International Journal of Automation and Computing 9(5), October 2012, 501-505
DOI: 10.1007/s11633-012-0673-x

Identification Algorithm Based on the Approximate Least Absolute Deviation Criteria

Bao-Chang Xu Xin-Le Liu
Department of Automation, China University of Petroleum, Beijing 102249, China

Abstract: Considering the situation that the least-squares (LS) method for system identification has poor robustness and the least absolute deviation (LAD) algorithm is hard to construct, an approximate least absolute deviation (ALAD) algorithm is proposed in this paper. The objective function of ALAD is constructed by introducing a deterministic function to approximate the absolute value function. Based on the function, the recursive equations for parameter identification are derived using Gauss-Newton iterative algorithm without any simplification. This algorithm has advantages of simple calculation and easy implementation, and it has second order convergence speed. Compared with the LS method, the new algorithm has better robustness when disorder and peak noises exist in the measured data. Simulation results show the efficiency of the proposed method.

Keywords: System identification, least absolute deviation (LAD), Gauss-Newton algorithm, robustness, disorder and peak noise.

1 Introduction

The least-squares (LS) method has been widely used in the field of system identification because of its mature theories[1−4]. A series of new identification methods based on the LS method have been proposed by many researchers[5−8]. It is well known that if random noises are normally distributed, the LS criteria can obtain excellent performance[9]. However, if abnormal data exist in the measured data (such as disorder and peak noises), the LS criteria is unable to obtain good identification result. In fact, abnormal data existing in the measured data can cause large error and the squared value of the error becomes larger. Thus, the impact of the abnormal data is inflated when using the LS objective function. The estimated parameters by minimizing the sum of squared errors have to fit with these abnormal data, which worsen the identification robustness and convergence speed. Moreover, in many cases, abnormal points just reflect some specific information, so they cannot be removed arbitrarily[10].

For the shortcoming of LS methods, least absolute deviation (LAD) criteria which uses the minimum sum of absolute errors as the objective function is adopted by many researchers in order to improve the robustness and convergence speed of identification algorithm[11−13]. However, the main weakness of LAD criteria is that the objective function is non-differentiable. We need to solve a non-smooth optimization problem, which is more difficult. Reference [14] shows an iterative algorithm to find the linear LAD regression coefficient, which lists all the possibilities. But the convergence condition of the algorithm is difficult to determine, and the initial iteration vector is given only depending on experiences. Linear model is analyzed by translating it to linear programming problem, and the linear regression coefficients under LAD criteria can be given by solving linear programming in Matlab quickly and correctly[15]. However, the computation cost of the method is too high in the presence of large samples. Xu et al.[16] used the majorize-minimize (MM) algorithm to solve parameter identification problem under LAD criteria, but the calculation process is too complicated and not suitable for practical application. For nonlinear Wiener-Hammerstein model, a new method under LAD criteria is proposed by using particle swarm optimization (PSO) algorithm for nonlinear system identification[17]. However, it is not convenient to apply because it needs to set many parameters.

Considering the drawbacks mentioned above, identification algorithm based on approximate least absolute deviation criteria (ALAD) is proposed in this paper. The objective function is constructed by using a deterministic function to replace the absolute value function. The deterministic function has similar characteristics as the absolute value function is differentiable. Model parameter identification is done based on the objective function, and the recursive equations for parameters identification are derived using Gauss-Newton iterative algorithm without any simplification. This algorithm has the advantages of simple calculation and easy implementation, and it compensates the drawbacks of the LAD method whose objective function is non-differentiable and involves complicated calculation. Furthermore, compared with the LS method, this algorithm is more robust when disorder and peak noises exist in the measured data. Simulation results show the efficiency of the new algorithm.

The remaining part of the paper is organized as follows: A deterministic function used to approximate the absolute value function is described in Section 2. Section 3 gives the details of the new algorithm proposed in this paper. Simulation results and analysis are presented in Section 4. Finally, the conclusions are presented in Section 5.

