A New Denoising Technique for Capillary Electrophoresis Signals

WANG, Ying(王瑛) MO, Jin-Yuan*(莫金垣)

School of Chemistry and Chemical Engineering, Zhongshan University, Guangzhou, Guangdong 510275, China

Capillary electrophoresis (CE) is a powerful analytical tool in chemistry. Thus, it is valuable to solve the denoising of CE signals. A new denoising method called MWDA which employs Mexican Hat wavelet is presented. It is an efficient chemometrics technique and has been applied successfully in processing CE signals. Useful information can be extracted even from signals of $S/N = 1$. After denoising, the peak positions are unchanged and the relative errors of peak height are less than 3%.

Keywords Mexican Hat wavelet denoising algorithm, curve fitting, denoising, signal processing

As a powerful analytical tool, capillary electrophoresis (CE) has been widely used in chemistry.[1] But signals obtained from CE instruments often have random noise. The noise degrades the accuracy and precision of an analysis, so it is crucial for the further application of CE that the noise can be removed from the signals effectively. Besides improving instruments, the chemometrics means has become an important tool to remove noise. There are different possible approaches to signal denoising and the most recent methods[2,3,4] are based on wavelet transform[5] (WT), which is a novel and powerful signal processing tool developed from Fourier transform. The principle of methods based on WT is extracting the low frequency components from the noisy signals and taking them as the denoised results, because the frequency of useful signals is usually much lower than that of the noise. But when CE signals are processed by WT, the useful peaks will be removed as noise. Because the peaks of CE signals are very sharp and their frequencies are high. So the useful peaks and noise can not be distinguished by frequencies. Spline wavelet least square[6] (SWLS) is another kind of new denoising technique and can process signals with sharp peaks. But the baseline of the processed curve is not smooth enough and repeated denoising is generally required.

Therefore, though there are many denoising methods, no one is very suitable for CE signals. The purpose of this paper is to present a new method named Mexican Hat wavelet denoising algorithm (MWDA). MWDA does not denoise by frequencies, so it is effective for CE signals. In the field of signal denoising, Mexican Hat wavelet has not been reported, it is a really new tool.

MWDA is derived from curve fitting, which is to choose or construct a new function $f(x)$ as fitting function, and $f(x)$ can be used to approximate the original signal points $\{(x_i, y_i)\}_{i=1}^{k}$. In course of fitting, a principle called least square must be obeyed, which is to make the sum of square errors between the new function and the original data points to be minimum. That is

$$D_{min} = \sum_{i=1}^{k} [y_i - f(x_i)]^2 \cdot P_i \quad (1)$$

y_i, $f(x_i)$, represent the ith point on the original signal and the fitting function respectively, k is the number of signal points, P_i is the weight coefficient which always equals to 1 in this paper.

Mexican Hat wavelet[7] has a simple explicit expression

* E-mail: cesmjy@zsu.edu.cn

Received February 4, 2002; revised and accepted June 4, 2002.

Project supported by the National Natural Science Foundation of China (No. 29975033) and the Natural Science Foundation of Guangdong Province (No. 980340).

$$\Psi(t) = (1 - t^2)e^{-t^2/2} \tag{2}$$

and its figure is smooth. Its value reduces rapidly with the attenuation of coefficients, so it accords with the characteristic of seeing in space. For these reasons, Mexican Hat wavelet is suitable for a fitting function. So it is used in this paper, then the fitting function $f(x)$ can be expressed as follows:

$$f(x_i) = \sum_{i=1}^{n} c_i\varphi_i \tag{3}$$

where φ_i is Mexican Hat wavelet function, c_i is the wanted coefficients for φ_i, n is the number of the wavelet functions used in curve fitting. The expression of φ_i is shown as:

$$\varphi_i(x) = \Psi\left(\frac{x - x_{(i-1/2)}}{h}\right), \; i = 1, 2, \cdots, n \tag{4}$$

where h is the step length of fitting.

From Eqs. (1) and (3), it can be seen that Eq. (1) can be regarded as the function of c_i:

$$D(c_1, c_2, \cdots, c_n) = \sum_{i=1}^{k}[f(x_i) - y_i]^2, \quad i = 1, 2, \cdots, n \tag{5}$$

When D gets its minimum value, its partial derivatives are zero,

$$\frac{\partial D}{\partial c_i} = 0, \; i = 1, 2, \cdots, n \tag{6}$$

From Eqs. (5) and (6), the following linear equations can be obtained,

$$\sum_{j=1}^{n} \langle \varphi_i, \varphi_j \rangle c_j = \langle \varphi_i, y \rangle, \quad i = 1, 2, \cdots, n \tag{7}$$

where

$$\langle \varphi_i, \varphi_j \rangle = \sum_{v=1}^{k} \varphi_i(x_v)\varphi_j(x_v) \tag{8}$$

$$\langle \varphi_i, y \rangle = \sum_{v=1}^{k} \varphi_i(x_v)y_v \tag{9}$$

Use Eqs. (8) and (9) to calculate the values of $\langle \varphi_i, \varphi_j \rangle$, and $\langle \varphi_i, y \rangle$, then the wanted coefficients c_i can be obtained by Eq. (7). Put the values of c_i into Eq. (3), a fitting function $f(x)$, which accords with least square, is constructed. The noise in signals is random, so it will not exist in the fitted curve. Then, $f(x)$ is the denoised signal.

