Modified stochastic gradient estimation algorithms for Box-Jenkins model based on auxiliary model

Jianxia Feng
Yingtian College, Nanjing, 210023, China
Email: lynette_feng@sina.com

Keywords: Parameter estimation, Stochastic gradient, Auxiliary model, Box-Jenkins model, Convergence rate

Abstract. An auxiliary model based stochastic gradient estimation algorithm in proposed in this paper. The unknown variables in the information vector can be estimated by using the auxiliary model. Then the unknown parameters can be estimated by the stochastic gradient algorithm. Furthermore, in order to increase the convergence rate, a modified stochastic gradient algorithm is also proposed. The simulation results indicate that the proposed algorithm has good performances.

Introduction

Consider a Box-Jenkins model [1]:

$$y(t)=\frac{B(z)}{A(z)}u(t)+\frac{D(z)}{C(z)}v(t), \tag{1}$$

where $u(t)$ and $y(t)$ are the input and output of the Box-Jenkins model, respectively, $v(t)$ is a noise with zero mean, $A(z)$, $B(z)$, $C(z)$ and $D(z)$ are scalar polynomials in the unit backward shift operator $z^{-1}[z^{-1}y(t)=y(t-1)]$ and

$$A(z):=1+a_1z^{-1}+a_2z^{-2}+\cdots+a_nz^{-n},$$

$$B(z):=b_1z^{-1}+b_2z^{-2}+\cdots+b_nz^{-n},$$

$$C(z):=1+c_1z^{-1}+c_2z^{-2}+\cdots+c_nz^{-n},$$

$$D(z):=1+d_1z^{-1}+d_2z^{-2}+\cdots+d_nz^{-n}.$$

In [2], Chen and Zhang presented an auxiliary model based multi-innovation extended stochastic gradient (SG) algorithm to estimate the OEMA systems. The SG algorithm has less computational effort but slower convergence rate than the recursive least squares (RLS) algorithm [3, 4, 5, 6]. In order to improve the convergence rate of the SG algorithm, we presented a modified SG algorithm in this paper.

The auxiliary model method is a useful method which is usually utilized to identify systems with unknown variables. [7, 8, 9, 10, 11]. For example, Ding et al proposed an auxiliary model based RLS algorithm for dual-rate state space systems with time-delay, the inner variables can be predicted by the auxiliary model and the parameters can be estimated by the RLS algorithm [12]. Chen provided a missing-output estimation model based SG algorithm for a class of dual-rate linear systems [13], the basic idea is to keep the parameter estimate updating at the slow rate, then to replace the unavailable outputs in the two slow samples with the outputs of the auxiliary model by using the parameter estimate.

In this paper, we will use a auxiliary model based modified SG algorithm for Box-Jenkins systems. The unknown inner variables can be estimated by an auxiliary model, then the parameters can be estimated by the SG algorithm. In order to increase the convergence rate, a modified SG algorithm is also proposed. Briefly, the paper is organized as follows. Section 2 introduces the identification model related to Box-Jenkins model. Section 3 derives a modified SG algorithm for the Box-Jenkins model. Section 4 provides an illustrative example. Finally, concluding remarks are given in Section 5.

The system description and identification model

First, let us introduce some notations first. The symbol I stands for an identity matrix of the appropriate sizes; the norm of a matrix X is defined as $\| X \| := tr[XX^T] = tr[X^T X]$; the superscript T denotes the matrix transpose.

