi_t^T \theta_{t-1}) \tag{6}$$
$$e_t = y_t - \varphi_t^T \theta_t \tag{7}$$

where φ_t is constructed by using e_{t-i} instead of w_{t-i} in φ_t^o, $i = 1, 2, \ldots, n_w$ and $e_t = 0$ when $t \le 0$. θ is a $n \times n$ matrix and φ_t is h-dimensional vector, where $h = n_y \times n + n_f \times m_f + n_w \times n$. We select $R_0 = hI_h, \theta_0 = 0_{h \times n}$. Actually, R_0 can be any hth order positive definite matrix and θ_0 can be any h-dimensional vector.
Several notations and definitions are given in the following, which will be used in the system and algorithm convergence analysis later in the paper.

- The norm of the vector x is defined as $\|x\| = (x^T x)^{1/2}$ and the norm of the matrix X is defined as maximum eigenvalue of the matrix $X^T X$.
- The series of the trace of the matrices R_t is given as follows;

$$r_0 = Tr(R_0^{-1}), \quad r_t = r_{t-1} + \sum_{i=1}^{t} \|\varphi_i\|^2 \tag{8}$$

where $Tr(X)$ denotes the trace of the matrix X.

- The set F_t is the $\sigma-$algebra set generated by $\{w_s,\ s \leq t\}, i.e.\ F_t = \sigma\{w_s,\ s \leq t\}$.
- The estimate error matrix of the parameter $\tilde{\theta}_t = \theta - \theta_t$.
- $b_t = e_t - w_t$.
- The λ_t^{max} and λ_t^{min} are denoted as the maximum and minimum eigenvalues of the matrix R_t^{-1} respectively.
- We denote $det(X)$ as the determinant of the matrix X.
- We define two functions as follows:

$$\begin{aligned} \log k(x) &= \underbrace{\log\log\ldots\log(x)}_{k \quad \text{times}} \\ L_k^{\delta}(x) &= \log(x)\log 2(x)\ldots \\ &\qquad \log(k-1)(x)[\log k(x)]^{\delta} \end{aligned} \quad (9)$$

  where x is a sufficient large positive number and $\delta > 1$.

In order to prove the strong consistency of the algorithms, several assumptions are needed for the system as follows:

- *A1*: The system (1) is a BIBO system.
- *A2*: The noise w_t is assumed to be a martingale difference sequence and for all $t = 2, 3, \ldots$, the following properties hold:

$$\begin{aligned} E(w_t/F_{t-1}) &= 0, \quad a.s. \quad \forall t \geq 1, \quad F_0 = \{\emptyset, \Omega\} \\ E(w_t^2/F_{t-1}) &\leq \xi_0 r_{t-1}^{\epsilon}, \quad a.s \quad \forall t \geq 1, \\ & \qquad 0 < \xi_0 < \infty,\ 0 \leq \epsilon < 1 \end{aligned} \quad (10)$$

  wher E is an expectation operator.
- *A3*: The input u_t is F_t-measurable and $E(\|u_t\|^2) \leq \infty, \quad t = 0, 1, 2, \ldots$. If u_t is a deterministic signal, then u_t is bounded.
- *A4*: f is a bounded function
- *A5*: $r_t \to \infty$ as $t \to \infty$
- *A6*: Positive real condition:

  $C^{-1}(z) - \frac{1}{2}I$ is strictly positive real and all the zeros of $C(z)$ are outside the unit circle.
- *A7*: Improved persistent excitation condition:

  There exists a natural number k and constant $c > \delta > 1$ such that

$$\lim_{t\to\infty} r_t^{\epsilon} L_k^c(r_t)/\lambda_t^{min} = 0 \qquad a.s. \quad (11)$$

Remark 1. It should also assume that the system is stable when its parameters take values at the current parameter estimate.

4 Convergence Analysis

In this section, we discuss the issue of the strong consistency and convergence rate of the Extended Recursive Least Squares (ERLS) Identification for the system (1).