In a word, fitting the signals by Mexican Hat wavelet and obeying the least square, to achieve the purpose of removing noise, this is MWDA.

According to the above algorithm, a program was made, and the whole process can be accomplished by the computer automatically.

Simulating signals are processed at first to test MWDA's performance. When the signal-to-noise ratio (S/N) varies from 4 to 1, the processed curves are similar to the theoretical ones, the peak positions are unchanged and the relative errors of peak height are less than 3%. Then the actual signals with noise are processed successfully by MWDA. Fig. 1 gives an example to demonstrate the efficiency of MWDA. CE signal has a typical form as Fig. 1a. When denoised by WT, the sharp peaks containing useful information will be removed as noise. But if MWDA is employed, the result is satisfactory (Fig. 1b). Comparing the processed curve by SWLS (Fig. 1c) and the one by MWDA, we can see that MWDA is better. As seen from the figure, MWDA is a powerful signal processing tool.

Something that must be explained is the selection of parameters during processing. There are only two alterable parameters, step length h and number of wavelet functions n. For a certain type of signals, h is decided by noise. When the noise is high, the value of h should be big; when it is low, the value is small. In practice, it was found that n has an optimum value corresponding to a settled h. We can match h and its optimum n, so selecting h according to the noise is the whole thing. We can even match h and the level of the noise, then the parameters will be decided by the signals themselves.

MWDA is a very simple and efficient denoising technique. There is no limitation of the number of sampling points. Satisfactory results are got after denoising only once. No parameters are set beforehand, so there is few man-induced factors. And MWDA is easy to realize automatization and on-line processing. It is of great practical value.

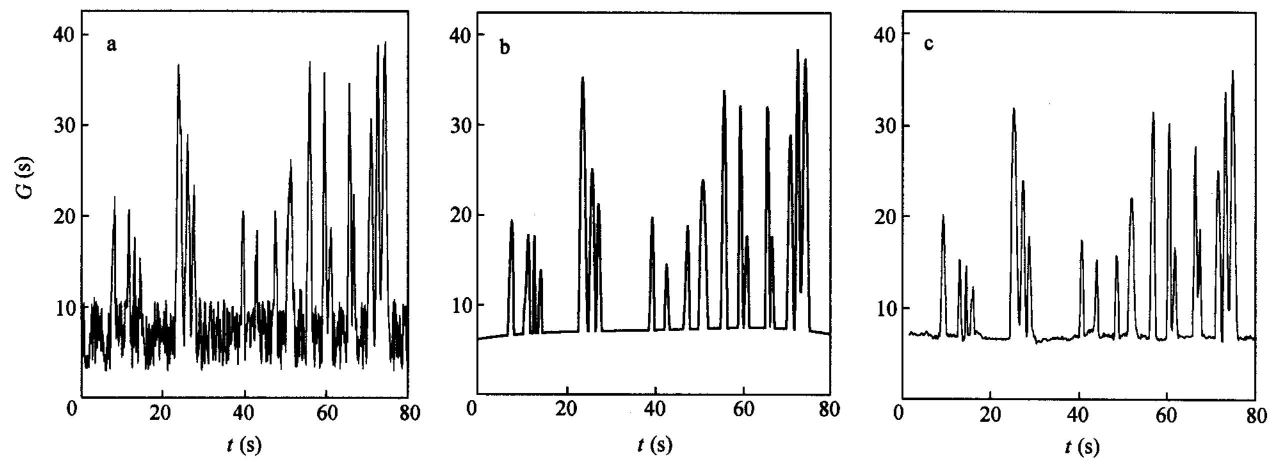

Fig. 1 Capillary electrophoresis signal and processing results. a: original signal b: processing result of MWDA c: processing result of SWLS. The 15 peaks from the most right side to left are respectively: cystine, glutamic acid, asparagic acid, tyrosine, proline, methionine, threonine, serine, leucine, valine, lactamic acid, glycine, histidine, lysine, arginine

References

1 Altria, K. D. *J. Chromatogr.*, *A* **1999**, *856*, 443.

2 Cherkassky, V.; Shao, X. *Neural Networks* **2001**, *14*, 37.

3 Yuan, Z. B.; Yu, K. Q. *Chin. J. Anal. Lab.* **2001**, *20*, 7 (in Chinese).

4 Dai, Y. S. *Solid-State Electron.* **2000**, *44*, 2147.

5 Daubechies, I. *IEEE Transaction on Information Theory* **1990**, *36*, 961.

6 Zheng, X. P.; Mo, J. Y.; Xie, T. Y. *Talanta* **1999**, *48*, 425.

7 Yang, F. S. *The Engineering Analysis and Application of Wavelet Transform*, Science Press, Beijing, **1999**, p. 22 (in Chinese).

(E0202048 ZHAO, X. J.; DONG, L. J.)