Define the inner variables $x(t)$ and $w(t)$ as follows:

$$x(t) := \frac{B(z)}{A(z)} u(t), \qquad w(t) := \frac{D(z)}{C(z)} v(t).$$

Define the parameter vectors θ_1 and θ_2 as

$$\theta_1 := [a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n]^T \in R^{2n}, \tag{2}$$

$$\theta_2 := [c_1, c_2, \cdots, c_n, d_1, d_2, \cdots, d_n]^T \in R^{2n}, \tag{3}$$

and the information vectors $\varphi_1(t)$ and $\varphi_2(t)$ as

$$\varphi_1(t) := [-x(t-1), -x(t-2), \cdots, -x(t-n), u(t-1), u(t-2), \cdots, u(t-n)]^T \in R^{2n}, \tag{4}$$

$$\varphi_2(t) := [-w(t-1), -w(t-2), \cdots, -w(t-n), v(t-1), v(t-2), \cdots, v(t-n)]^T \in R^{2n}. \tag{5}$$

Then Equation (1) can be simplified as

$$y(t) = \varphi_1^T(t)\theta_1 + \varphi_2^T(t)\theta_2 + v(t) = \varphi^T(t)\theta \quad , \tag{6}$$

where

$$\varphi(t) := [\varphi_1^T(t), \varphi_2^T(t)]^T \in R^{4n}, \tag{7}$$

$$\theta := [\theta_1^T, \theta_2^T]^T \in R^{4n}. \tag{8}$$

Using the SG algorithm to estimate the parameter vector θ gets

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\varphi(t)}{r(t)} e(t), \tag{9}$$

$$e(t) = y(t) - \varphi^T(t)\hat{\theta}(t-1), \tag{10}$$

$$r(t) = r(t-1) + \| \varphi(t) \|^2, r(0) = 1. \tag{11}$$

Clearly, $\varphi(t)$ contains unknown variables $x(t-i)$, $w(t-i)$ and $v(t-i)$. Thus the SG algorithm cannot estimate the unknown parameter vector. We will apply the auxiliary model to overcome this difficulty. The unknown variables can be replaced by the outputs of two auxiliary models.

$$\hat{x}(t) = \hat{\varphi}_1^T(t)\hat{\theta}_1(t), \tag{12}$$

$$\hat{w}(t) = \hat{\varphi}_2^T(t)\hat{\theta}_2(t), \tag{13}$$

$$\hat{v}(t) = y(t) - \hat{x}(t) - \hat{w}(t). \tag{14}$$

where $\hat{x}(t)$, $\hat{w}(t)$ and $\hat{v}(t)$ are the estimates of $x(t)$, $w(t)$ and $v(t)$ at time t, respectively, and

$$\hat{\varphi}_1(t) = [-\hat{x}(t-1), -\hat{x}(t-2), \cdots, -\hat{x}(t-n), u(t-1), u(t-2), \cdots, u(t-n)]^T \in R^{2n}, \tag{15}$$

$$\hat{\varphi}_2(t) = [-\hat{w}(t-1), -\hat{w}(t-2), \cdots, -\hat{w}(t-n), \hat{v}(t-1), \hat{v}(t-2), \cdots, \hat{v}(t-n)]^T \in R^{2n}. \tag{16}$$

Then we can get the following auxiliary model based SG algorithm (AM-SG)

$$\hat{\theta}(t)=\hat{\theta}(t-1)+\frac{\hat{\varphi}(t)}{\hat{r}(t)}e(t), \tag{17}$$

$$\hat{\varphi}(t)=[\hat{\varphi}_1^T(t),\hat{\varphi}_2^T(t)]^T, \tag{18}$$

$$\hat{\varphi}_1(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n),u(t-1),u(t-2),\cdots,u(t-n)]^T, \tag{19}$$

$$\hat{\varphi}_2(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n)]^T. \tag{20}$$

$$e(t)=y(t)-\hat{\varphi}^T(t)\hat{\theta}(t-1), \tag{21}$$

$$\hat{v}(t)=y(t)-\hat{x}(t)-\hat{w}(t), \tag{22}$$

$$\hat{r}(t)=\hat{r}(t-1)+\|\hat{\varphi}(t)\|^2,\hat{r}(0)=1. \tag{23}$$

The AM-M-SG algorithm

Compared with the RLS algorithm, the SG has slow convergence rates. In order to increase the convergence rate but not to increase the computational effort. In this section, We will introduce the modified SG algorithm. The auxiliary model based modified SG (AM-M-SG) algorithm is summarized as follows:

