Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

A New Recursive Identification Method for Weighted Criterion

Guiming Luo, Yue Zhao, Yulai Zhang

School of Software, Tsinghua University
Tsinghua National Laboratory for Information Science and Technology,
Beijing, 100084, China,(email:gluo@tsinghua.edu.cn).

Abstract: In time-series system modeling traditional criterions only consider current estimation error while the past and global prediction errors are usually overlooked. By integrating both the estimation error and the difference between neighbor prediction errors, a novel weighted identification criterion was presented. Based on this criterion the alternations of the two-step algorithm is constructed through separating the system parameters estimation and noise parameters estimation for the system disturbed by colored noise, which could result in oscillation and instability. An extended one-step recursive algorithm for the weighted identification criterion is introduced in this paper. For the input-output system disturbed by colored noise, the prediction gradients and the gradient of the pseudo linear regression vector are given. The gradient iterative algorithm and the direct adaptive method (DAM), the new one-step recursive algorithm are proposed by a series of estimation process optimizations. Finally, a simulation example is conducted to demonstrate the efficiency of this new method.

1. INTRODUCTION

With the rapid development of computer technology, the systems quantitative analysis of real-time processing has been made possible. It has been used extensively to explore the theories and methods for signal modeling, and many modeling algorithms have been proposed. Combining the advantages of time-domain estimates and frequency-domain estimates, the classical frequency-domain optimal parameter (EFOP) method was recently introduced (Luo and Kwon, 2002, 2003; Luo, et. al. 2006; Luo and Huang, 2008). System updates are divided into two parts: the current prediction error and the past estimation error. The classical adaptive algorithms are time-invariant and the adaptive systems that are updated only depend on the prediction errors at the current time. Previous simulation results have shown that if the disturbed system is identified by such an adaptive algorithm, large estimation errors would be produced (Luo and Kimura, 2003). An effective algorithm should be able to deal with complex and comprehensive information, rather than just relying on a single source of data. Extended recursive algorithm (ERA), a new adaptive signal processing method which integrates the current and past prediction errors, is given by Luo, et al. (2006), Luo and Huang (2008), Luo and Kimura (2003, 2008). The version of the algorithm introduced a weighted matrix which updated in real-time (Luo and Huang, 2008; Luo and Kimura, 2003; and Luo, et al., 2008). The recursive algorithm was based on minimizing both the current and previous estimation errors. Compared to only considering recursive processes in the current estimation error, it has an advantage of anti-interference. The methods mentioned above consist of a weighted matrix and result in different recursive algorithms. However, how to choose the weighted matrix is a major issue which affects the system estimation. Adaptive algorithms are required to have certain guidelines to improve the weight selection and a macro-controlled way to verify the selection. A novel criterion was proposed that considers both estimation error and tuning for prediction errors of neighbor points (Zhao and Luo, 2009). Based on this criterion, an adaptive algorithm was presented to estimate the system parameters and noise parameters for the ARMAX model disturbed by colored noise (Zhao and Luo, 2008, 2009). A two-step adaptive algorithm was constructed through separating the parameters in both system parameters estimation and noise parameters estimation. In the recursive process, the estimation of system parameters was calculated through the previous moment, and the estimation of noise parameters was calculated relative to the ERA algorithm, and vice versa. However, the alternations of the two-step algorithm will result in oscillation and instability. There is a vulnerability that the algorithm will fluctuate within the system and result in convergence of the estimation of the two kinds of parameters. An extended one-step adaptive algorithm for the new criterion is introduced in this paper. By integrating system parameters with noise parameters, the prediction gradients can be given. For the model of a system disturbed by coloured noise, the regression vector contains the system parameters and noise parameters. Based on a series of estimation process optimization, the gradient of the regression vector is determined and a one-step adaptive algorithm is proposed. Finally, simulation examples were conducted to demonstrate the efficiency and accuracy of this new method.

⋆ This work was supported by the Funds NSFC61171121 and the Science Foundation of Chinese Ministry of Education - China Mobile 2012.

2. GRADIENT ITERATIVE ALGORITHM

Consider the following time-series model ARMAX:

$$A(q)y(t) = B(q)u(t) + C(q)w(t) \tag{1}$$

where $y(t)$, $u(t)$, and $w(t)$ are system output, input, and interference noise, respectively. $A(q)$, $B(q)$, and $C(q)$ are polynomials in the backward operator q^{-1}:

$$\begin{aligned} A(q) &= 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{n_a} q^{-n_a} \\ B(q) &= b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_b} q^{-n_b} \\ C(q) &= 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_{n_c} q^{-n_c} \end{aligned} \tag{2}$$

Let θ be the vector of the system parameters and noise parameters, and $\varphi_0(t)$ be the regression vector:

$$\theta = (a_1, a_2, \cdots, a_{n_a}, b_1, b_2, \cdots, b_{n_b}, c_1, c_2, \cdots, c_{n_c})^\tau$$

$$\begin{aligned}\phi_0(t) = (&-y(t-1), -y(t-2), \cdots, -y(t-n_a),\\ &u(t-1), u(t-2), \cdots, u(t-n_b),\\ &w(t-1), w(t-2), \cdots w(t-n_c))^\tau\end{aligned}$$

