XVI IMEKO World Congress
Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future
Vienna, AUSTRIA, 2000, September 25-28

STEP FORM DETECTION: A COMPARISON OF CURVE FITTING TECHNIQUES

***H. Jokinen*[1], *H. Ihalainen*[2], *M. Moisio*[3]**
[1,2] Tampere University of Technology, Measurement and Information Technology, FIN-33101 Tampere, Finland.
[3] Oy Keskuslaboratorio, FIN-02151 Espoo, Finland.

Abstract: Controlling of process flow is problematic if the valves do not behave ideally. The non-ideal behaviour of a valve causes the valve position to differ from the action given by the control system, which produces discontinuities in the process flow. This is shown as stick-slip steps in the measurement signals. A curve fitting technique is used in step form detection. In this paper we compare this curve fitting technique to well-known Haar wavelet method and present their ability in reliable step form detection.
Keywords: signal processing, stick-slip steps, Haar wavelets

1 INTRODUCTION

The non-ideal behaviour of a flow control valve causes the valve position to differ from the action given by the control system, which produces discontinuities in the process flow. This is clearly illustrated as a step form characteristics in the measured error signal, i.e. the difference of the process output signal and the set point value. These forms are called stick-slip steps and they are caused by phenomena such as hysteresis and differences between static and dynamic friction forces [1].

The main target of this signal processing task is to identify the stick-slip tendency of the process. It is done by computing a measure called *step index* [2], which is applied in a process control maintenance system to decide whether the signals really consist of steps or not. This measure is calculated as a relation of two different fits created both linear and non-linear smoothing methods. The linear fit is produced by a zero phase smoothing filter and the non-linear one is a combination of several piecewise constant polynomial fits connecting abruptly to each other. In this paper we present procedures in which the original curve fitting technique based on non-linear smoothing methods is compared to well-known Haar and Wavelet transform techniques [3].

The detection of step form characteristics is problematic in practise, because there are lots of different disturbing effects including noise, impulsive disturbances, and periodicities, which must not be confused with the actual abrupt changes. In order to be able to handle with all these problems a special preprocessing of the signals is needed: the outlying samples as well as the periodic sequences are detected and rejected before the index calculation [2].

The step index has proved to be very useful for checking control valves. It is applied to technical diagnostic purposes used to monitor the developing stick-slip step characteristic of the valve operation [4].

2 PRESENT CURVE FITTING TECHNIQUE

The aim of the method is to fit a step form signal x_{step} to the measured error signal x that is the difference of the output measurement and the set point value. x is expected to be normally distributed with standard deviation σ_x and zero mean value. The fitting is initialised by introducing evenly spaced step breaks in the error signal and adjusting them by the method of minimising the square sum of the fitting error (MSSE). The number of breaks is as large as $N/5$ at the beginning, producing $N_{\mathrm{i}} = N/5 + 1$ intervals of length 5. These breaks are then adjusted one by one by computing the square sums Q of the intervals and finding out the minimum Q of two successive intervals in the following manner. Let's assume three successive break points i, j, and k, where $i < j < k$, and the minimum value for $i = 1$ and the maximum value for $k = N + 1$. The break points define two successive intervals $[i, j)$ and $[j, k)$. The square sum of the two intervals between $[i, k)$ can be calculated as

$$Q_{j[i,k)} = \sum_{[i,k)} x^2 - (j-i)\cdot m^2{}_{[i,j)} - (k-j)\cdot m^2{}_{[j,k)} \tag{1}$$

where j varies in the interval $[i+1, k-1)$, and $m_{[i,j)}$ and $m_{[j,k)}$ denote the sample means of the intervals $[i, j)$ and $[j, k)$, respectively. The new adjusted break point is the point j producing the minimum value of $Q_{j,[i,\,k)}$.

The algorithm continues by iteratively trying to find out the optimum number of intervals with proper break point locations. This is done by merging intervals and inserting new break points when needed.