$$\hat{\theta}(t)=\hat{\theta}(t-1)+\frac{\hat{\varphi}(t)}{\hat{r}^{\varepsilon}(t)}e(t),\ \ \frac{1}{2}\le\varepsilon<1, \tag{24}$$

$$\hat{\varphi}(t)=[\hat{\varphi}_1^T(t),\hat{\varphi}_2^T(t)]^T, \tag{25}$$

$$\hat{\varphi}_1(t)=[-\hat{x}(t-1),-\hat{x}(t-2),\cdots,-\hat{x}(t-n),u(t-1),u(t-2),\cdots,u(t-n)]^T, \tag{26}$$

$$\hat{x}(t)=\hat{\varphi}_1^T(t)\hat{\theta}_1(t), \tag{27}$$

$$\hat{w}(t)=\hat{\varphi}_2^T(t)\hat{\theta}_2(t), \tag{28}$$

$$\hat{\varphi}_2(t)=[-\hat{w}(t-1),-\hat{w}(t-2),\cdots,-\hat{w}(t-n),\hat{v}(t-1),\hat{v}(t-2),\cdots,\hat{v}(t-n)]^T, \tag{29}$$

$$e(t)=y(t)-\hat{\varphi}^T(t)\hat{\theta}(t-1), \tag{30}$$

$$\hat{v}(t)=y(t)-\hat{x}(t)-\hat{w}(t). \tag{31}$$

$$\hat{r}(t)=\hat{r}(t-1)+\|\hat{\varphi}(t)\|^2,\hat{r}(0)=1. \tag{32}$$

The steps of computing the parameter estimation vector $\hat{\theta}(t)$ by the AM-M-SG algorithm are listed in the following

1. Let $u(t)=0$, $y(t)=0$, $\hat{x}(t)=0$, $\hat{w}(t)=0$, and $\hat{v}(t)=0$, $:t\le 0$, and choose a small constant ε satisfied $\frac{1}{2}\le\varepsilon<1$.

2. Collect the measured data $\{u(t),y(t):t=0,1,2,\cdots\}$.

3. To initialize, let $t=1$, $\hat{\theta}(0)=1/p_0$.

4. Compute $\hat{x}(t)$ according to (27).

5. Compute $\hat{w}(t)$ by (28).

6. Compute $\hat{v}(t)$ by (31).

7. Form $\hat{\varphi}_1(t)$ and $\hat{\varphi}_2(t)$ by (26) and (29), respectively.

8. Form $\hat{\varphi}(t)$ according to (25).

9. Compute $\hat{r}(t)$ and $e(t)$ by (32) and (30), respectively.

10. Update the parameter estimation vector $\hat{\theta}(t)$ by (24), and compare $\hat{\theta}(t)$ and $\hat{\theta}(t-1)$: if they are sufficiently close, or for some preset small ε, if $\|\hat{\theta}(t)-\hat{\theta}(t-1)\|\leq\varepsilon$, then terminate the procedure and obtain the estimate $\hat{\theta}(t)$; otherwise, increase t by 1 and go to step 4.

Example

Consider the following Box-Jenkins system,

$$y(t)=\frac{0.2z^{-1}+0.8z^{-2}}{1+0.9z^{-1}}u(t)+\frac{1+0.3z^{-1}}{1+0.4z^{-1}}v(t), \tag{33}$$

the $\{v(t)\}$ is taken as a white noise sequence with zero mean and variance $\sigma^2=0.10^2$.

Firstly, apply the AM-SG algorithm to estimate the parameters of this system, the parameter estimates and their errors are shown in Table 1, the parameter estimation errors $\delta:=\|\hat{\theta}-\theta\|/\|\theta\|$ versus t are shown in Figure 1.