Notations.
β: A controllable parameter.
$u(k)$: System input.
$y(k)$: System input output.
$v(k)$: System noise.
$A(z^{-1}), B(z^{-1})$: Polynomials of unit delay operator z^{-1}.

Manuscript received November 23, 2011; revised March 14, 2012
This work was supported by Important National Science & Technology Specific Projects (No. 2011ZX05021-003)

a_i, b_j: Parameters to be estimated.
n_a, n_j: Order of polynomials.
θ: System parameter vector.
$\hat{\theta}$: Estimated parameter vector.
$\varphi(k)$: Data vector.
$J(*)$: Cost function.
$e(k)$: Estimation error.
G: Gradient of $J(\hat{\theta})$.
H: Hessian matrix of $J(\hat{\theta})$.

2 Description of approximate absolute function

A deterministic function which has characteristics of the absolute value function is introduced in this paper to approximate the absolute value function, which compensates the shortcomings of LAD method. The function can be describe as

$$\phi(x) = \beta \ln[\cosh(\frac{x}{\beta})] \tag{1}$$

where β is a controllable parameter, and nonlinear function $\phi(*)$ is related with β.

When β is small enough, nonlinear function $\cosh(\frac{x}{\beta}) \approx \frac{1}{2}\mathrm{e}^{|x/\beta|}$, $\beta \ln[\cosh(\frac{x}{\beta})] \approx |x|$, i.e., $\phi(x) \approx |x|$ apprioximates the absolute value function.

When $\beta = 0.01$, the function curve is shown in Fig. 1.

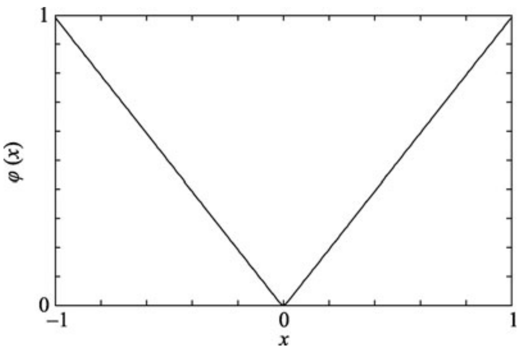

Fig. 1 Curve of the function, $\beta = 0.01$

The curve shows that when β is small enough, nonlinear function $\phi(x) = \beta \ln[\cosh(\frac{x}{\beta})]$ can approximate the absolute value function $|x|$ effectively.

3 The identification algorithm based on approximate LAD criteria

3.1 The model description

The system under consideration is assumed to be represented by

$$A(z^{-1})y(k) = B(z^{-1})u(k) + v(k) \tag{2}$$

where $u(k)$ and $y(k)$ are the system input, and output respectively, $A(z^{-1})$ and $B(z^{-1})$ are polynomials of the unit delay operator z^{-1} (i.e., $z^{-1}y(k) = y(k-1)$), $A(z^{-1})$ and $B(z^{-1})$ are defined as

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{na} z^{-na}$$

$$B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{nb} z^{-nb}$$

where $a_i (i = 1, 2, \cdots, n_a)$, $b_i (i = 1, 2, \cdots, n_b)$ are the parameters to be estimated.

The system represented by (2) can be rewritten in a signal-regressive form

$$y(k) = \varphi^{\mathrm{T}}(k)\theta + v(k) \tag{3}$$

where

$$\varphi(k) = [-y(k-1), \cdots, -y(k-n_a), u(k-1), \cdots, u(k-n_b)]^{\mathrm{T}}$$
$$\theta = [a_1, \cdots, a_{na}, b_1, \cdots, b_{nb}]^{\mathrm{T}}.$$

The problem is to estimate the parameter vector θ from the available data $y(k)$, $\varphi(k)$, $k = 1, 2, \cdots, m$. The estimated θ should converge to the true value as the sample size tends to infinity. The criteria used in this paper is defined as

$$J(\hat{\theta}) = \sum_{k=1}^{m} \beta \ln[\cosh(\frac{e(k)}{\beta})] \tag{4}$$

where $e(k) = y(k) - \varphi^{\mathrm{T}}(k)\hat{\theta}$.