System (1) can then be rewritten as

$$y(t) = \theta^\tau \phi_0(t) + w(t) \tag{3}$$

If $w(t)$ is a white noise, the prediction of system (1) is (Ljung, 2002)

$$\hat{y}(t|\theta) = \frac{B(q)}{C(q)} u(t) + \left(1 - \frac{A(q)}{C(q)}\right) y(t)$$

or

$$C(q)\hat{y}(t|\theta) = B(q)u(t) + \left(C(q) - A(q)\right) y(t) \tag{4}$$

Then, output prediction of system (1) is

$$\hat{y}(t|\theta) = \phi(t, \theta)^\tau \theta \tag{5}$$

where

$$\begin{aligned}\phi(t, \theta) = (&-y(t-1), -y(t-2), \cdots, -y(t-n_a),\\ &u(t-1), u(t-2), \cdots, u(t-n_b),\\ &\bar{w}(t-1), \bar{w}(t-2), \cdots, \bar{w}(t-n_c))^\tau\end{aligned}$$

and $\bar{w}(t)$ is the estimation of the noise $w(t)$. Therefore, system (1) can be expressed as:

$$y(t) = \theta^\tau \phi(t, \theta) + w(t) \tag{6}$$

The prediction error $\varepsilon(t, \theta)$ is defined as

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta) \tag{7}$$

Consider the following performance index (Luo and Huang, 2008):

$$\begin{aligned}J(N) = &\lambda \sum_{t=1}^{N} \varepsilon(t, \theta)^2\\ &+\mu \left(\varepsilon(N, \theta)^2 + \sum_{t=1}^{N-1} \left(\varepsilon(t+1, \theta) - \varepsilon(t, \theta)\right)^2\right)\end{aligned} \tag{8}$$

Criterion (8) involves not only the current estimation error but also the change rate of estimation error at each step. λ is the weight for the current estimation error, while μ is the weight of the difference of estimation errors. Their choices depend on the specific requirements. Performance index (8) can be rewritten as Zhao and Luo (2009),

$$\begin{aligned}J(N, \theta) = &\left(\varepsilon(1, \theta), \varepsilon(2, \theta), \cdots \varepsilon(N, \theta)\right)\\ &\cdot Q(N) \left(\varepsilon(1, \theta), \varepsilon(2, \theta), \cdots \varepsilon(N, \theta)\right)^\tau\end{aligned} \tag{9}$$

where the weighted matrix $Q(N)$ is:

$$Q(N) = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN} \end{pmatrix} \in N \times N$$

and

$$\begin{cases} q_{ii} &= \lambda + 2\mu, \\ q_{i+1,i} &= \quad -\mu, \qquad i = 1, 2, \cdots, N \\ q_{i,i+1} &= \quad -\mu, \\ q_{jk} &= \quad 0, \qquad\qquad \text{others} \end{cases} \tag{10}$$

From relationships (5) and (6), it is easy to see that the prediction of system (1) is a pseudo linear regression rather than a linear regression. It is hard to obtain the optimal system parameters estimate directly from a weighted quadratic performance index. This is also the reason the two-step modeling method was adopted in Zhao and Luo (2009),. Our goal is to derive a reasonable direct adaptive algorithm. From (4)-(7), the gradient of the prediction with respect to the parameter θ is given as follows Ljung (2002):

$$\psi(t, \theta) = \frac{d}{d\theta}\hat{y}(t|\theta) = -\frac{d}{d\theta}\varepsilon(t|\theta) \tag{11}$$

and

$$\begin{aligned}C(q)\psi(t, \theta) &= C(q)\frac{d}{d\theta}\hat{y}(t|\theta)\\ &= \phi(t, \theta)\end{aligned} \tag{12}$$

where $\psi(t, \theta)$ is the gradient of system (1) or (4). Let

$$Z^t = \{(u(k), y(k)) | k \leq t\}$$

be a data set on input signals and output signals at time t. At time $t+1$, the data set is defined as

$$\begin{aligned}Z^{t+1} &= \{Z^t, (u(t+1), y(t+1))\}\\ &= \{(u(k), y(k)) | k \leq t+1\}\end{aligned}$$

Using numerical minimizing criterion (8) or (9), we can update the estimate of the optimal point iteratively. Then, parameter identification corresponding to (1) and (8) is given by the substitution method in Ljung (2002) and Dennis Jr and Schnabel (1983):

$$\hat{\theta}_t^{(i+1)} = \hat{\theta}_t^i + \delta f^{(i)}(t, \hat{\theta}_t^{(i)}, Z^t), \ \ i = 1, 2, 3, ... \tag{13}$$

where the subscript t denotes that the estimate is based on t data. The superscript (i) denotes the i-th iteration of the minimization procedure. $f^{(i)}$ is a search direction based on information about $J(N, \theta))$ acquired at previous iterations, and $\delta(0 < \delta \leq 1)$ is a positive constant determined so that an appropriate decrease in the value of $J(N, \theta))$ is obtained. The function gradient and Hessian matrix are chosen in the Newton direction:

$$f^{(i)}(t, \hat{\theta}_t^{(i)}, Z^t) = -\left(J''(t, \hat{\theta}_t^{(i)})\right)^{-1} J'(t, \hat{\theta}_t^{(i)}) \tag{14}$$

Where J' stands for the gradient w.r.t θ.The gradient iterative algorithm for system (1) based on criterion (8) consists of relationships (13) and (14).

