Adaptive Abnormal Signal Extraction Based on Iterative Computations

Wang Fang, Lin Weiguo*, He Zhaoyan, Wu Haiyan

College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029 E-mail: linwg@mail.buct.edu.cn

Abstract: An adaptive approach of abnormal signal extracting is proposed for pipeline leak detection based on acoustic method in this paper. The dynamic pressure signal is divided into many intervals based on zero crossing points, and each frame of signal is regarded as the superposition of the interval signals. The characteristics of the interval signals and their Signal-to-Noise Ratios (SNRs) are analyzed first. Based on the analysis, the number of abnormal signal intervals is determined depending on the position of inflection point in the Root-Mean-Square error sequence obtained by SNR difference sequences, and then abnormal signals are isolated. In this method, the feature extraction of signals and leakage diagnosis model training are all cancelled, and the problem of missing alarm can be overcome. The field experiment testified the effectiveness of the approach.

Key Words: Iterative computations, Signal-to-Noise Ratio, Abnormal signal, Adaptive extraction, Leak detection

1. INTROUCTION

At present, many domestic and overseas scholars have done a lot of research on the pipeline leak detection technologies and pipeline leak signal feature extraction ^[1-3]. Among them, the study of pipeline leak detection method based on acoustic ^[4] is relatively outstanding.

Pipeline leakage diagnosis is the basis of the leak location. With the development of science and technologies as well as the needs of engineering practice, people propose many leakage diagnosis methods and models. Jiedi Sun et al put forward the concept of root mean square entropy, and regarded it as the feature of abnormal signals. Then, local mean decomposition (LMD) method combined with support vector machine (SVM) was adopted to establish the leakage diagnosis model. However, there are still some problems need be solved in the LMD method, such as the choice of moving average step, the criterion of pure frequency modulation signal and the endpoint effect ^[5]. Yu Zhang et al decomposed the detected signal into a sum of finite intrinsic mode functions (IMFs) by empirical mode decomposition (EMD). The normalized kurtosis value of the main IMF components was analyzed and extracted as feature of the abnormal signals. However, boundary effect, modal aliasing and so on are problems to be solved ^[6]. Yuanhua Oi et al proposed a method for feature extraction based on the time domain statistical characteristics of acoustic signal, and only normal signal samples were used to train the leakage diagnosis model of PCA-SVDD, but the model based on PCA-SVDD also has some problems, such as the selection of kernel function should be made by experience and the model allows a certain probability of false and missing alarm ^[7]. Wei Liang et al used wavelet packet nodes' energies of normal signal samples as the input space of Gaussian mixture model (GMM) to train leakage diagnosis model, and the number of Gaussian function in the GMM was determined by Bayesian information criterion (BIC)^[8]. Lei Ni et al used wavelet packet characteristics entropy to describe abnormal signals, and built leakage diagnosis model based on PSO-SVM ^[9]. Likun Wang et al built leakage diagnosis model based on Back-Propagation neural network using kurtosis and energy as the input feature space ^[10]. In addition to the above methods, there are many leakage diagnosis approaches. These approaches have some common problems: some features (kurtosis, energy and so on) are used to describe abnormal signals, samples (normal or abnormal signal samples) are used to train leakage diagnosis model. Whereas, these extracted features and samples cannot fully represent signals under all conditions, then missing alarm is inevitable. In recent years with the rapid development of data driven [11-13] and its application in process monitoring ^[14], some new ideas are provided to solve these drawbacks mentioned above.

An approach of adaptive abnormal signal extraction based on iterative Signal-to-Noise Ratio (SNR) computations is proposed in this paper. A frame of signal is firstly divided into positive and negative intervals according to the zero crossing points and is regarded as the superposition of all interval signals. With the increasing of the assumed number of abnormal signals, the SNR of every interval signal is then calculated iteratively. The abnormal signals are extracted by adaptively finding the inflection point of Root-Mean-Square (RMS) error sequence calculated from difference sequences of successive two SNR sequences. The approach does not need feature extraction and leak diagnosis model training; the missing alarms aroused by these procedures are then eliminated. And it can not only determine the number of abnormal signals, but also locate the abnormal signals in the original signal which provides support for the leak location with correlation.