The identification algorithm proposed in this paper is used to search system parameters θ_{LS} to satisfy $J(\theta_{LS})$ is minimum.

3.2 Identification algorithm based on ALAD criteria

Lemma 1. Given the system represented by (3), the recursive equations for estimated system parameters $\hat{\theta}$ that minimize the objective function (4) are derived as follows

$$\hat{\theta}(m) = \hat{\theta}(m-1) + \alpha \Delta\hat{\theta}(m) \tag{5}$$

$$\Delta\hat{\theta}(m) = \Delta\hat{\theta}(m-1) - W(m)\{\varphi^{\mathrm{T}}(m)\Delta\hat{\theta}(m-1) - z(m)\tanh[\frac{e(m)}{\beta}]\} \tag{6}$$

$$W(m) = P(m-1)\varphi(m)[z(m) + \varphi^{\mathrm{T}}(m)P(m-1)\varphi(m)]^{-1} \tag{7}$$

$$P(m) = P(m-1) - W(m)\varphi^{\mathrm{T}}(m)P(m-1) \tag{8}$$

$$z(m) = \frac{1}{1 - \tanh^2 \frac{e(m)}{\beta}}$$

$$e(m) = y(m) - \varphi^{\mathrm{T}}(m)\hat{\theta}(m-1) \tag{9}$$

where α is the learning rate, $P(0) = \omega I$, $\omega = 10^6$, $\hat{\theta}(0)$ and $\Delta\hat{\theta}(0)$ are initialized as small real vectors.

Proof. Using Gauss-Newton optimization method to solve the minimization problem (4), the recursive equation for parameter identification is derived as

$$\hat{\theta}(m) = \hat{\theta}(m-1) + \alpha \Delta\hat{\theta}(m) \tag{10}$$

where α is the learning rate, $\Delta\hat{\theta}(m)$ is "Gauss-Newton" search direction. $\Delta\hat{\theta}(m)$ can be defined as

$$\Delta\hat{\theta}(m) = -[H_m]^{-1} G(m) \tag{11}$$

where $G(m)$ is a gradient of $J(\hat{\theta})$, and H_m is Hessian matrix of $J(\hat{\theta})$.

We can obtain from (4) that

$$J_m = J_{m-1} + \beta \ln[\cosh(\frac{e(m)}{\beta})]. \quad (12)$$

Thus,

$$G(m) = \frac{\partial J_m}{\partial \theta}|_{\theta=\hat{\theta}(m-1)} = \frac{\partial J_{m-1}}{\partial \theta}|_{\theta=\hat{\theta}(m-1)} - \varphi(m)\tanh[\frac{e(m)}{\beta}] = G(m-1) - \varphi(m)\tanh[\frac{e(m)}{\beta}]. \quad (13)$$

Then, we can get

$$H_m = \frac{\partial^2 J_m}{\partial \theta^2} = \frac{1}{\beta}\sum_{k=1}^{m} \varphi(k)\varphi^{\mathrm{T}}(k)[1 - \tanh^2(\frac{e(k)}{\beta})]. \quad (14)$$

Apparently, Hessian matrix is a symmetric matrix, and it should be kept positive in the process of recursive computing. (14) can be rewritten as

$$H_m = \frac{\partial^2 J_m}{\partial \theta^2} = H_{m-1} + \frac{1}{\beta}\varphi(m)\varphi^{\mathrm{T}}(m)[1 - \tanh^2(\frac{e(m)}{\beta})]. \quad (15)$$

Define

$$P(m) = H_m^{-1}.$$

Then,

$$P(m-1) = H_{m-1}^{-1}.$$

Define

$$z(m) = \frac{1}{1 - \tanh^2 \frac{e(m)}{\beta}}.$$

Then, we can obtain that

$$P(m) = P(m-1) - P(m-1)\varphi(m)[z(m) + \varphi^{\mathrm{T}}(m)P(m-1)\varphi(m)]^{-1}\varphi^{\mathrm{T}}(m)P(m-1). \quad (16)$$