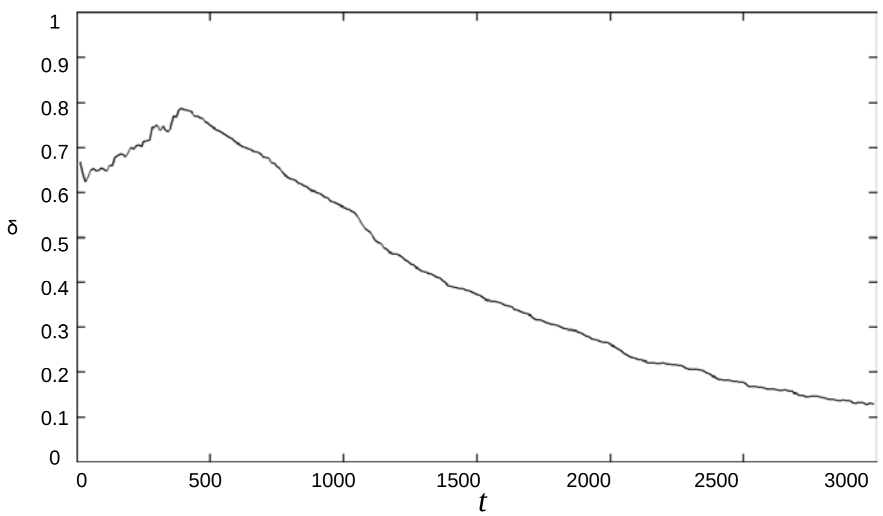

Figure 1: The parameter estimation errors δ versus t (AM-SG)

Table 1: The AM-SG estimates and errors

t	a_1	b_1	b_2	c_1	d_1	δ (%)
500	0.86109	0.31421	0.71005	0.49593	-0.80690	74.90847
1000	0.90250	0.31211	0.76006	0.39004	-0.60584	56.73341
1500	0.90225	0.29384	0.75008	0.25827	-0.40101	37.29563
2000	0.88398	0.32537	0.77878	0.15924	-0.29753	26.11562
2500	0.90106	0.26217	0.72520	0.08954	-0.22675	17.80354
3000	0.89523	0.27859	0.71247	0.04439	-0.17767	12.73540
True Values	0.90000	0.29000	0.73000	-0.10000	-0.13000	

Secondly, apply the AM-M-SG algorithm to estimate the parameters of this system, the parameter estimates and their errors are shown in Table 2, the parameter estimation errors $\delta := \|\hat{\theta} - \theta\| / \|\theta\|$ versus t are shown in Figure 2.

Table 2: The AM-M-SG estimates and errors

t	a_1	b_1	b_2	c_1	d_1	δ (%)
500	0.90544	0.31021	0.74915	0.39139	-0.67252	60.75144
1000	0.90464	0.31711	0.75616	0.23511	-0.44858	38.47598
1500	0.90594	0.28721	0.76135	0.07587	-0.25203	17.95033
2000	0.88593	0.34340	0.78071	-0.01146	-0.17715	10.38625
2500	0.89664	0.24796	0.73099	-0.06227	-0.14532	4.86302
3000	0.88325	0.28136	0.71766	-0.09332	-0.13257	1.96073
True Values	0.90000	0.29000	0.73000	-0.10000	-0.13000	

