Data Mining Algorithm for Off-Group Points on Noise Polluted Time Series Based on ESO

Fuxin WANG
Engineering Research Center on Shipping Simulation by the Ministry of Education, Shanghai Maritime University, Shanghai 201306, China

Zhijian HUANG*
Engineering Research Center on Shipping Simulation by the Ministry of Education, Shanghai Maritime University, Shanghai 201306, China

Yanyan ZHANG
Central Lab of the 10th affiliated people’s hospital, Tongji University, Shanghai 200072, China

Liang QIAO
Engineering Research Center on Shipping Simulation by the Ministry of Education, Shanghai Maritime University, Shanghai 201306, China

Pinyou LI
Engineering Research Center on Shipping Simulation by the Ministry of Education, Shanghai Maritime University, Shanghai 201306, China

Email: zjhuang@shmtu.edu.cn

Abstract – The practically measured signals always contain wild values deviating far from true values. How to remove these wild values is an important research project for data mining of off-group points. In active disturbance rejection controller (ADRC), it is difficult to acquire accurate signal, since the signal is vulnerable to the influence of the wild values. Therefore, this paper puts forward extend state observer (ESO) algorithm to replace tracking differentiator (TD) method. The performances are compared between them under equal conditions and for different rang of wild values. The result suggests that the ESO algorithm will remove the wild values effectively and better, when it’s relatively small.

Keywords – ADRC, TD, ESO, Remove Wild Values, Data Mining.

I. INTRODUCTION

It is an inevitable problem that wild values exist in signals. So related algorithms on the elimination of wild values are extensively applied to many fields, such as image processing. In recent years, more and more experts begin to pay attention to mining algorithm[1-5] similar to the elimination of wild values. Mining algorithm can avoid the difficulties of analyses resulted from detailed consideration of the consumption of energy. Now we present extend state observer (ESO) algorithm to the elimination of wild values.

Signals with high precision, small fluctuation and better effect will be acquired if wild values existed in the signals are eliminated. Therefore, there are many research methods of the elimination of wild values. Some experts raise the method of restoring signals to eliminate wild values by improving damage identification techniques of signals[6], and then applying advanced techniques to restoring signals. But this approach is too idealized and has large technique barriers. Some experts suggest the segmentation of signals[7], that is to say, dividing the signal processing flow into several phases, processing respectively, and finally and completely the unified USB of procedures. But this method is too complicated to apply to the practice. Other experts use Tracking Differentiator (TD) to extract differential signals and arrange the process of transition, since TD can be more accessible to extract differential signals and remove wild values at the same time[8-9]. But using TD to remove wild values will lead to high losses of phases and distortion, so sometimes accurate signals can not be acquired.

On the basis of analyzing the previous related researches and deeply studying Extend State Observer (ESO) algorithm[10-13], we raises a new method to eliminate wild values, that is, by using ESO to remove wild values existed in signals. This method relies less on the system, and removes wild values to the most extent, so that the most accurate signals can be acquired.

II. THE ESO ALGORITHM

The ESO is the key part of the active disturbance rejection controller active disturbance rejection controller (ADRC)[14], by observing external variables to determine state variables inside the system, expanding disturbing actions able to affect controlled output into new state variables and using special feedback system to observe the extend state. Actually, The ESO is a linear differential equation, whose principle is as follows:

For a given second-order linear control system

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= a_1 x_1 + a_2 x_2 + bu \\
y &= x_1
\end{align*}
\]

(1)

The form of ESO corresponding to this linear system is

\[
\begin{align*}
\dot{e}_1 &= z_1 - y \\
\dot{z}_1 &= z_2 - l e_1 \\
\dot{z}_2 &= (a_1 z_1 + a_2 z_2) - l e_1 + bu
\end{align*}
\]

(2)

As regard to non-linear control system, let

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= f (x_1 + x_2) + bu \\
y &= x_1
\end{align*}
\]

(3)

When the function \( f (x_1 + x_2) \) and \( b \) are known, then the form of ESO is

This work was supported by the NSFC Projects of China under Grant No. 61403250.
\[ e_i = z_i - y \]
\[ \dot{z}_i = z_i - l_i e_i \]
\[ \dot{z}_2 = f^i(z_1, z_2) - l_i e_i + bu \] (4)

In order to eliminate the unknown function \( f(x_1, x_2) \) and remove errors effectively, we acquire feedback form as follows:

\[ e_i = z_i - y \]
\[ \dot{z}_i = z_i - \beta_i g_i(e) \]
\[ \dot{z}_2 = \beta_2 g_2(e) + bu \] (5)

In the form, \( e \) is the error between expected objective and actual behavior of system; \( z_1, z_2 \) are state variables of controlled system; \( \beta_i \) and \( \beta_2 \) are magnification factors of variables of ESO; \( g_i(e), i=1,2 \) (satisfying \( g_i(e) \geq 0 \); \( b \) is the magnification factor of controlled variable; \( u \) is the final controlled variable of output of controller.