3. DIRECT ADAPTIVE METHOD

The model gradient and the gradient iterative algorithm for system (1) based on performance index (8) have been given in the previous section. The search direction $f^{(i)}(t, \theta(i))$ is calculated by the numerical minimization, which is composed of the following gradients:

$$J'(t, \hat{\theta}_t^{(i)}), \quad \left(J''(t, \hat{\theta}_t^{(i)})\right)^{-1}, \quad i = 1, 2, 3, \cdots$$

If we want to establish the direct adaptive algorithm, it is necessary to compute the gradients for any estimation value of θ. From (14), we can see that the calculation of the search direction, especially the Hessian matrix calculation, is very complicated and difficult to use in practical modeling (Walther 2008). Moreover, it is difficult to determine the step number of the iterative algorithm. The accuracy of system modelling cannot be guaranteed, either. In addition, the method of implementation is off-line, so the model cannot be updated in real-time on a data set. The following examination establishes a direct adaptive algorithm for system (1) and performance index (8). For iteration $t+1$, we introduce the following:

$$\hat{\theta}^{t+1}_{t+1} = \theta(t+1),\ \hat{\theta}^{t}_{t+1} = \theta(t)$$

Then, gradient algorithms (13) and (14) can be approximately expressed as follows:

$$\theta(t+1) = \theta(t) - \delta(J''(t+1,\theta(t)))^{-1}J'(t+1,\theta(t)) \quad (15)$$

From criterion (8), we have

$$\begin{aligned} &J(t,\theta) \\ &= J(t-1,\theta) + (\lambda+2\mu)\varepsilon(t,\theta)^2 - 2\mu\varepsilon(t,\theta)\varepsilon(t-1,\theta) \end{aligned}$$

Then, from (11) and the above relationship, the gradient of performance index $J(t,\theta)$ with respect to the parameter θ is

$$\begin{aligned} J'(t,\theta) &= \frac{dJ(t,\theta)}{d\theta} \\ &= J'(t-1,\theta) - 2(\lambda+2\mu)\varepsilon(t,\theta)\psi(t,\theta) \\ &\quad + 2\mu\left(\varepsilon(t,\theta)\psi(t-1,\theta)\right. \\ &\quad \left.+ \varepsilon(t-1,\theta)\psi(t,\theta)\right) \end{aligned} \quad (16)$$

Since the parameter estimates of system (1) are obtained by minimizing performance index $J(t,\theta)$, when $\theta = \theta(t)$ it should hold that

$$J'(t,\theta(t)) = 0$$

From (16), we have

$$\begin{aligned} &J'(t+1,\theta(t)) \\ &= \left.\frac{dJ(t+1,\theta)}{d\theta}\right|_{\theta=\theta(t)} \\ &= 2\mu\left(\varepsilon(t+1,\theta(t))\psi(t,\theta(t))\right. \\ &\quad - 2(\lambda+2\mu)\varepsilon(t+1,\theta(t))\psi(t+1,\theta(t)) \\ &\quad \left.+\varepsilon(t,\theta(t))\psi(t+1,\theta(t))\right) \end{aligned} \quad (17)$$

Combining (15) and (17), it follows that

$$\begin{aligned} &\theta(t+1) = \theta(t) \\ &- \delta(J''(t+1,\theta(t)))^{-1}\left[2\mu\left(\varepsilon(t+1,\theta(t))\psi(t,\theta(t))\right)\right) \\ &- 2(\lambda+2\mu)\varepsilon(t+1,\theta(t))\psi(t+1,\theta(t)) \\ &+ \varepsilon(t,\theta(t))\psi(t+1,\theta(t))] \end{aligned} \quad (18)$$

To estimate the system parameters, it is necessary to calculate Hessian matrix $J''(t+1,\theta(t))$. First, we want to discuss the gradient of $\psi(t,\theta)$ with respect to the parameter θ. Lemma 1. For system (1), the gradient of the vector $\psi(t,\theta)$ with respect to the parameter θ is

$$\frac{d}{d\theta}\psi(t,\theta) = -\frac{1}{C(q)^2}\left(T + T^{\tau}\right)$$

Here, T is the matrix

$$T = \left(\underbrace{0,\cdots,0}_{n_a+n_b}, \phi(t-1,\theta), \phi(t-2,\theta), \cdots, \phi(t-n_c,\theta)\right)$$

Proof: From (12):

$$\begin{aligned} \psi(t,\theta)^{\mathrm{T}} = &-\frac{y(t-1)}{C(q)}, -\frac{y(t-2)}{C(q)}, \cdots, \\ &-\frac{y(t-n_a)}{C(q)}, \frac{u(t-1)}{C(q)}, \frac{u(t-2)}{C(q)}, \cdots, \\ &\frac{u(t-n_b)}{C(q)}, \frac{\varepsilon(t-1,\theta)}{C(q)}, \frac{\varepsilon(t-2,\theta)}{C(q)}, \cdots, \\ &\frac{\varepsilon(t-n_c,\theta)}{C(q)} \end{aligned} \quad (19)$$