For the merging of intervals the sum of the square sums of two successive intervals $[i, j)$ and $[j, k)$ is compared with the square sum of a combined interval $[i, k)$ of these two intervals. The merging is done if the sum of the square sums of the intervals is only slightly smaller than the square sum of the combined interval satisfying the following equation [5]:

$$\frac{Q_{[i,k)} - \left(Q_{[i,j)} + Q_{[j,k)}\right)}{\sum_{i=1}^{N_i-1} Q_{[i,j)}} < \frac{F_{merge}}{N - N_i}$$

F_{merge} is a value of $F(1,\infty)$-distribution, N denotes the number of samples, N_{i} denotes the number of intervals, N–N_{i} denotes the number of degrees of freedom for the testing, and $Q_{[\cdot)}$:s are the square sums of the intervals e.g.

$$Q_{[i,j)} = \sum_{[i,j)} x^2 - (j-i) \cdot m^2{}_{[i,j)} \tag{3}$$

New break points are inserted if the square sum of an interval is large, relative to the sum of the square sums of the intervals. Insertion of a new break point between the points i and j is done if the following equation becomes true:

$$\frac{Q_{[i,j)}}{\sum_{i=1}^{N_i-1} Q_{[i,j)}} > \frac{F_{insert}}{N - N_i} \tag{4}$$

F_{insert} is an experimental value. After insertion the new break point is adjusted in a similar manner presented before in this paper.

3 WAVELET BASED CURVE FITTING TECHNIQUE

The another curve fitting technique is based on discrete time wavelet transform (DTWT) [3, 6] especially applying its multiresolution analysis properties. DTWT procedure contains two stages: decomposition and reconstruction. DTWT decomposition is used to extract two different aspects of the signal: the low-frequency as well as the high-frequency contents, called approximations and details, respectively. These forms are created using a successive filtering technique, in which the original signal is filtered by high-pass and low-pass filters. These filters are called the discrete mother wavelet h_a and its mirror version g_a, respectively. In order to eliminate redundancy, the filtered signals are downsampled by a factor of two creating the first scale (m = 1) coefficients of the transform. The procedure may then be repeated on the downsampled low-pass output to create the next scale ($m = 2, \ldots, M$) coefficients. After successive filtering process, the frequency content of the signal is divided into several components, called wavelet coefficients $w(m,n)$, where w denotes a wavelet coefficient, m is the scale index, and n is the translation index.

By inverting the process the original signal may be reconstructed. It is done by upsampling and filtering the coefficients with reconstruction high-pass and low-pass filters h_s and g_s, respectively. Applying this procedure separately to the coefficients of each scale m, band-pass multiresolution structures of the signal are produced. Multiresolution analysis may be applied e.g. to signal detection [6] by combining the structures of some scales only. The perfect reconstruction of the original signal is achieved by summing all the structures together.

In this application, the Haar wavelets [3] or orthogonal Daubechies wavelets [7] may be used as filters h_a, g_a, h_s, and g_s. Haar wavelet is used to approximate signals with sudden changes and piecewise constant areas because it has only one vanishing moment. The decomposition Haar filter is $h_a = \left[1/\sqrt{2}, 1/\sqrt{2}\right]$. The Haar wavelet is defined as

$$y(t) = \begin{cases} 1 & \text{if } 0 \le t < ½ \\ -1 & \text{if } ½ \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The multiresolution technique with simply taking into account only some scales is not satisfactory because exact locations of edges may not be detected properly. Some large signal components at the lower scales are needed. Thus the following methods are used with the multiresolution technique in order to enhance significant edges and suppress noise. The total curve fitting algorithm combines DTWT decomposition, spatial coefficient filtering, coefficient thresholding, and DTWT reconstruction, applied in this order.

3.1 Spatial coefficient filtering

Sharp edges have large signals over many wavelet scales while noise is present mainly on lower scales. This is taken into account in spatially selective noise filtration technique [8, 9] that is applied to sharpen and enhance significant edges while suppressing noise. Instead of using dyadic wavelet transform [7] the Haar wavelet coefficients are filtered in this application. The method may be called spatial coefficient filtering.