Define

$$W(m) = P(m-1)\varphi(m)[z(m) + \varphi^{\mathrm{T}}(m)P(m-1)\varphi(m)]^{-1}. \quad (17)$$

Then,

$$P(m) = P(m-1) - W(m)\varphi^{\mathrm{T}}(m)P(m-1). \quad (18)$$

Equation (17) can be rewritten

$$W(m)z(m) = P(m-1)\varphi(m) - W(m)\varphi^{\mathrm{T}}(m)P(m-1)\varphi(m). \quad (19)$$

We obtain (20) from (18)

$$P(m)\varphi(m) = P(m-1)\varphi(m) - W(m)\varphi^{\mathrm{T}}(m)P(m-1)\varphi(m). \quad (20)$$

Thus,

$$W(m)z(m) = P(m)\varphi(m). \quad (21)$$

Then,

$$\begin{aligned}
&\Delta\hat{\theta}(m) = -[H_m]^{-1}G(m) - \\
&P(m)\{G(m-1) - \varphi(m)\tanh[\frac{e(m)}{\beta}]\} = \\
&[W(m)\varphi^{\mathrm{T}}(m) - I]P(m-1)G(m-1) + \\
&P(m)\varphi(m)\tanh[\frac{e(m)}{\beta}] = \\
&[I - W(m)\varphi^{\mathrm{T}}(m)]\Delta\hat{\theta}(m-1) + P(m)\varphi(m)\tanh[\frac{e(m)}{\beta}] = \\
&\Delta\hat{\theta}(m-1) - W(m)\{\varphi^{\mathrm{T}}(m)\Delta\hat{\theta}(m-1) - \\
&z(m)\tanh[\frac{e(m)}{\beta}]\}
\end{aligned} \quad (22)$$

where $P(0) = \omega I$, $\omega = 10^6$, $\hat{\theta}(0)$ and $\Delta\hat{\theta}(0)$ are real vectors initialized small enough. □

3.3 Analysis of algorithm performance

1) A deterministic function which has characteristics of absolute value function is introduced in this paper. It compensates the shortcomings of the LAD method whose objective function is non-differentiable. Traditional optimize algorithm is used to solve the problem of system identification. It has advantages of simple calculation and easy implementation, and is valuable for practical applications.

2) The method proposed in this paper uses first and second order derivative fully, Gauss-Newton optimization algorithm is used to derive recursive equations that reduce storage and computation effectively and it has better convergence performance than gradient algorithm. Moreover, recursive equations for Hessian matrix are derived in the method.

4 Simulation and analysis

In order to prove the good performance of the proposed ALAD method in this paper, many simulation experiments have been done. The proposed method is compared with LS method for two cases, one is that only white noises exist in measured data, the other is that not only white noises but disorder and peak noises exist in the measured data. simulation results show that the new algorithm has better robustness and faster convergence speed. The simulation and analysis of a typical system are shown as follows.

4.1 Simulation example and evaluation criteria

Consider the typical auto regression with control item (CAR) model

$$A(z^{-1})y(k) = B(z^{-1})u(k) + v(k)$$

where

$$A(z^{-1}) = 1 - 1.5z^{-1} + 0.7z^{-2}$$

$$B(z^{-1}) = z^{-1} + 0.5z^{-2}$$

where $\{u(k), k = 1, 2, \cdots, m\}$ are input values of M series with the longest cycle period of 31 bits, $\{v(k), k = 1, 2, \cdots, m\}$ are additive random disturbances, input and output data for identification are obtained from this model. Comparison with LS method for the two cases is as follows:

Case 1. Only white noises exist in measured data, $\{v(k), k = 1, 2, \cdots, m\}$ are white noises with variance one and zero mean value.