The rest of the paper is organized as follows. In Section 2, the algorithm of abnormal signal extraction is proposed. In Section 3, numerical simulation is carried out to evaluate the performance of the approach proposed. In Section 4, field data test is used to verify the effectiveness of this approach. Conclusions are given in Section 5.

This work is supported by State Key Laboratory of NBC Protection for Civilian (SKLNBC2014-10), National Natural Science Foundation of China (61403017) and Fundamental Research Funds for the Central Universities (YS0104)

2. ADAPTIVE ABNORMAL SIGNAL EX-TRACTION BASED ON ITERATIVE COMPUTATIONS

As the dynamic pressure signals are bipolar signals with positive and negative signals alternatively, a frame of signal is firstly divided into positive and negative intervals according to zero crossing points. Assuming that there are M intervals after dividing, each interval is regarded as an independent signal, namely a frame signal is a superposition of M interval signals. The beginning and ending time of each interval are named SSt(k) and SEnd(k)respectively, and k is interval index number, as shown in the Fig. 1. Assuming that the absolute value of signal peak in each interval is sp(k), the peak position is Ppos(k), $1 \le k \le M$, then:

$$x(n) = \sum_{k=1}^{M} s_{k}(n)$$
(1)

where $S_k(n)=x(n) \cdot W_k(n)=x(n) \cdot [u(n-SEt(k))-u(n-SEt(k))]$, $sp(k)=max\{|S_k(n)|\}$. $W_k(\cdot)$ is window function and $u(\cdot)$ is step function.



Fig. 1: Interval dividing of signal

Ordering peaks vector ranging from big to small as: $P = [sp(1), sp(2), \dots, sp(M)]^T$.

The corresponding original interval indexes are stored in an array *Pos*, which can be used to trace the peak back to the interval position in the original signal.

Based on the experience of pipeline leak detection, the following assumptions can be made:

A1: The number of noise signal intervals in a signal frame is greater than that of abnormal signal intervals, namely: $m \le M/2$.

*A***2**: The absolute value of abnormal signal peak is greater than the normal signal, namely: $sp_a > sp_n$.

According to the experience, the number of abnormal signal intervals in a frame dynamic pressure signal is far less than the total intervals number. That is the number of abnormal signal intervals is far less than the noise signal intervals, namely $m \ll M-m$. Then the assumption of A1 is founded. Assumption A2 is a typical conclusion in pipeline leakage detection based on dynamic pressure signal because of signal de-noising.

Generally, Signal-to-Noise Ratio (SNR) is defined as:

$$SNR = 10 \lg(P_s / P_u)(dB)$$
(2)

Where $P_s = A_s^2 / 2$, as is the peak of signal, P_u is the noise variance.

Combining with the assumption in Eq. (1) and A1, the P_u in the SNR definition formula is modified as:

$$P_u = \left(\sum_{i=1}^{M-1} \frac{A_u^2(i)}{2}\right) / (M-1)$$
(3)

Where $A_u(i)$ is the *i*th interval signals' maximum peak (background signals).

In the following, the algorithm of adaptive abnormal signal extraction is given. And there are some lemmas used in this algorithm. The following lemmas are developed first.

Lemma1: If sp(i) > sp(j), then $SNR_i > SNR_j$.

Proof1. Vector P in descending order, namely sp(i)>sp(i+1), available SNR of each interval is given based on Eq. (2) and Eq. (3):

$$SNR(i) = 10 \lg \frac{sp^{2}(i)/2}{((\sum_{j=1}^{M} \frac{sp^{2}(j)}{2}) - \frac{sp^{2}(i)}{2})/(M-1)}$$
$$= 10 \lg \frac{sp^{2}(i)(M-1)}{(\sum_{j=1}^{M} sp^{2}(j)) - sp^{2}(i)}$$

$$\cdot \quad sp(l) > sp(l+1)$$

$$\cdots \quad sp^{2}(i) > sp^{2}(i+1)$$

$$(\sum_{j=1}^{M} sp^{2}(j)) - sp^{2}(i) < (\sum_{j=1}^{M} sp^{2}(j)) - sp^{2}(i+1)$$

$$\cdot \frac{sp^{2}(i)}{(\sum_{j=1}^{M} p^{2}(j)) - sp^{2}(i+1)} > \frac{sp^{2}(i+1)}{(\sum_{j=1}^{M} p^{2}(j)) - sp^{2}(i+1)}$$

$$(\sum_{j=1}^{M} sp^{2}(j)) - sp^{2}(i) \quad (\sum_{j=1}^{M} sp^{2}(j)) - sp^{2}(i+1)$$

: M-1>0, and y=10 lgx is a monotone increasing function.