When $i = 1,2,3,\cdots,n_a$, it follows that

$$\frac{\partial}{\partial a_i}\left(\frac{y(t-m)}{C(q)}\right) = 0, \quad m = 1,2,\cdots,n_a$$

$$\frac{\partial}{\partial a_i}\left(\frac{u(t-r)}{C(q)}\right) = 0, \quad r = 1,2,\cdots,n_b$$

$$\begin{aligned} &\frac{\partial}{\partial a_i}\left(\frac{\varepsilon(t-s,\theta)}{C(q)}\right) \\ &\qquad = \frac{1}{C(q)}\frac{\partial\varepsilon(t-s,\theta)}{\partial a_i} \\ &\qquad = -\frac{1}{C(q)}\frac{\partial}{\partial a_i}\hat{y}(t-s|\theta), \\ &s = 1,2,\cdots,n_c \end{aligned}$$

When $j = 1,2,3,\cdots,n_b$, it follows that

$$\frac{\partial}{\partial b_j}\left(\frac{y(t-m)}{C(q)}\right) = 0, \quad m = 1,2,\cdots,n_a$$

$$\frac{\partial}{\partial b_j}\left(\frac{u(t-r)}{C(q)}\right) = 0, \quad r = 1,2,\cdots,n_b$$

$$\begin{aligned} \frac{\partial}{\partial b_j}\left(\frac{\varepsilon(t-s,\theta)}{C(q)}\right) &= \frac{1}{C(q)}\frac{\partial\varepsilon(t-s,\theta)}{\partial b_j} \\ &= -\frac{1}{C(q)}\frac{\partial}{\partial b_j}\hat{y}(t-s|\theta), \end{aligned}$$

$$s = 1,2,\cdots,n_c$$

If $k = 1,2,3,\cdots,n_c$, it follows that

$$\frac{\partial}{\partial c_k}\left(\frac{y(t-m)}{C(q)}\right) = -\frac{q^{-k}y(t-m)}{C(q)^2}, \quad m = 1,2,\cdots,n_a$$

$$\frac{\partial}{\partial c_k}\left(\frac{u(t-r)}{C(q)}\right) = -\frac{q^{-k}u(t-r)}{C(q)^2}, \quad r = 1,2,\cdots,n_b$$

$$\begin{aligned} &\frac{\partial}{\partial c_k}\left(\frac{\varepsilon(t-s,\theta)}{C(q)}\right) \\ &= -\frac{1}{C(q)}\frac{\partial}{\partial c_k}\hat{y}(t-s|\theta) - \frac{q^{-k}\varepsilon(t-s,\theta)}{C(q)^2}, \\ &s = 1,2,\cdots,n_c \end{aligned}$$

Combining the above derivation, for $m = 1,2,3,\cdots,n_a$, $r = 1,2,3,\cdots,n_b$, and $s = 1,2,3,\cdots,n_c$ we have

$$\begin{aligned} &\frac{d}{d\theta}^{\mathrm{T}}\left(\frac{y(t-m)}{C(q)}\right) \\ &= \left[\underbrace{0,\cdots,0}_{n_a+n_b}, -\frac{q^{-1}y(t-m)}{C(q)^2}, -\frac{q^{-2}y(t-m)}{C(q)^2},\right. \\ &\qquad \left.\cdots -\frac{q^{-n_c}y(t-m)}{C(q)^2}\right]^{\tau} \end{aligned} \quad (20)$$

$$\begin{aligned} &\frac{d}{d\theta}^{\mathrm{T}}\left(\frac{u(t-r)}{C(q)}\right) \\ &= \left[\underbrace{0,\cdots,0}_{n_a+n_b}, -\frac{q^{-1}u(t-r)}{C(q)^2}, -\frac{q^{-2}u(t-r)}{C(q)^2},\right. \\ &\qquad \left.\cdots -\frac{q^{-n_c}u(t-r)}{C(q)^2}\right]^{\tau} \end{aligned} \quad (21)$$

$$\begin{aligned}
&\frac{d}{d\theta}\left(\frac{\varepsilon(t-s,\theta)}{C(q)}\right)\\
&=\Big[-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial a_1},-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial a_2},\cdots,\\
&\quad-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial a_{n_a}},-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial b_1},\\
&\quad-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial b_2},\cdots,-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial b_{n_b}},\\
&\quad\cdots,-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial c_1}-\frac{q^{-1}\varepsilon(t-s,\theta)}{C(q)^2},\\
&\quad-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial c_2}-\frac{q^{-2}\varepsilon(t-s,\theta)}{C(q)^2},\cdots,\\
&\quad-\frac{1}{C(q)}\frac{\partial\hat{y}(t-s|\theta)}{\partial c_{n_c}}-\frac{q^{-n_c}\varepsilon(t-s,\theta)}{C(q)^2}\Big]^\tau\\
&=-\frac{1}{C(q)^2}\phi(t-s,\theta)-\frac{1}{C(q)^2}\\
&\quad\cdot\Big[\underbrace{0,\cdots,0}_{n_a+n_b},q^{-1}\varepsilon(t-s,\theta),\cdots,q^{-n_c}\varepsilon(t-s,\theta)\Big]^\tau
\end{aligned}\tag{22}$$