The central part of the algorithm is spatial correlation

$$C_k(m,n) = \prod_{i=0}^{k-1} w(m+1,n) \quad n = 1,2,\dots,N_m \tag{6}$$

where $w(m,n)$ denotes the wavelet coefficients, $k < M - m + 1$, and M is the total number of scales. Usually $k = 2$. The algorithm proceeds from higher scales to lower scales starting with the computation of $C_k(m,n)$ for each scale m. In this application only scales $m = 4,3,2,1$ are used. Because the amount of coefficients N_m is different at successive scales, the coefficients at scale $m+1$ are upsampled for the computation of $C_k(m,n)$. Additionally, in order to improve spatial coverage we use $w(m+1,n) = \max_n\left(|w(m+1,n)|, |w(m+1,n-1)|\right)$.

The power of $C_k(m,n)$ is then rescaled to that of $w(m,n)$, i.e.

$$NC_k(m,n) = C_k(m,n)\sqrt{P_w / P_C} \tag{7}$$

where $P_w = \sum_n w(m,n)^2$ and $P_C = \sum_n C_k(m,n)^2$. If $|NC_k(m,n)| > p(m)|w(m,n)|$, point n is accepted as an edge, $w(m,n)$ is passed to $w_S(m,n)$, and $w(m,n)$ and $C_k(m,n)$ are set to 0. Otherwise, $w(m,n)$ is expected to be produced by noise, and no zeroing is done. $p(m)$ is a weight used to reduce the amount of noise to be extracted as edges at lower scales. $p(1) \geq p(2) \geq 1$, other weights are equal to 1. The algorithm is repeated until the power of $w(m,n)$ is nearly equal to the noise power of the scale. In practice, two or three iterations are needed. The signal is finally reconstructed by inverse Wavelet transform from the coefficients $w_S(m,n)$.

3.2 Wavelet coefficient thresholding

Wavelet coefficient thresholding [7, 8] is a technique used in signal smoothing and noise removal. It sets to zero all wavelet coefficients that are below certain threshold $T(m) = ks_m$, where k is a constant, and s_m denotes the standard deviation of the coefficients of the mth scale. The smoothing is thus adaptive depending on the amplitude of the coefficients. To estimate the standard deviation s_m of the coefficients, a robust estimator is calculated by sorting the absolute values of the coefficients in ascending order and choosing the value at the index $0.68 \cdot N_m$ as the standard deviation estimate. N_m denotes the number of coefficients of the mth scale. In this application hard thresholding is applied using the threshold $T(1)$ computed based on the standard deviation of the first scale coefficients, i.e.

$$\hat{w}(m,n) = \begin{cases} w(m,n), & |w(m,n)| \geq T(1) \\ 0, & |w(m,n)| < T(1) \end{cases} \tag{8}$$

4 STEP INDEX COMPUTATION

As a result of the both curve fitting techniques presented in chapters 2 and 3, the break points I_{step} are found and a signal x_{step} is produced. Next two squared sequences $u_1(i) = \left(x(i) - x_{\text{step}}(i)\right)^2$ and $u_2(i) = \left(x(i) - \bar{x}(i)\right)^2$ are produced. $\bar{x}$ is derived from signal x smoothed with a gauss window function of length N_{GW}. The default value for $N_{GW} = 128$. This is done in order to distinguish steps from smooth signal forms.

The step index c is computed based on the following equation

$$c = \frac{\sum_i \left(u_2(i) \cdot V(i) - u_1(i) \cdot V(i)\right)}{\sum_i \left(u_2(i) \cdot V(i)\right) + D} \tag{9}$$

The weight vector V is a sequence composed of values 0 or 1. $V(i) = 0$, if sample $x(i)$ is found to be either erroneous or situated within the periodical region of the signal. Otherwise, $V(i) = 1$. V is used in suppressing the values of $x(i \mid V(i) = 0)$, so that they have no effect on step index calculations. D denotes the amount of variance of the periodical region of x multiplied by the amount of samples N_s i.e. $D = N_s \cdot \sigma_x^2\left(i \mid s_1(i) > 2 \cdot s_2(i)\right)$. D is needed in deteriorating the step index value due to the periodical region of the signal, because the effect of this region on the index value should be similar to that of the normal region of the data. Without this deterioration the step index value would be too large. When the value of c is between 0.5 and 0.7, or between 0.7 and 1, moderate or high-level stick-slip tendency is detected, respectively.