:
$$10 \lg \frac{sp^2(i)(M-1)}{(\sum_{j=1}^{M} sp^2(j)) - sp^2(i)} > 10 \lg \frac{sp^2(i+1)(M-1)}{(\sum_{j=1}^{M} sp^2(j)) - sp^2(i+1)}$$

$$\therefore$$
 SNR(i) > SNR(i+1)

Thus if sp(i) > sp(j), then $SNR_i > SNR_i$.

Lemma2: If sp(i)>sp(j), then p(sp(i))>p(sp(j)), p is the probability that sp's interval is abnormal signal.

Proof2: From the A2, it can be got that the absolute value of abnormal signal peak is greater than the normal signal. That is to say, if sp(i) is not abnormal signal, then sp(i+1), sp(i+2), ..., sp(M) are certainly not abnormal signals. In other words, if sp(i) > sp(j), then p(sp(i)) > p(sp(j)).

The elements in vector P represent the absolute value of interval peak, and are ordered from big to small. From the Lemma2, it can be seen that if there are m abnormal interval signals, the others are noise signals, the vector P can be expressed as $P = [sp_a(1),...,sp_a(m), sp_n(1),..., sp_n(M-m)]^T$.

Based on the A1, A2, Lemma1, Lemma2, Eq. (2) and Eq. (3), this paper proposes an adaptive approach for extracting abnormal signals based on iterative computation:

Step1: Assuming that there is one abnormal signal corresponding to the first element in vector P, the SNR of the abnormal interval signal is calculated according to the Eq. (4). The SNRs of the other background interval signals are calculated according to the Eq. (5), where j=2,...,M.

$$SNR_{1}(1) = 10 \lg(\frac{sp^{2}(1)}{(\sum_{i=2}^{M} sp^{2}(i)) / (M-1)})$$
(4)

$$SNR_{1}(j) = 10 \lg(\frac{sp^{2}(j)}{(\sum_{i=2, i \neq j}^{M} sp^{2}(i)) / (M-2)})$$
(5)

Step2: Increasing the number of abnormal signals to m in turn, the SNRs of the abnormal interval signals are calculated according to Eq. (6), and the SNRs of the other background interval signals are calculated according to Eq. (7).

$$SNR_{m}(j) = 10 \lg(\frac{sp^{2}(j)}{(\sum_{i=m+1, i \neq j}^{M} sp^{2}(i)) / (M-m)})$$
(6)

$$SNR_{m}(j) = 10 \lg(\frac{sp(j)}{(\sum_{i=m+1, i\neq j}^{M} sp^{2}(i)) / (M - (m+1))})$$
(7)

where in Eq. (6) j=1,2,...,m, and in the Eq. (7) j=m+1,...,M, m=1, 2,...,L. *L* is the maximum iterative times which represents the largest possible number of abnormal signals in a frame.

Step3: The difference equation and its Root-Mean-Square error (denoted with *RMSED*) are defined as following:

$$D_{SNR_{m-1}}(j) = SNR_m(j) - SNR_{m-1}(j), j=1,...,M; m=2,...,L$$
(8)

$$MD_{SNRm-1} = \frac{1}{M} \sum_{j=1}^{M} D_{SNRm-1}(j)$$
(9)

$$RMSED_{m-1} = \sqrt{\frac{\sum_{j=1}^{M} (D_{SNRm-1}(j) - MD_{SNRm-1})^2}{M-1}}$$
(10)

Step4: Each *RMSED* obtained by difference sequence D_{SNR} is regarded as an element of the vector \mathbf{R} , $\mathbf{R} = [RMSED_1, RMSED_2, ..., RMSED_{L-1}]$. Because $RMSED_i$ is greater than zero, \mathbf{R} is processed as the following in order to make the sequence \mathbf{R} have bipolar characteristic.