and

$$\begin{aligned}
&\Big[\frac{q^{-s}y(t-1)}{C(q)^2},\frac{q^{-s}y(t-2)}{C(q)^2},\cdots,\\
&\quad\frac{q^{-s}y(t-n_a)}{C(q)^2},-\frac{q^{-s}u(t-1)}{C(q)^2},\\
&\quad-\frac{q^{-s}u(t-2)}{C(q)^2}\cdots,-\frac{q^{-s}u(t-n_b)}{C(q)^2},\\
&\quad-\frac{q^{-s}\varepsilon(t-1,\theta)}{C(q)^2},-\frac{q^{-s}\varepsilon(t-2,\theta)}{C(q)^2},\cdots,\\
&\quad-\frac{q^{-s}\varepsilon(t-n_c,\theta)}{C(q)^2}\Big]^\tau\\
&=-\frac{q^{-s}}{C(q)^2}\phi(t,\theta)
\end{aligned}\tag{23}$$

From (19)-(23), we have

$$\begin{aligned}
&C(q)^2\frac{d}{d\theta}\psi(t,\theta)\\
&=-\left(\underbrace{0,\cdots,0}_{n_a+n_b},\phi(t-1,\theta),\phi(t-2,\theta),\cdots,\phi(t-n_c,\theta)\right)\\
&\quad-\begin{pmatrix}\left.\begin{matrix}0\\ \vdots\\ 0\end{matrix}\right\}n_a+n_b\\ \phi(t-1,\theta)^\tau\\ \phi(t-2,\theta)^\tau\\ \vdots\\ \phi(t-n_c,\theta)^\tau\end{pmatrix}
\end{aligned}\tag{24}$$

The result of Lemma 1is followed. Denote $\mathcal{F}_n$ as σ -algebra generated by the data $\{w_i,u_j\|i,j\le n\}$. From (3)-(6), it is easy to see that the pseudo linear regression vector $\phi(t,\theta)$ and prediction $\hat{y}(t,\theta)$ are $\mathcal{F}_{n-1}$ measurable, so the gradient $\psi(t,\theta)$ is also $\mathcal{F}_{n-1}$ measurable. According to Lemma 1, the gradient of $\psi'(t,\theta)$ is $\mathcal{F}_{n-2}$ measurable. From (16):

$$\begin{aligned}
&J''(t,\theta)=J''(t-1,\theta)-\\
&2(\lambda+2\mu)\left(\frac{d\varepsilon(t,\theta)}{d\theta}\psi(t,\theta)^\tau+\varepsilon(t,\theta)\frac{d\psi(t,\theta)}{d\theta}\right)\\
&+2\mu\frac{d\varepsilon(t,\theta)}{d\theta}\psi(t-1,\theta)^\tau+2\mu\varepsilon(t,\theta)\frac{d\psi(t-1,\theta)}{d\theta}\\
&+2\mu\frac{d\varepsilon(t-1,\theta)}{d\theta}\psi(t,\theta)^\tau+2\mu\varepsilon(t-1,\theta)\frac{d\psi(t,\theta)}{d\theta}
\end{aligned}\tag{25}$$

Using (5) and (7) again:

$$\psi(t,\theta)=\frac{d}{d\theta}\hat{y}(t|\theta)=-\frac{d}{d\theta}\varepsilon(t|\theta)$$

Then, by (25) and the above relationship, for $k=1,2,\cdots,r$:

$$\begin{aligned}
J''(k,\theta)&=J''(k-1,\theta)+2(\lambda+2\mu)\psi(k,\theta)\psi(k,\theta)^\tau\\
&-2\mu\left(\psi(k,\theta)\psi(k-1,\theta)^\tau+\psi(k-1,\theta)\psi(k,\theta)^\tau\right)\\
&+\left(2\mu\varepsilon(k-1,\theta)-2(\lambda+2\mu)\varepsilon(k,\theta)\right)\frac{d\psi(k,\theta)}{d\theta}\\
&+2\mu\varepsilon(k,\theta)\frac{d\psi(k-1,\theta)}{d\theta}
\end{aligned}$$

Summing up with respect to k on both sides of the above relation, we get

$$\begin{aligned}
&J''(t,\theta)=2(\lambda+2\mu)\sum_{k=1}^{\tau}\psi(k,\theta)\psi(k,\theta)^\tau\\
&-2\mu\sum_{k=1}^{t}\left(\psi(k,\theta)\psi(k-1,\theta)^\tau+\psi(k-1,\theta)\psi(k,\theta)^\tau\right)\\
&-\sum_{k=1}^{t}\left(2(\lambda+2\mu)\varepsilon(k,\theta)-2\mu\varepsilon(k-1,\theta)\right)\frac{d\psi(k,\theta)}{d\theta}\\
&+2\mu\sum_{k=1}^{t}\varepsilon(k,\theta)\frac{d\psi(k-1,\theta)}{d\theta}
\end{aligned}$$