Lemma 1: *Under the assumptions of A1-A7, the ERLS algorithm (4-7) of the system (2) has the following properties, for t=1,2,3,...*

$$r_t = Tr(R^{-1}), r_t < \infty \quad (12)$$
$$E(\|y_t\|^2) < \infty, E(\|\theta_t\|^2) < \infty, E(\|e_t\|^2) < \infty, a.s. \quad (13)$$
$$E(\|\tilde{\theta}_t\|^2) < \infty, E(\|b_t\|^2) < \infty,\ a.s. \quad (14)$$
$$\tilde{\theta}_t = \tilde{\theta}_{t-1} - R_{t-1}\varphi_t(b_t + w_t) \quad (15)$$
$$\varphi_t^T R_t \varphi_t = (detR_t^{-1} - detR_{t-1}^{-1})/detR_t^{-1} \quad (16)$$
$$C(z)b_t = \tilde{\theta}_t^T \varphi_t \quad (17)$$

Proof:
It could be proved in a vary similar way in (Chen *et al*, 1996).

Lemma 2: *Under the assumptions of A1-A7, for any natural number k and some positive number $\delta > 1$, there exists an integer N(k) such that*

$$\sum_{t=N(k)}^{\infty} \varphi_t^T R_t \varphi_t / L_k^{\delta}(r_t) < \infty \quad a.s. \quad (18)$$

Proof:
It could be proved in a vary similar way in (Chen *et al*, 1996).

Lemma 3: *Under the assumptions of A1-A7, for any natural number k, some positive number $\delta > 1$ and integers k_1, k_2 and $N(k)$, and let:*

$$V_t = Tr(\tilde{\theta}_t R_t^{-1} \tilde{\theta}_t)/r_t^{\epsilon} L_k^{\delta}(r_t),\ \forall t \geq N(k) \quad (19)$$

where ϵ is given by (12). Then

$$\lim_{t\to\infty} V_t = V < \infty \quad (20)$$

The outline of the Proof:
1. First we prove that:

$$\begin{aligned} & E(V_t/F_{t-1}) \leq V_{t-1} + 2\xi_0 \varphi_t^T R_t \varphi_t / L_k^{\delta}(r_t) \\ & + E(\varphi_t^T \tilde{\theta}_t (\varphi_t^T \tilde{\theta}_t - 2b_t)/F_{t-1})/r_t^{\epsilon} L_k^{\delta}(r_t),\ \forall t \geq N(k) \end{aligned} \quad (21)$$

2. Second we prove that :

$$E(\xi_{t+1}/F_t) \leq \xi_t + \eta_t - \zeta_t \quad (22)$$

where

$$S_t = \sum_{i=N(k)}^{t} \varphi_i^T \tilde{\theta}_i \left(C^{-1}(z) - \frac{1+k_1}{2} \right) \varphi_i^T \tilde{\theta}_i + k_2 \geq 0, \qquad \forall t \geq N(k) \tag{23}$$

and

$$\xi_t = V_t + 2S_t / r_t^{\epsilon} L_k^{\delta}(r_{t+1}) \tag{24}$$

$$\eta_t = 2\xi_0 \varphi_{t+1}^T R_{t+1} \varphi_{t+1} / L_k^{\delta}(r_{t+1}) \tag{25}$$

$$\zeta_t = k_1 E \left(\|(\varphi_{t+1}^T \tilde{\theta}_{t+1})\|^2 / F_t \right) / r_{t+1}^{\epsilon} L_k^{\delta}(r_{t+1}) \tag{26}$$

3. Thirdly we prove that

$$\lim_{t\to\infty} V_t = V < \infty \tag{27}$$

Theorem 1: *Under the assumptions of A1-A7, the ERLS algorithm (4-7) of the system (2) has results as followings:*

$$1. \quad \lim_{t\to\infty} \theta_t = \theta \ a.s. \tag{28}$$

$$2. \quad \|\tilde{\theta}_t\| = O(r_t^{\epsilon} L_k^{c}(r_t)/\lambda_t^{min})^{1/2} \quad a.s. \tag{29}$$

$$3. \quad \lim_{t\to\infty} \frac{1}{r_t^{\epsilon} L_k^{\delta}(r_t)} \sum_{i=1}^{t} \|b_i\|^2 = 0 \quad a.s. \tag{30}$$

Proof:
Since $\tilde{\theta}_t^T R_t^{-1} \tilde{\theta}_t \geq \lambda_t^{min} \|\tilde{\theta}_t\|^2$, and from (19), we have:

$$\|\tilde{\theta}_t\|^2 \leq V_t r_t^{\epsilon} L_k^{\delta}(r_t)/\lambda_t^{min}, \quad a.s. \tag{31}$$

From A7 and (27), both of the equations (28) and (29) can be deduced. The equation (30) can be obtained directly from (21).