The element which is the first one out of the range $[RMSED^-, RMSED^+]$ in vector **R** by reverse order is denoted by $RMSED_r$. Then calculating the mean value (denoted by mr) of elements from r+1 to L-1 in vector **R**, the new vector **R**' is obtained by every element minus mr:

$$MR = \frac{1}{L-1} \sum_{m=1}^{L-1} RMSED_m$$
(11)

$$\sigma = \sqrt{\frac{\sum_{m=1}^{L-1} (RMSED_m - MR)^2}{L-2}}$$
(12)

$$RMSED^+ = MR + \sigma, RMSED^- = MR - \sigma$$
 (13)

$$mr = \frac{1}{L - 1 - r} \sum_{m=r+1}^{L-1} RMSED_m$$
(14)

$$\boldsymbol{R}' = [RMSED_1 - mr, RMSED_2 - mr, \dots, RMSED_{L-1} - mr]$$
(15)

Similarly, as the elements in \mathbf{R} also have the bipolar characteristic, \mathbf{R} is divided into positive and negative intervals according to zero crossing points.

Assuming that there are *l* intervals after dividing, the beginning and ending time of each interval are named SSt'(k) and SEnd'(k) respectively, and *k* is interval index number, namely k=1,...,l. Then each interval's absolute value of sum (denoted by sum_k) can be got, and the maximum value (denoted by sum_{pos}) in all sums also can be found out.

Finally, the interval index *pos* indicate the interval which the abnormal signal locates in.

The number of abnormal signals can be got by tracing to the source. In this paper, the number of extracted abnormal signals is denoted by *Num*:

$$sum_{k} = \left| \sum_{i=SSt^{'}(k)}^{SEnd^{'}(k)} (RMSED_{i} - mr) \right|$$
(16)

$$sum_{pos} = max(sum_1, sum_2, \dots, sum_l)$$
(17)

$$Num = SEnd'(pos) + 1 \tag{18}$$

The following Proof3 presents the reason why Eq. (18) plus one.

Proof3:

: $RMSED_i = f(SNR_{i+1}, SNR_i)$, i=1,...,L-1, and the corresponding number of iterations m=2,...,L

$$\therefore i = m - 1$$

Also: The number of iterations expresses the number of abnormal signals, namely *m=Num*

$$\therefore i = Num - 1$$
Also
$$\therefore i = SEnd'(pos)$$

$$\therefore Num = SEnd'(pos) + 1$$

Step5: After the *Num* is determined, every abnormal signal's position in the original signal can be got by tracing the peak back to the interval position in the original signals. The beginning time of every abnormal signal in original signal is denoted by $t_s(i)$, and the ending time is denoted by $t_e(i)$, where i=1,2,...,Num:

$$t_s(i) = SSt(Pos(i)) \tag{19}$$

$$t_e(i) = SEnd(Pos(i)) \tag{20}$$

For the convenience of following-up leak location, the normal signals are made zero clearing.

3. NUMERICAL SIMULATION

Considering the following equations:

$$g = e + \boldsymbol{\beta} * \boldsymbol{h} \tag{21}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{bmatrix}$$
(22)
$$\begin{bmatrix} [A.sin(2\pi f.x)]^* [u(randi) - u(randi + n. -1)] \end{bmatrix}$$

$$\boldsymbol{h} = \begin{bmatrix} [A_2 \sin(2\pi f_2 x)]^* [u(randi) - u(randi + n_2 - 1)] \\ [A_3 \sin(2\pi f_3 x)]^* [u(randi) - u(randi + n_3 - 1)] \\ [A_4 \sin(2\pi f_4 x)]^* [u(randi) - u(randi + n_4 - 1)] \end{bmatrix}$$
(23)

Where the data points N=6000, sampling frequency $f_s=50Hz$ (the dominant frequency of acoustic leak signal is less than 20Hz). *e* is Gaussian and randomly produced with mean 0 and standard deviation 0.5. $x=(0: N-1)/f_s$, $u(\cdot)$

is unit step function. *randi* is a random integer from 0 to *N*-1. n_i is the data points of every sinusoidal signal in a single cycle. Amplitude/*V* and frequency/*Hz* of sinusoidal signal are set as follows: $A_1=7$, $A_2=12$, $A_3=5$, $A_4=3$, $f_1=0.50$, $f_2=0.45$, $f_3=0.43$, $f_4=0.39$. $\beta_i=0$ or 1. If one and only one element is 1 in vector β , it means that there is only one sinusoidal signal. If at least two elements get 1 in vector β , it means that there are two sinusoidal signals. In this paper $\beta = [1 \ 1 \ 1 \ 1]$, namely, the number of sinusoidal signals is 4 and abnormal signals is 8, as shown in Fig.2.