Case 2. Disorder and peak noises are added with the probability of 5%, 10% and 20% respectively to the measured data that are disturbed by white noises. Relative error of estimated parameter is used as the evaluation criteria of the algorithm, which is defined as

$$\delta = \frac{||\theta - \hat{\theta}(t)||}{||\theta||}.$$

4.2 Identification results and analysis

Simulation experiments are done for the mentioned two cases respectively. The corresponding results obtained by the ordinary LS method and the proposed ALAD method are also included for comparison.

1) If only white noises exist in measured data, the results are shown in Table 1.

Table 1 Simulation results of LS method and ALAD method

Parameter	a_1	a_2	b_1	b_2
True value	–1.5	0.7	1	0.5
LS	–1.4988	0.6998	1.0044	0.4951
ALAD	–1.5011	0.6983	1.0101	0.4983

2) Disorder and peak noises are added with the probability of 5%, 10% and 20% respectively to the measured data that are disturbed by white noises, curves of relative errors are shown in Figs. 2–4.

Results analysis: The results of Table 1 demonstrate that, when only white noises exist in measured data, both the ALAD algorithm and the LS method are able to achieve good performance with high accuracy. It is shown from Figs. 2–4 that the ALAD algorithm has higher identification accuracy and faster convergence speed than the LS method when disorder and peak noises exist in the measured data. And the more noisy, the better identification results the new algorithm gets. Simulation results show that the new algorithm has better robustness, faster convergence speed and higher identification accuracy than the LS method, when the measured data are disturbed by not only white noises but also disorder and peak noises.

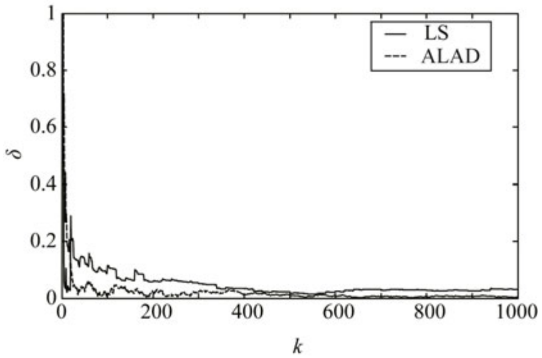

Fig. 2 Comparison of δ curves with the 5% probability of disorder and peak noises

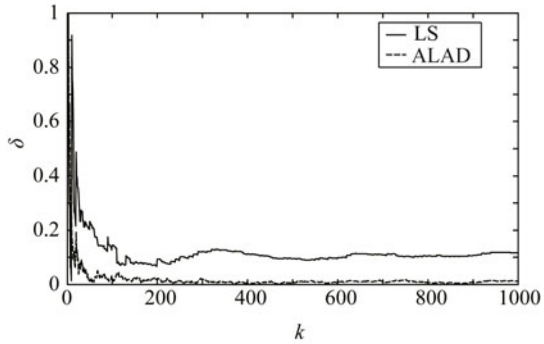

Fig. 3 Comparison of δ curves with the 10% probability of disorder and peak noises

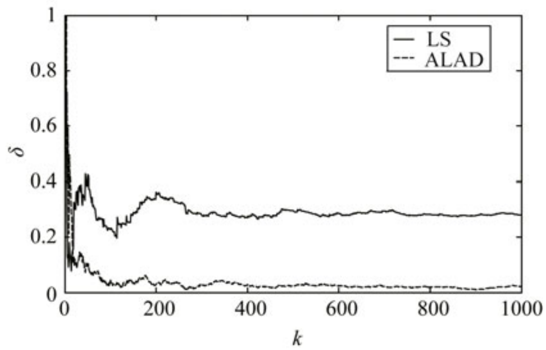

Fig. 4 Comparison of δ curves with the 20% probability of disorder and peak noises

5 Conclusions

A novel identification algorithm is proposed in this paper by minimizing an approximate LAD function. The approximate LAD function has characteristics of absolute value function and is differentiable. Then, the recursive equations can be derived using Gauss-Newton iterative algorithm. The novel algorithm has advantages of simple calculation and easy implementation with second-order convergence speed, and it avoids matrix inversion problem. The algorithm compensates the shortcomings of LAD method whose objective function is non-differentiable. And it is more robust than the LS method. Finally, the simulation results demonstrate the superiority of the proposed algorithm.