4.1 Verification of the stick-slip tendency

When the stick-slip steps are present in the error signal, the control output signal x_c has its own typical characteristic, too [10]. At the time the step change occurs in the error signal x, the sign of the slope factor of the control output signal x_c changes. When the changes occur continuously, the form of x_c is close to triangular wave. This is due to the fact that the control system tries to drive the valve position into a proper state after too large a step change. The interaction is used in verifying the stick-slip tendency of the process.

The sequence I_{step} contains the indices of the step changes in x_{step}. Because the sequences $x_{step}(I_{step})$ and $x_c(I_{step})$ may be locally monotonous, a sequence I_c is created based on I_{step} by removing the indices causing $x_c(I_{step})$ to be monotonous. As a result, a sequence I_c is created so that $x_c(I_c)$ oscillates without locally monotonous periods.

The phasing of sequences $x(I_c)$ and $x_c(I_c)$ is computed for the verification of the stick-slip tendency. It expresses the amount of phase similarity between the error signal break points and the control output signal and it is estimated according to the following procedure. $x_c(I_c)$ is first differentiated in order to remove a trend from the sequence i.e. $x_{cd}(i) = \{0, x_c(i) - x_c(i-1)\}, i \in I_c$. The amount of phasing P is computed as

$$P = 100 \cdot \frac{\sum_i \left[\left(\left|\operatorname{sgn}(x(i)) + \operatorname{sgn}(x_{cd}(i))\right| - 1\right) \cdot V(i)\right]}{N_{Ic} - N_e} \tag{10}$$

where $i \in I_c$, $\operatorname{sgn}(\cdot)$ denotes the sign function, N_{Ic} denotes the number of indices in sequence I_c, and N_e denotes the number of samples getting zeroed by V. The range of P is $-100 \ldots 100$ % illustrating the total out of phase and in-phase alternatives, respectively. When the absolute value of P is between $50 \ldots 100$ %, the probability of stick-slip tendency is high.

5 COMPARISON OF CURVE FITTING TECHNIQUES

The presented curve fitting techniques are compared in the following. This is done by investigating their applicability in finding out step form characteristic from three different error signals. Additionally, the measures revealing the stick-slip tendency, the step index c and the phasing P are also computed. In all the cases the control valve was expected to suffer from stick-slip phenomenon. The signals and the two different fits in cases A, B, and C are presented in Figures 1, 2, and 3, respectively. The computed values of c and P for both the fitting techniques are shown in Table 1. What is common among all the cases is that the fit produced by the Wavelet based technique follows the local mean of the original signal accurately – including the short time changes of the signal – at times where large step changes occur. This is not satisfactory because the generality of the fitting is impeded. In fact, because the general optimality is important the minimum square sum method produces better results.

Signal A is a good example of clear stick-slip phenomenon. The two fits as well as the step index values are almost identical. Due to the increased number of break points the phasing produced by the Wavelet based technique is slightly smaller. In the second case (signal B) the stick-slip steps are not as clear as in case A. In fact, some of the large level changes are due to normal variation of the system and thus they should not become mixed with stick-slip steps. However, the Wavelet based fit follows the slow variation and the extraction of real steps becomes impossible. The values of step index and phasing produced by the MSSE technique are acceptable in this case.

In the last case (signal C) stick-slip steps could not be detected, but the large changes are due to the oscillations of the system. The step index value 0.34 produced by the MSSE technique is certainly closer to the true value in this case. The Haar wavelet fit seems to contain too many lower scale coefficients, but, if the algorithm were changed so that these coefficients are removed, the break points would not be located optimally even in case A. Although Haar wavelets are composed of abrupt changes and constant areas, the extraction of that kind of gross features only is not possible with the presented algorithm.