Fig. 3 shows the RMSED curve of simulated abnormal signals respect to different iteration times. From the Fig.3, it can be seen that the RMSED curve changes intensively in the beginning and then tends to stabilize with the increasing of the number of iterations. When RMSED becomes stable, the corresponding number of iterations is the number of abnormal signals. While after RMSED centralized processing, as shown in Fig. 4, R' can be divided into l (l=3) intervals. From the Fig. 4, it can be seen that the maximum value of each interval's accumulation is in the first interval, namely pos=1. And this interval corresponding to iteration cut-off time is SEnd(pos)=7. So the number of abnormal signals is Num= SEnd(pos)+1=8, which is consistent with the fact. The positions of abnormal signals can be traced from the position of the original signal with Eq. (19) and (20). And the Fig. 5 is the abnormal signals extracted.



Fig. 3: The *RMSED* curve of the simulated abnormal signals respect to different iteration times



Fig. 4: The *RMSED-mr* curve of the simulated abnormal signals respect to different iteration times



Fig. 6: The statistical results of abnormal signals

In order to test whether any missing alarm will occur with the approach proposed above. A big data sample is generated randomly, particular procedure is as follows:

Based on based on Eq. (21), Eq. (22) and Eq. (23), corresponding to the amount of the data of one year in pipeline leak detection, a signal sample including 525600 data is generated. The amplitudes (A1, A2, A3, A4) and position of abnormal signals (simulated with sinusoidal signal) in every frame signal sample are randomly generated. The results is shown in Fig. 6, the horizontal axis shows the number of abnormal signals extracted, and the vertical axis shows the frequency of abnormal signals' number extracted. According to Fig. 6, it can be seen that

the number of abnormal signals extracted is always greater than or equal to 8. In other words, there is not missing alarm occurred.

4. FIELD DATA TEST

Field data from a naphtha pipeline are also used for test. The pipeline was 15.511km long, and the pressure was 2.18MPa at upstream, 0.48MPa at downstream. Pipeline diameter was 150 mm, artificial leaks was 9.476km away from the upstream of the pipeline. The sampling frequency was 50 Hz.

As the dynamic pressure signals detected by the piezoelectric pressure transducer were coupled with a great number of background noise signals, in order to improve the SNR of signals, wavelet soft-threshold de-noising is used to de-noise the signals collected from the field (*DB9* in Daubechies wavelet group is selected as the wavelet basis and the decomposition scale is 5).





There was an artificial leak signal simulated on this pipeline on November 21, 2013, 14:24, and the leakage hole-size was less than 4mm. Dynamic pressure signals were collected by piezoelectric pressure transducers installed at upstream and downstream of the pipeline respectively.



Fig. 8: The *RMSED* curve of the true abnormal signals respect to different iteration times



Fig. 9: The *RMSED-mr* curve of the true abnormal signals respect to different iteration times



Fig. 10: The extracted abnormal signals

As shown in Fig. 7, upstream and downstream have four abnormal signals respectively. Fig. 8 shows the *RMSED* curve corresponding to different iteration times, from which it can be seen that the inflection points of upstream and downstream are all around 4. Then the accumulation's maximum values of upstream and downstream are all in the first interval. The *SEnd(pos)* of upstream is 4, and the downstream *SEnd(pos)* is 3, as shown in Fig. 9. So the number of abnormal signals in upstream and downstream is 5 and 4. And the Fig. 10 is the abnormal signals extracted.

Multiple artificial leaks had been simulated on the naphtha pipeline in November 20 to 21, 2013, and the leak hole-size are 4mm, 8mm, less than 4mm, less than 8mm respectively. The proposed approach is tested using these historical data. The extracted abnormal signal numbers are compared with actual numbers. The results are shown in Table 1.

From Table 1, it can be seen that the number of extracted abnormal signals are always bigger than or equal to the actual numbers. That means all abnormal signals are found, and there is no missing alarm occurred.