5 Simulations

In this section several examples are given to show the performance and application of the algorithm developed in this paper. We considered a nonlinear system with correlated noise $\epsilon_t = e_t - 0.4e_{t-1}$ as follows:

$$y_t = 0.8y_{t-1}\cos(\frac{1}{2}u_{t-1}) - 0.8u_{t-1}\cos(\frac{1}{2}y_{t-1}) + e_t - 0.4e_{t-1} \tag{32}$$

We use different model strutures with unknown parameters to estimate the model respectively and the simulation details are given as follows:
In the following examples, the system input $u_t = U[-0.5, 0.5]$, which is a uniform distribution between -0.5 to 0.5 and $e_t = N(0, \sigma^2)$ is a normal distribution with the zero mean and σ^2 variance.

The estimations are based on 1000 samples of u_t and y_t under several different noise conditions. Here the percentage noise is defined in terms of the standard deviation of the noise e_t in relation to the standard deviation of the noise-free output.

Example 1 The model to be identified is in the third order approximation form of any nonlinear system of $y_t = f(u_{t-1}, y_{t-1}) + e_t + c_1 e_{t-1}$ as follows:

$$y_t = a_1 y_{t-1} + b_{1,1}u_{t-1} + b_{2,1}u_{t-1}y_{t-1} + b_{3,1}u_{t-1}^2 + b_{4,1}y_{t-1}^2 + b_{5,1}u_{t-1}^2 y_{t-1} + b_{6,1}u_{t-1}y_{t-1}^2 + b_{7,1}u_{t-1}^3 y_t + b_{8,1}y_{t-1}^3 + e_t + c_1 e_{t-1} \tag{33}$$

For the model (32) and Taylor expansion, we can know the true value of the parameters of the third order approximation form of (33). *i.e.* $a_1 = 0.8, b_{1,1} = -0.8, b_{5,1} = -0.1, b_{6,1} = 0.1, c_1 = -0.4$ and $b_{2,1} = b_{3,1} = b_{4,1} = b_{7,1} = b_{8,1} = 0$.
The details of the simulation result are given in the following Table 1.

Table 1. Parameter estimation for Model given by Example 1

Noise	Parameter Estimation				
σ	$\hat{a}_1$	$\hat{b}_{1,1}$	$\hat{b}_{2,1}$	$\hat{b}_{3,1}$	$\hat{b}_{4,1}$
10%	0.80 (0.03)	-0.81 (0.03)	0.00 (0.01)	0.00 (0.01)	-0.00 (0.00)
50%	0.81 (0.04)	-0.80 (0.03)	-0.01 (0.04)	-0.02 (0.03)	0.02 (0.03)
100%	0.82 (0.04)	-0.81 (0.05)	-0.00 (0.05)	0.03 (0.13)	0.01 (0.04)
	$\hat{b}_{5,1}$	$\hat{b}_{6,1}$	$\hat{b}_{7,1}$	$\hat{b}_{8,1}$	$\hat{c}_1$
10%	-0.10 (0.01)	0.10 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.41 (0.02)
50%	-0.11 (0.03)	0.11 (0.03)	-0.02 (0.02)	-0.02 (0.03)	-0.41 (0.02)
100%	-0.09 (0.09)	0.09 (0.04)	0.02 (0.04)	-0.01 (0.03)	-0.39 (0.06)

where the number inside the bracket denoting the standard deviation of the relevant parameter, which is calculated based on the samples of the relevant estimated parameters. From Table 1 and Figure 1, it may be noticed that the parameter estimates $\hat{b}_{2,1}, \hat{b}_{3,1}, \hat{b}_{4,1}, \hat{b}_{7,1}$ and $\hat{b}_{8,1}$ are very close to zero. It can be made the assumption that they are zero parameters and a new model candidate can be suggested to investigate as in Chen (2003).

Example 2.
The model to be identified is assumed as the following form:

$$y_t = a y_{t-1}\cos(\frac{1}{2}u_{t-1}) + b u_{t-1}\cos(\frac{1}{2}y_{t-1})$$

$$+e_t + ce_{t-1} \tag{34}$$

where $a = 0.8, b = -0.8, c = -0.4$ The details of the simulation result are given in the following Table 2.