Title:
swtfig1.eps
Creator:
MATLAB, The Mathworks, Inc.
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

Figure 1. Example A: a measured error signal (gray) with the fits produced by the MSSE technique (–) and the Wavelet based technique (· –).

Title:
swtfig2.eps
Creator:
MATLAB, The Mathworks, Inc.
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

Figure 2. Example B: a measured error signal (gray) with the fits produced by the MSSE technique (–) and the Wavelet based technique (· –).

Title:
swtfig3.eps
Creator:
MATLAB, The Mathworks, Inc.
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

Figure 3. Example C: a measured error signal (gray) with the fits produced by the MSSE technique (–) and the Wavelet based technique (· –).

Table 1: Values of the step index c and phasing P of three different error signals. MSSE and DTWT denote the curve fitting techniques used.

Signal example	c MSSE	c DTWT	P [%] MSSE	P [%] DTWT
A	0.72	0.73	68	59
B	0.60	0.79	57	78
C	0.34	0.76	49	63

6 CONCLUSION

In this paper we have compared two curve fitting techniques in connection with the computation of a step index, a measure that is used to reveal the stick-slip step phenomenon in flow control valve measurements. The present technique applying minimum square sum error computation was compared with Haar wavelet method that is used to approximate signals with sudden changes and piecewise constant areas. However, the signals may contain oscillations that are close to abrupt changes. It is important that these slow features are separated from real abrupt ones. Based on the study, the present MSSE method turned out to be more reliable compared to the technique based on Haar wavelet method.

REFERENCES

[1] J. Piipponen, Controlling processes with nonideal valves: tuning of loops and selection of valves, *Control Systems ‘96*, Halifax, Canada, 1996, p. 179-186.
[2] H. Jokinen, H. Ihalainen, M. Moisio, Detection of step form in signals with application to flow control maintenance, *Proceedings of the IMEKO XV World Congress*, Osaka, Japan, 1999.
[3] I. Daubechies, *Ten lectures on Wavelets*, SIAM, Philadelphia, 1997, 357 p.
[4] M. Moisio, R. Ritala, L. Nyström, Less process variations through continuous control loop performance assessment by using KCL-CoPA, *VTT Symposium 188*, Espoo, 1998, p. 73-85.
[5] N.R. Draper, H. Smith, Applied Regression Analysis, 2nd edition, John Wiley & Sons, New York 1981, 709 p.
[6] L. Angrisani, P. Daponte, M. D’ Apuzzo, A method for the automatic detection and measurement of transients. Part I: the measurement method, *Measurement* **25** (1) (1999) 19-30.
[7] S.G. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998, 577 p.
[8] Q. Pan, L. Zhang, G. Dai, H. Zhang, Two denoising methods by wavelet transform, *IEEE Transactions on Signal Processing* **47** (12) (1999) 3401-3406.
[9] Y. Xu, J.B. Weaver, D.M. Healy, Jr., J. Lu, Wavelet transform domain filters: a spatially selective noise filtration technique, *IEEE Transactions on Image Processing* **3** (6) (1994) 747-758.
[10] K. Forsman, Performance monitoring of sensors, controllers and actuators with application to paper making, *Control Systems ‘98*, Porvoo, Finland, 1998, p. 275-281.

AUTHORS: Researcher Heikki JOKINEN, Laboratory engineer Heimo IHALAINEN, Tampere University of Technology, Measurement and Information Technology, P.O. Box 692, FIN-33101 Tampere, Finland, Fax +358 3 365 2171, E-mail: heikki.jokinen@mit.tut.fi, heimo.ihalainen@mit.tut.fi, MSc Marko MOISIO, Oy Keskuslaboratorio, P.O. Box 70, FIN-02151 Espoo, Finland, E-mail: Marko.Moisio@kcl.fi