Table 1. The comparison of actual and extracted abnormal signals number

Leakage time		Leakage hole- size/mm	Actual number Up/Down stream	Extracted number Up/Down stream
On No- vember 20, 2013	16:14	4	2/2	5/3
	16:21		2/2	3/4
	16:24		3/2	7/4
	16:27	8	2/2	4/3
	16:31		2/2	8/3
	17:51	<8	2/2	7/3
	17:54		2/2	2/5
	17:57		2/2	4/5
	18:01		2/2	3/4
On No- vember 21, 2013	14:24	<4	4/4	5/4
	14:39		2/2	2/8
	14:43		2/2	4/5
	14:46		2/2	4/3

5. CONCLUSIONS

This paper proposes an adaptive abnormal signal extraction approach based on iterative computation of SNR, no feature extraction and leakage diagnosis modeling are needed. Furthermore, threshold setting can be eliminated too. It has been testified with filed data that all of the abnormal signals can be extracted without missing. However, there are some certain defects, such as the number of abnormal signals extracted is often bigger than the actual numbers, which may result in false alarming. These problems need to be explored next.

REFERENCES

- Pedro J. Lee, John P. Vitkovsky, Martin F. Lambert, Angus R. Simpson, James A. Liggett, Leak location using the pattern of the frequency response diagram in pipelines: a numerical study, Journal of Sound and Vibration, Vol.284, No.3-5, 1051-1073, 2005.
- [2] Zhao Yang, Zhuang Xiong and Min Shao, A new method of leak location for the natural gas pipeline based on wavelet analysis, Energy 35 (2010), pp. 3814-3820.
- [3] Yang Jie, Wang Guizeng, The summarization of gas pipeline leakage diagnosis technology, Control and Instruments in Chemical Industry, Vol.03, 01-05, 2004.
- [4] Lin Weiguo, Zheng Zhishou, Research on pipeline leak detection based on dynamic pressure signal, Chinese Journal of Scientific Instrument, Vol.27, No.8, 907-910, 2006.
- [5] Jiedi Sun, Qiyang Xiao, Jiangtao Wenb, Ying Zhang, Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis, Measurement, Vol.79, 147–157, 2016.
- [6] Zhang Yu, Shijiu Jin, Jingjing He, Shili Chen, Jian Li, Extraction method for pipeline leakage feature based on dynamic pressure signal, Shiyou Xuebao, Vol.31, No.2, 338-342, 2010.
- [7] Qi Yuanhua, Lin Weiguo, Wu Haiyan, A leak detection method for natural gas pipelines based on time-domain sta-

tistical features, Shiyou Xuebao, Vol.34, No.6, 1195-1199, 2013.

- [8] Wei Liang, Laibin Zhang, A wave change analysis (WCA) method for pipeline leak detection using Gaussian mixture model, Journal of Loss Prevention in the Process Industries, Vol.25, 60-69, 2012.
- [9] Lei Ni, Juncheng Jiang, Yong Pan, Zhirong Wang, Leak location of pipelines based on characteristic entropy, Journal of Loss Prevention in the Process Industries, Vol.30, 24-36, 2014.
- [10] Likun Wang, Jinyun Zhao, Songguang Fu, Dongjie Tan, Jian Li, Shijiu Jin, Pipeline leakage acoustic signal feature recognition based on neural network, Chinese Journal of Scientific Instrument, Vol.27, No.6, 2247-2249, 2006.
- [11] Zhongsheng Hou, Xuhui Bu, Model free adaptive control with data dropouts, Expert Systems with Applications, Vol.28, 10709-10717, 2011.
- [12] Dong Shen, Zhongsheng Hou, Iterative Learning Control with Unknown Control, IEEE Trans. on Neural Networks, Vol.22, No. 12, 2011.
- [13] Dong Shen, Youqing Wang, Survey on stochastic iterative learning control, Journal of Process Control, Vol.24, 64-77, 2014.
- [14] Shen Yin, Steven X. Ding, Xiaochen Xie, Hao Luo, A Review on Basic Data-Driven Approaches for Industrial Process Monitoring, IEEE Trans. on Industrial Electronics, Vol.61, No.11, 6418-6428, 2014.