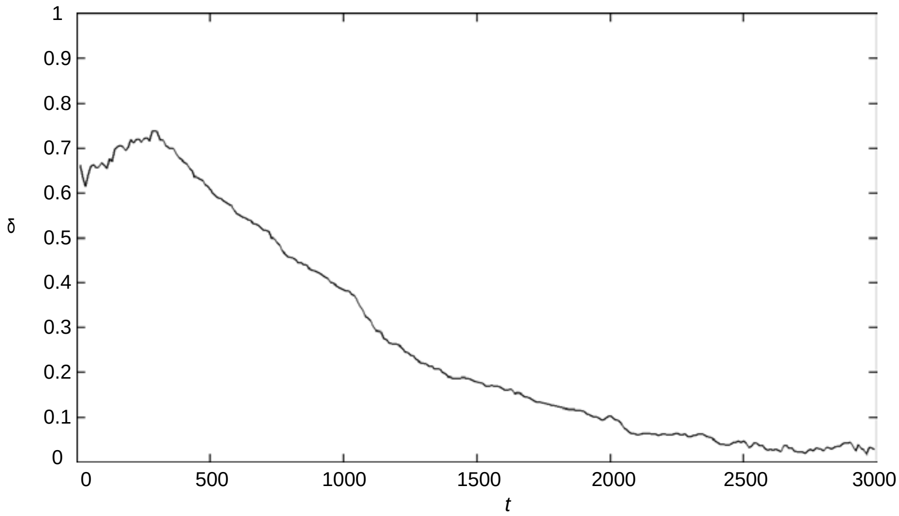

Figure 2: The parameter estimation errors δ versus t (AM-M-SG)

Finally, we can get the following conclusions:

1. From Figures 1 and 2, we can get that the AM-M-SG algorithm has a quicker convergence rate than the AM-SG algorithm.
2. Tables 1 and 2 witness that the AM-M-SG is more accurate than the AM-SG algorithm.

Conclusions

This paper proposes an AM-M-SG algorithm for Box-Jenkins model. By using the auxiliary model, the inner variables can be estimated. Then the parameters can be estimated by the M-SG algorithm. Compared with the SG algorithm, the M-SG algorithm can increase the convergence rates and can keep the computational effort unchanged. Thus this algorithm can be widely used in system identification.

References

[1] X.H. Wang, F. Ding, Convergence of the auxiliary model based multi-innovation generalized extended stochastic gradient algorithm for Box-Jenkins systems, Nonlinear Dynamics, 2015, 82 (1-2) 269-280.

[2] J. Chen, Y. Zhang, R.F. Ding, Auxiliary model based multi-innovation algorithms for multivariable nonlinear systems, Mathematical and Computer Modelling, 2010, 52 (9-10)

1428-1434.

[3] G.C. Goodwin, K.S. Sin, Adaptive Filtering Prediction and Control, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[4] F. Ding, H.Z. Yang, F. Liu, Performance analysis of stochastic gradient algorithms under weak conditions, Science in China Series F–Information Sciences, 2008, 51 (9) 1269-1280.

[5] F. Ding, X.M. Liu, M.M. Liu, The recursive least squares identification algorithm for a class of Wiener nonlinear systems, Journal of the Franklin Institute, 2016, 353 (7) 1518-1526.

[6] F. Ding, P.X. Liu, G. Liu, Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises, Signal Processing, 2009, 89 (10) 1883-1890.

[7] F. Ding, T. Chen, Combined parameter and output estimation of dual-rate systems using an auxiliary model, Automatic, 2004, 40 (10) 1739-1748.

[8] F. Ding, T. Chen, Parameter estimation of dual-rate stochastic systems by using an output error method, IEEE Transactions on Automatic Control, 2005, 50 (9) 1436-1441.

[9] F. Ding, Y. Shi, T. Chen, Auxiliary model based least-squares identification methods for Hammerstein output-error systems, Systems & Control Letters, 2007, 56 (5) 373-380.

[10] Y.J. Liu, L. Xie, F. Ding, An auxiliary model based recursive least squares parameter estimation algorithm for non-uniformly sampled multirate systems, Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2009, 223 (4) 445-454.

[11] F. Ding, X.P. Liu, H.Z. Yang, Parameter identification and intersample output estimation for dual-rate systems, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 2008, 38 (4) 966-975.

[12] F. Ding, X.M. Liu, Y. Gu, An auxiliary model based least squares algorithm for a dual-rate state space system with time-delay using the data filtering. Journal of the Franklin Institute, 2016, 353 (2) 398-408.

[13] J. Chen, Several gradient parameter estimation algorithms for dual-rate sampled systems, Journal of the Franklin Institute, 2014, 351 (1) 543-554.