Table 2. Parameter estimation for Model given by Example 2

Noise	Parameter Estimation		
σ	$\hat{a}$	$\hat{b}$	$\hat{c}$
10%	0.8023 (0.0255)	-0.8039 (0.0262)	-0.4057 (0.0172)
50%	0.8043 (0.0335)	-0.8050 (0.0340)	-0.4083 (0.0192)
100%	0.8095 (0.0341)	-0.8079 (0.0430)	-0.3962 (0.0656)

From above mentioned Examples 1-2, it is shown that the proposed algorithm has a very good performance for the nonlinear system and the wide range applications. The algorithm can be either used as the primary model structure identification, or further model structure identification and validation since it has guaranteed the strong consistency for the parameter estimations. It can also be suggested to select an appropriate model structure to make the comprimise between model accuracy and calculation cost according to the data and model requirement.

The 200 sample segments (800-1000) of the parameter estimations from Example 1 in the case of 10% noise level are presented in Figure 1.

6 Comments and Conclusion

The model we discussed here is a quite general one. The algorithm we introduced in this paper has a major advantage of the results of the strong consistency and covergence rate as long as the system meets certain assumptions listed as A1-A7. In order to identify the nonlinear system properly, we can either use the algorithm and system model (1) as a primary model structure identification and do the model modification and validation afterwards; or use the identification results from some other mentioned methods such as Young *et al* (2001), Pelt and Bernstein (2001), Hu *et al* (2001) and Coca and Billings (2001) to do some further parameter estimation or model validations. The simulation results given by the examples 1-2 have shown the good convergence results of the algorithm.

Acknowledgement:

The author would like to represent the sincere thanks for the three anonymous referees for their valuable comments and suggestions for the revision of the paper.

7 References

Chen, H.F., 1981a, Strong consistency of recursive identification under correlated noise. *Journal of Systems Science and Mathematical Science*, **1**(1), 34-52.

Chen, H.F.,1981b, Quasi Least-squares identification and its strong consistency. *International Journal of Control*, **34**, 921-936.

Chen, H.F., 1982, Strong consistency and convergence rate of least squares. *Scientia Sinica, (Series A)*, **25**, 771-784.

Chen, H.F., and Guo, L., 1985, Convergence rate of least-squares identification and adaptive control for linear stochastic systems *Int.J.Control,* Vol.**44**, No.5, 1459-1476.

Chen, H., 2003, Strong Consistency and Convergence Rate for Stochastic Nonlinear Systems. Submitted for *Automatica*.

Chen, H.X., Zinober, A.S.I., and Ruan, R.Y., 1996, Strong consistency and convergence rate of parameter identification for bilinear systems, *Int.J.Control,* Vol.**63**, No.5, 907-919.

Coca, D. and Billing, S.A., 2001, Non-linear System Identification Using Wavelet Multiresolution Models, *Int. J. Control*, Vol. 74, No. 18, 1718-1736

Guo, L., and Chen, H.F., 1991, The Astrom Wittenmark self-tuning regulator revisited and ELS-based adaptive trakers, *IEEE Trans. Automat. Contr.* Vol.**36**, No.7, 802-812.

Hu, J., Kumamaru, K. and Hirasawa, K., 2001, A Quasi-AMAX approach to modelling of non-linear systems, *Int.J.Cotrol*, Vol74, No 18, 1754-1766.

Le Caillec J-M. and Garello, R., 2001, Nonlinear system identification using autoregressive quadratic models. *Signal Processing* Vol. 81, 357-379

Ljung, L., and Soderstorm, T., 1983, Theory and Practice of Recursive Identification (MIT Press).

Ljung, L., 1999, System Identification: Theory for the User, 2nd edition (Upper Saddle River, NJ:Prentice Hall).

Soderstrom, T., and Stoica, P., 1989, System Identification (Upper Saddle River, NJ: Prentice Hall).

Solo, V., 1979, The convergence of AML. *IEEE Trans. Auto. Control*, AC-24,958-962.

Solo, V., 1980, Some aspects of recursive parameter estimation, *Int. J. Control*, 32, 395-410

Van Pelt, T.H. and BernStein, D.S., 2001, Non-linear system identification using Hammerstein and nonlinear feedback models with piuecewise linear static maps. *Int. J. Control*, Vol. 74, No. 18, 1807-1823.