where we denote that $J''(0,\theta)=0$. Then

$$\begin{aligned}
&\frac{1}{t}J''(t,\theta)=\frac{2(\lambda+2\mu)}{t}\sum_{k=1}^{t}\psi(k,\theta)\psi(k,\theta)^\tau\\
&-\frac{2\mu}{t}\sum_{k=1}^{t}\left(\psi(k,\theta)\psi(k-1,\theta)^\tau+\psi(k-1,\theta)\psi(k,\theta)^\tau\right)\\
&-\frac{2}{t}\sum_{k=1}^{t}\left((\lambda+2\mu)\varepsilon(k,\theta)-\mu\varepsilon(k-1,\theta)\right)\frac{d\psi(k,\theta)}{d\theta}\\
&+\frac{2\mu}{t}\sum_{k=1}^{t}\varepsilon(k,\theta)\frac{d\psi(k-1,\theta)}{d\theta}\\
&\approx\frac{2(\lambda+2\mu)}{t}\sum_{k=1}^{t}\psi(k,\theta)\psi(k,\theta)^\tau\\
&-\frac{2\mu}{t}\sum_{k=1}^{t}\left(\psi(k,\theta)\psi(k-1,\theta)^\tau+\psi(k-1,\theta)\psi(k,\theta)^\tau\right)\\
&-E\left(2\left((\lambda+2\mu)\varepsilon(t,\theta)-\mu\varepsilon(t-1,\theta)\right)\frac{d\psi(t,\theta)}{d\theta}\right)\\
&+2\mu E\left(\varepsilon(t,\theta)\frac{d\psi(t-1,\theta)}{d\theta}\right)
\end{aligned}\tag{26}$$

From (5)-(7) it can be seen that the prediction error vector $\varepsilon(t,\theta)$ is consistent with noise $w(t)$. According to Lemma 1, the gradient $\psi'(t,\theta)$ is $\mathcal{F}_{n-2}$-measurable and $\psi'(t-1,\theta)$ is $\mathcal{F}_{n-3}$-measurable. If the error sequences are white noises, then

$$E\left(\varepsilon(t)\frac{\partial\psi(t,\theta)}{\partial\theta}\right)=0,$$
$$E\left(\varepsilon(t-1)\frac{\partial\psi(t,\theta)}{\partial\theta}\right)=0,$$
$$E\left(\varepsilon(t)\frac{\partial\psi(t-1,\theta)}{\partial\theta}\right)=0.$$

If t is large enough, from (26) and the above relationships, we have the following:

$$\begin{aligned}\frac{1}{t}J''(t,\theta) \approx& \frac{2(\lambda+2\mu)}{t}\sum_{k=1}^{t}\psi(k,\theta)\psi(k,\theta)^{\tau} \\ &-\frac{2\mu}{t}\sum_{k=1}^{t}(\psi(k,\theta)\psi(k-1,\theta)^{\tau} \\ &+\psi(k-1,\theta)\psi(k,\theta)^{\tau})\end{aligned} \tag{27}$$

Combining (18) and (27), it follows that

$$\begin{aligned}\theta(t+1) =& \theta(t)- \\ &\delta(2(\lambda+2\mu)\sum_{k=1}^{t+1}\psi(k,\theta(t))\psi(k,\theta(t))^{\tau} \\ &-2\mu\sum_{k=1}^{t+1}(\psi(k,\theta(t))\psi(k-1,\theta(t))^{\tau} \\ &+\psi(k-1,\theta(t))\psi(k,\theta(t))^{\tau}))^{-1} \\ &[2\mu\left(\varepsilon(t+1,\theta(t))\psi(t,\theta(t))\right. \\ &\left.+\varepsilon(t,\theta(t))\psi(t+1,\theta(t))\right) \\ &-2\left(\lambda+2\mu\right)\varepsilon(t+1,\theta(t))\psi(t+1,\theta(t))]\end{aligned} \tag{28}$$

In algorithm (28), the parameter θ in $\psi(t,\theta)$ and $\hat{y}(t|\theta)$ is replaced by the recursively computed quantity $\theta(t)$, which is estimated on the data Z^t. Thus, we can denote that

$$\psi(t,\theta(t))=\phi(t),\ \hat{y}(t,\theta(t))=\hat{y}(t),\ \varepsilon(t,\theta(t-1))=\varepsilon(t)$$

Then, algorithm (28) can be expressed as

$$\begin{aligned}\theta(t+1) =& \theta(t)-\delta\left(2(\lambda+2\mu)\sum_{k=1}^{t+1}\phi(k)\phi(k)^{\tau}\right. \\ &\left.-2\mu\sum_{k=1}^{t+1}(\phi(k)\phi(k-1)^{\tau}+\phi(k-1)\phi(k)^{\tau})\right)^{-1} \\ &[2\mu\left(\varepsilon(t+1)\phi(t)+\varepsilon(t,\theta(t))\phi(t+1)\right) \\ &-2(\lambda+2\mu)\varepsilon(t+1)\phi(t+1)] \\ =& \theta(t)-\delta P^{-1}(t+1)E(t+1)\end{aligned} \tag{29}$$

where $0<\delta\leqslant 1$, and

$$\begin{aligned}P(t) =& \sum_{k=1}^{t}((\lambda+2\mu)\phi(k)\phi(k)^{\tau} \\ &-\mu(\phi(k)\phi(k-1)^{\tau}+\phi(k-1)\phi(k)^{\tau})) \\ E(t) =& \mu\left(\varepsilon(t)\phi(t-1)+\varepsilon(t-1)\phi(t)\right) \\ &-(\lambda+2\mu)\varepsilon(t)\phi(t)\end{aligned}$$