As long as the value of \( \beta_i \) and \( \beta_2 \), as well as the function form of \( g_i(e) \), are appropriate chose, The ESO can perfectly predict state variables and extract differential signal.

### III. Algorithm Design

Now there is signal
\[ v(t) = 5 \sin(0.5t) + n(t) \] (6)

Where, The noise is \( n(t) = n_1(t) + n_2(t) \). The \( n_1(t) \) is Brand-Limited white noise of uniform distribution between \([-0.2, 0.2]\). The \( n_1(t)(/300) \) is Brand-Limited white noise of uniform distribution between \([-10, 10]\). They serve as wild values.

At first, we deliver \( v(t) \) to TD, then

\[ fh(k) = fh \left( x_1(k) - v(k), x_2(k), r, h_n \right) \]
\[ x_1(k + 1) = x_1(k) + hx_2(k) \]
\[ x_2(k) = x_2(k) + hfh(k) \] (7)

Where, the form of above mentioned function \( fh = fh \left( x_1(t), x_2(t), r, h_n \right) \) are as follows (the \( sign \) suggests sign function):

\[ fh = -r \left( \frac{\mu}{d} \right) \left( \frac{a}{d} \right) \text{sgn}(a) \left( 1 - \text{sgn}(a) \right) \]
\[ d = rh^2 \]
\[ a = h_x x_1(t) \]
\[ y = x_1(t) + a_0 \]
\[ a_0 = \sqrt{d(1 + 8y)} \]
\[ a_2 = a_0 + \text{sgn}(y) (a_1 - d) / 2 \]
\[ a = (a_0 + y) \text{sgn}(y) + a_1 (1 - \text{sgn}(y) \right) \]
\[ \text{sgn}(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{if } x = 0 \\
-1 & \text{if } x < 0 
\end{cases} \]

Next, we deliver the same signal to ESO, then

\[ e_i = z_i - y \]
\[ fe = fa(e, \frac{1}{2}, \delta) \]
\[ z_i = z_i + h(z_2 - B_e) \]
\[ z_2 = z_2 + h(-B_2 fe) \] (9)

Where, the form of above mentioned function \( fe = fa(e(t), \alpha, \delta) \) are as follows:

\[ fa(e(t), \alpha, \delta) = \begin{cases} 
\frac{e(t)}{\sigma_+}, & \text{if } e(t) \leq \delta \\
\frac{|e(t)| \text{sgn}(e(t))}{\sigma_+}, & \text{else} 
\end{cases} \] (10)

In the formula, \( e \) is the error between expected objective and practical behavior of the system; \( \delta \) is the interval length of linear segment in this function; \( \alpha \) is the parameter of this non-linear function, \( 0 < \alpha < 1 \), usually take \( 0.5, 0.25 \) or \( 0.05 \); \( sign \) suggests sign function.

### IV. Simulation Result and Analysis

#### A. Use algorithm of TD to eliminate wild values

After delivering the signal to TD, take parameters \( h = 0.001, r = 50, h_n = n_1, n_1 = 10 \), meanwhile setting Brand-Limited white noise \( n_1(t) = [-0.2, 0.2] \), setting Brand-Limited white noise \( n_1(t) = [-10, 10] \) which serves as wild values.

Signals without the elimination of wild values are shown in the chart of Fig1. From the chart, obviously, we can see that since the signal is affected by the noise of wild values, it has a large fluctuation and a serious damage of the phase.