References

[1] X. F. Wu, Z. Q. Lang, S. A. Billings. An orthogonal least squares based approach to FIR designs. *International Journal of Automation and Computing*, vol. 2, no. 2, pp. 163–170, 2005.

[2] F. Ding, T. W. Chen. Hierarchical least squares identification methods for multivariable systems. *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 397–402, 2005.

[3] M. Lovera, T. Gustafsson, M. Verhaegen. Recursive subspace identification of linear and non-linear wiener state-space models. *Automatica*, vol. 36, no. 11, pp. 1639–1650, 2000.

[4] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.

[5] F. Ding, T. W. Chen, L. Qiu. Bias compensation based recursive least-squares identification algorithm for MISO systems. *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 5, pp. 349–353, 2006.

[6] X. H. Huang, X. J. Zeng, M. Wang. SVM-based identification and un-calibrated visual servoing for micromanipulation. *International Journal of Automation and Computing*, vol. 7, no. 1, pp. 47–54, 2010.

[7] F. Ding, T. W. Chen. Performance bounds of forgetting factor least-squares algorithms for time-varying systems with finite measurement data. *IEEE Transactions on Circuits and Systems*, vol. 52, no. 3, pp. 555–566, 2005.

[8] W. X. Zheng. Fast identification of autoregressive signals from noisy observations. *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 52, no. 1, pp. 43–48, 2005.

[9] C. Z. Fang, D. Y. Xiao. *Process Identification*, Beijing, China: Tsinghua University Press, pp. 457–460, 1988. (in Chinese)

[10] S. N. Elian, C. D. S. Andre, S. C. Narula. Influence measure for the L_1 regression. *Communications in Statistics. Theory and Methods*, vol. 29, no. 4, pp. 837–849, 2000.

[11] C. Bogani, M. G. Gasparo, A. Papini. Pattern search method for discrete L_1-approximation. *Journal of Optimization Theory and Applications*, vol. 134, no. 1, pp. 47–59, 2007.

[12] X. Q. Zhou, J. D. Wang. A genetic method of LAD estimation for models with censored data. *Computational Statistics and Data Analysis*, vol. 48, no. 3, pp. 451–466, 2005.

[13] L, Peng, Q. W. Yao. Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, vol. 90, no. 4, pp. 967–975, 2003.

[14] H. R. Cao. System identification based on the least-absolute criteria. *Techniques of Automation and Applications*, vol. 28, no. 7, pp. 8–10, 2009. (in Chinese)

[15] J. H. Xiao, W. L. Yang, J. Li. Realization of least-absolute regression using linear programming based on Matlab. *Journal of Wuyi University (Natural Science Edition)*, vol. 20, no. 1, pp. 5–9, 2006. (in Chinese)

[16] L. Xu, Y. H. Zhou, B. C. Wei. Influence analysis of LAD regression based on MM algorithm. *Applied Mathematics A Journal of Chinese Universities (Series A)*, vol. 22, no. 1, pp. 48–58, 2007. (in Chinese)

[17] X. P. Xu, F. C. Qian, F. Wang. New method for identification of Wiener-Hammerstein model. *Control and Decision*, vol. 23, no. 8, pp. 929–933, 2008. (in Chinese)

Bao-Chang Xu graduated from Northeast Petroleum University (NPU), China in 2000. He received his master degree from NPU in 2000, and received his Ph. D. degree from Beijing University of Aeronautics and Astronautics, China in 2005. He is currently an associate professor of control theory and control engineering in China University of Petroleum, Beijing, China.

His research interests include system identification and advanced control, image processing, multisensor data fusion, and soft-sensor technology.

E-mail: xbcyl@163.com (Corresponding author)

Xin-Le Liu is currently a post graduate student in China University of Petroleum, Beijing, China.

His research interests include intelligent control and system identification.

E-mail: liuxinle1987825@163.com