Lemma 2. If $A\in R^{n\times n}$, $C,D\in R^{n\times 1}$, $q\in R$, and matrix A is nonsingular, then (Luo and Kimura, 2003)

$$\begin{aligned}&(A+CD^{\tau}+DC^{\tau}+qDD^{\tau})^{-1} \\ =& A^{-1}+(ab)^{-1}A^{-1}\left(D^{\tau}A^{-1}DCC^{\tau}-\sigma DD^{\tau}\right)A^{-1} \\ &-b^{-1}A^{-1}\left(CD^{\tau}+DC^{\tau}\right)A^{-1}\end{aligned}$$

where

$$a=1+D^{\tau}A^{-1}C$$
$$b=a+a^{-1}\sigma D^{\tau}A^{-1}D$$
$$\sigma=q-C^{\tau}A^{-1}C$$

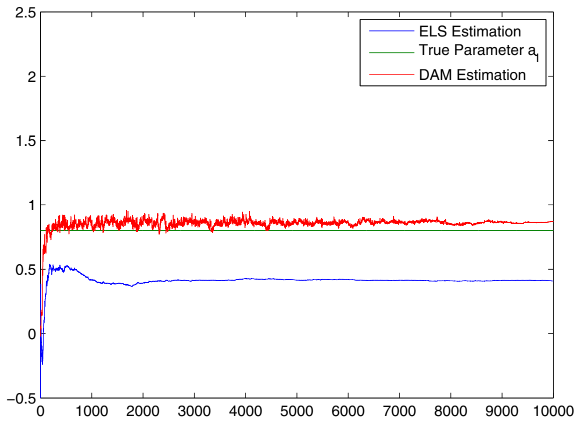

Fig. 1. Estimate for Parameters a_1

Since

$$\begin{aligned}P(t) =& P(t-1)+(\lambda+2\mu)\phi(t)\phi(t)^{\tau} \\ &-\mu\left(\phi(t)\phi(t-1)^{\tau}+\phi(t-1)\phi(t)^{\tau}\right)\end{aligned}$$

we can denote that $P(t-1)=A$, $D=\phi(t)$, $C=\mu\phi(t-1)$, and $q=\lambda+2\mu$. By Lemma 1 and 2, the matrix $P(t)$ and parameter estimation can be calculated. Then the following algorithm can be obtained.

DAM Algorithm: The direct adaptive method (DAM) algorithm of system (1) with respect criterion (8) yields

$$\begin{cases}P^{-1}(N)=P^{-1}(N-1) \\ \quad+\dfrac{\mu}{b(N)}P^{-1}(N-1)\left(\phi(N)\phi^{\tau}(N-1)\right. \\ \quad\left.+\phi(N-1)\phi^{\tau}(N)\right)P^{-1}(N-1) \\ \quad+\dfrac{P^{-1}(N-1)}{a(N)b(N)}\left(\mu^2\phi^{\tau}(N)\right. \\ \quad P^{-1}(N-1)\phi(N)\phi(N-1)\phi^{\tau}(N-1) \\ \quad\left.-\sigma(N)\phi(N)\phi^{\tau}(N)\right)P^{-1}(N-1) \\ \theta(N)=\theta(N-1)- \\ \quad\delta\left[\mu\left(\phi(N-1)(y(N)-\phi(N)\theta(N-1))\right.\right. \\ \quad\left.+\phi(N)(y(N-1)-\phi(N-1)\theta(N-2))\right) \\ \quad\left.-(\lambda+2\mu)\phi(N)(y(N)-\phi(N)\theta(N-1))\right]\end{cases} \tag{30}$$

where

$$a(N)=1-\mu\phi^{\tau}(N)P^{-1}(N-1)\phi(N-1)$$
$$b(N)=a(N)+a^{-1}(N)\sigma(N)\phi^{\tau}(N)P^{-1}(N-1)\phi(N)$$
$$\sigma(N)=\lambda+2\mu-\mu^2\phi^{\tau}(N-1)P^{-1}(N-1)\phi(N-1)$$

4. SIMULATIONS

Consider the following system:

$$(1+a_1q^{-1})y(t)=b_1q^{-1}u(t)+(1+c_1q^{-1})w(t) \tag{31}$$

The real parameters are: $a_1=0.8$, $b_1=1.2$, $c_1=0.1$. The input signal $u(t)$ was generated by a random number with variance $\sigma=1$. The disturbed noise $w(t)$ was also a random number with variance $\sigma=0.6573$. Both random numbers are not white noises. The sample number is 10000. The parameters were estimated according to both the ELS algorithm and the DAM algorithm (30), where the values of λ , μ in (30) are chosen as $\lambda=0.2$, $\mu=0.8$. The simulation results are shown in Figs. 1- 3.