![Fig.1. The original signal chart without the elimination of wild values. The lateral axis is time and the vertical axis is the value of signals.](image)

After the elimination of wild values by using TD, shown in the chart of Fig2. We can see the signal is affected by the noise of wild values by using TD, it has a little fluctuation and small loss of the phase.
B. Use algorithm of ESO to eliminate wild values

After delivering the signal to ESO, take parameters \( d = 5, \beta_1 = \beta_2 = 50, T = 0.0005 \), meanwhile setting Brand-Limited white noise \( n_1(t) = [-0.2, 0.2] \), setting Brand-Limited white noise \( n_2(t) = [-10, 10] \) which serves as wild values.

Signals without the elimination of wild values are shown in the chart of Fig1. Also from the chart, obviously, we can see that since the signal is affected by the noise of wild values, it has a large fluctuation and a serious damage of the phase.

After the elimination of wild values by using ESO, shown in the chart of Fig2. We can see the signal is affected by the noise of wild values by using ESO, it has a little fluctuation and small damage of the phase.

IV. Effect of Size of Wild Values on Algorithm

In the following, we will alter the size of the noise of wild values in order to acquire simulated results:

When \( n_2(t) \) is between \([-1, 1]\), signals without TD and with TD are shown in the charts of Fig4 and Fig5. From the charts we can see, TD is good at removing wild values with low noise.

When \( n_2(t) \) is between \([-30, 30]\), signals without TD and with TD are shown in the charts of Fig6 and Fig7. From the charts we can see, TD is not bad at (better than common effect), removing wild values with loud noise.
When $n_i(t)$ is between $[-30,30]$, signals with ESO and without ESO are shown in the charts of Fig6 and Fig9. From the charts we can see, ESO has a common effect on removing wild values with loud noise.

Fig.6. Original signal without elimination of wild values with noises between $[-30,30]$. The lateral axis is time and the vertical axis is the value of signals.

Fig.7. Signal with TD to eliminate wild values with noises between $[-30,30]$. The lateral axis is time and the vertical axis is the value of signals.

When $n_i(t)$ is between $[-1,1]$, signals with ESO and without ESO are shown in the charts of Fig4 and Fig8. From the charts we can see, ESO is excellent at removing wild values with loud noise.

Fig.8. Signal with ESO to eliminate wild values with noises between $[-1,1]$. The lateral axis is time and the vertical axis is the value of signals.

Fig.9. Signal with ESO to eliminate wild values with noises between $[-30,30]$. The lateral axis is time and the vertical axis is the value of signals.

Through comparing the simulated results from charts of Fig5 and Fig8, it is easy to see that the same signal respectively use algorithms of TD and ESO at the same time, the signal using ESO is more stable and its fluctuation is smaller. Through comparing the simulated results from charts of Fig7 and Fig9, it is easy to see that the same signal respectively use algorithms of TD and ESO at the same time, the fluctuation of the signal using TD is smaller.

Therefore, when the size of wild values are relatively small, ESO this paper mentioned is better at removing wild values than TD.

V. CONCLUSION

As regard to the elimination of wild values, this paper works out a new way to remove wild values, that is, replacing TD with ESO, perform the simulation and then acquiring data chart. The conclusion are as follows:
(1) Comparing with the method of restoring tiny signals to eliminate wild values, the technical difficulty of ESO this paper mentioned is small, and easy to implement.
(2) Comparing with the method of segmentation of signal to eliminate wild values, the ESO this paper mentioned is very simple and practical, can be better applied to practice.
(3) Comparing with the method of TD to eliminate wild values, the ESO this paper mentioned is more stable and the damage of the phase is comparatively small.

As an important research part in ADRC, ESO this paper mentioned is a simple and practical observer of mathematics model. We employ ESO to do signal extraction and removal of wild values, hoping to provide useful references for practical engineering.
ACKNOWLEDGEMENT

This work was supported by the NSFC Projects of China under Grant No. 61403250.

REFERENCES


AUTHOR’S PROFILE

Fuxin Wang (1990-)

is master from Shanghai Maritime University, Shanghai, China(email:1186505195@qq.com). From 2010 to 2014, studied in Merchant College of Shanghai Maritime University for marine engineering. From 2014 to now, study in Merchant College of Shanghai Maritime University for shipboard and ocean engineering, more about marine automation.

Zhijian Huang

received the B.S. and M.S. degrees in marine engineering from Shanghai Maritime University in 2002 and 2005. He received the Ph.D degree in marine engineering and automation from Shanghai Jiao Tong University in 2013. His research interests include optimal control, nonlinear control and adaptive control.

Email: zjhuang@shmtu.edu.cn