Westwick, D.T. and Kearney, R.E., 2001, Separable Least Squares Identification of Nonlinear Hammerstein Models: Application to Strech Reflex Dynamics, *Annals of Biomedical Engineering*, Vol 29, 707-718.
Young, P. C., 1968, The use of linear regression and related procedures for the identification of dynamic processes, *Proc. 7th IEEE Symposium on Adaptive processes*, UCLA, IEEE: New York, 5C1-5C5.
Young, P.C., McKenna, P. and Bruun, J., 2001, Identification of non-linear stochastic system by state dependent parameter estimation., *Int. J. Control*, Vol. 74, No. 18, 1837-1857.

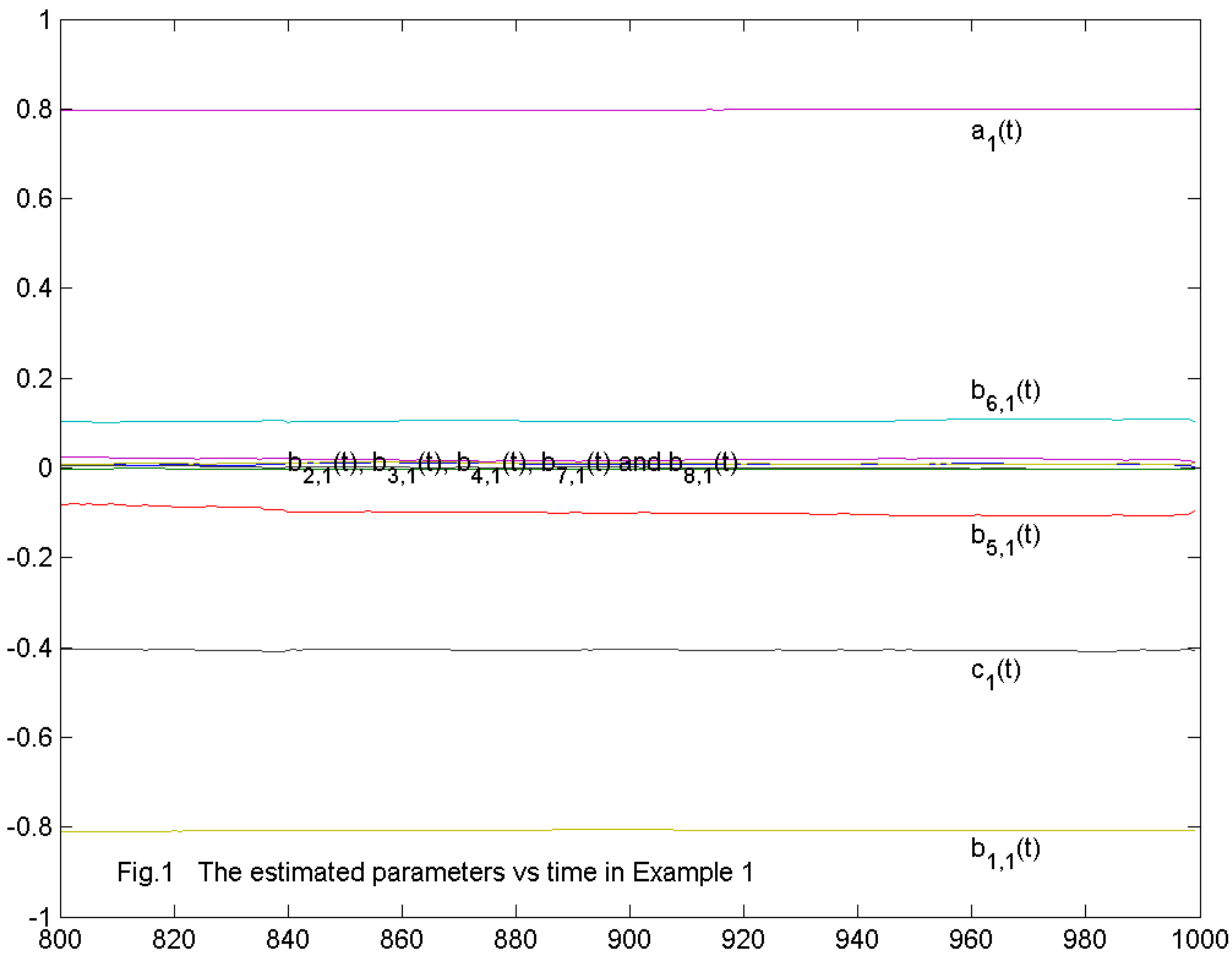

Fig.1 The estimated parameters vs time in Example 1