The DAM can be intuitively compared with the ELS method in Fig 1 - Fig 3, which clearly shows that the DAM algorithm

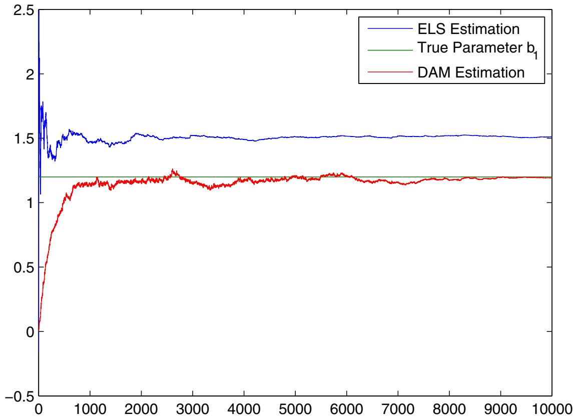

Fig. 2. Estimate for Parameters b_1

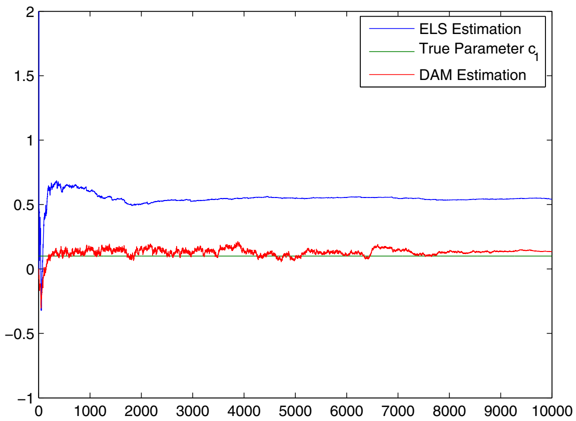

Fig. 3. Estimate for Parameters c_1

identifies real system (31) more efficiently than do the ELS method. These are also validated by the following calculated values:

$$\bar{\theta}_{ELS} = \begin{pmatrix} 0.4132 \pm 0.0030 \\ 1.5052 \pm 0.0029 \\ 0.5471 \pm 0.0045 \end{pmatrix}$$

$$\bar{\theta}_{DAM} = \begin{pmatrix} 0.8544 \pm 0.0031 \\ 1.1453 \pm 0.0174 \\ 0.1278 \pm 0.0014 \end{pmatrix}$$

$\bar{\theta}_{ELS}$ denotes the average estimate from the first LS estimate value to the 10000th LS estimate, while $\bar{\theta}_{DAM}$ denotes the average estimate from the first DAM estimate value to the 10000th DAM estimate, respectively. The calculational error is defined by the standard deviation. The accumulated estimate error in the DAM algorithm is 0.082, while the accumulated estimate error of the ELS method is 0.6653.

5. CONCLUSIONS

Estimation errors can significantly increase when systems are disturbed by complex noise. The two-step adaptive algorithm may be a divergence in the risk identification process because of the identification of the estimation error signal instead of the noise, and also because of the substitution of the noise parameters into the system parameter vector recursive algorithm. Identification for input-output systems disturbed by colored noise is the main focus of this paper. Based on integrating both the estimation error and the difference between neighbor prediction errors, a weighted identification criterion was recently introduced. An extended one-step adaptive algorithm based on this novel criterion was examined. Several prediction gradients were obtained for the pseudo linear regression vector. The gradient iterative algorithm and the direct adaptive algorithm, a new one-step algorithm for the time-series model based on the weighted criterion, were established. This new algorithm has the advantages of reducing operation costs, avoiding the risk of divergence of the parameters estimation, and strengthening the anti-interference properties of delivery systems with colored noise. In the end, several simulations demonstrated the efficiency and accuracy of this new method.

REFERENCES

Dennis Jr. J. E. and R. B. Schnabel Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Upper Saddle River, NJ: Prentice-Hall, 1983.

Lo, K. M. and W. H. Kwon A new identification approach for FIR models, *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol.49(6): 439-446, 2002.

Lo, K. M. and W. H. Kwon New Identification Approaches for Disturbed Models. *Automatica*, vol.39: 1627-1634, 2003.

Lo, K. M. H. Kimura, W. H. Kwon, and X. J. Yang . Empirical Frequency-domain Optimal Parameter Estimate for Black-box Processes. *IEEE Trans. Circuits and Systems I: Regular Papers*, vol.53(2): 419-430, 2006.

Luo, G. M. and J. Huang Recursive algorithm for the two-stage EFOP estimation method *Science in China, F*, vol.51(2): 145-157,2008.

Lo, K. M. and H. Kimura Recursive estimation methods for discrete systems. *IEEE Transactions on Automatic Control*, vol.48: 2019-2024, 2003.

Lo, K. M. H. Kimura, and W. H. Kwon Recursive identification algorithms based on minimizing estimation error. *Proceedings of 17th IFAC Congress*, Seoul, Korea, vol.17, pages 7510-7515, 2008.

L. Ljung, System Identification-Theory for the User, Upper Saddle River, NJ: Prentice-Hall, 2002.

Walther, A. Computing sparse Hessians with automatic differentiation. *ACM Transactions on Mathematical Software*, vol.34(3), 1-15, 2008.

Zhao, Y., K. M. Lo, and W. H. Kwon, Identification of multi-dimensional system based on a novel criterion, *Proceedings of ICINCO 2008*, pages 143-148, 2008.

Zhao, Y. and G. M. Luo Recursive algorithms for parameter identification of systems disturbed by coloured noise, *Journal of Tsinghua University*, vol.49(1): 135-141, 2009.