Wavelet-based baseline correction optimization algorithm for SF6 spectral signal

Liu Wenfeng¹, Wang Yanting¹, Liu Liang¹, Yao Mingshu², Xu Xiaofeng², Gao Jing² and Wang Liang²

¹ Taipingwan hydroelectric power plant, China
² Shenyang Institute of Engineering, China
E-mail: 2727820196@qq.com

Abstract. SF6 gas is widely used in various power equipment. In this paper, a baseline correction algorithm based on spectral detection method based on optimal wavelet basis is proposed. The algorithm selects the optimal wavelet base, removes the strong spectrum information by threshold method, eliminates the influence of continuum fitting, and performs the least squares fitting on the residual signal to obtain the continuum, and removes the continuum from the original spectrum to obtain the required Line spectrum. The accuracy of the quantitative analysis of the line spectrum intensity after processing is greatly improved, and the baseline is effectively estimated and corrected, and the fitting accuracy is higher than the traditional iterative wavelet baseline correction algorithm.

1. Introduction
The SF6 insulating gas is a mixture system with uncertain components. The concentration of many components to be tested is very low, which requires high sensitivity of the analytical method [1-3]. Absorbing spectral signals of insulating gases often exhibit baseline drift and tilt under the influence of source, instrument background, and sample size. A variety of derivatives also cause a complex background image of the spectral signal, coupled with the measurement environment and the background spectral drift caused by the change of conditions, so that in addition to noise, the spectral data also has a continuum of low-frequency gradual change. The line spectrum is superimposed on the continuum, so that the intensity of the line cannot be truly reflected, which is not conducive to the extraction of the line, which limits the accuracy of the quantitative analysis. Therefore, the use of baseline correction algorithms to remove continuum is one of the key techniques for spectral analysis.

2. The theoretical basis optimizing
Due to the complex physical causes of baseline drift, it is difficult to establish an accurate model of the continuum. Considering the environment and other factors, the background spectrum is a low-frequency signal with relatively stable trend, and the line spectrum is mainly located in the high frequency band [4]. Therefore, using the difference between the two, a curve fitting method is used to obtain an approximate continuous spectrum curve. These methods essentially extract the continuous slowly varying signal in the spectrum by smooth operation. The influence of the strong line makes the fitted continuum not accurate enough in the vicinity, away from the baseline, and deviates from the peak of the line, thus limiting the extraction of the line and the determination of the parameters. Wavelet with its tight support and multi-resolution characteristics, has attracted researchers at home and abroad to study the application of baseline correction [5].
In this paper, a baseline correction algorithm based on the optimal wavelet basis is proposed. By constructing the cost function, the best wavelet base is selected to represent the signal in the wavelet library generated by the SF6 mixed gas spectral signal, so that the spectral signal is decomposed by the wavelet base. According to the characteristics of the continuum and spectral peaks in the energy distribution, the decomposition coefficients of the spectral signals under the optimal wavelet basis are thresholded, the peak signal components of the energy concentration are eliminated, and the residual signal is subjected to least squares fitting to extract the continuum signal. After removing it, a spectral line spectrum can be obtained. Applying the algorithm to the sr absorption spectrum signal of insulating gas can accurately estimate the baseline fluctuation potential and effectively correct the baseline drift phenomenon.

Figure 1 shows a schematic diagram of the wavelet library space for signal decomposition. Each of these small squares represents the time-frequency space occupied by the decomposition of the signal under a particular basis function. It can be seen that the width of the entire line represents the room space, and the frequency from low to high is shown from left to right. For the spatial decomposition scale, I is the sequence number of the space V0. Each column represents the decomposition scale of the previous column, and the decomposition scale plus one. As long as the selected subspace does not cover the length of one line, it can be used as a set of orthogonal bases of space to completely represent the signal of the space.

![Wavelet frequency domain space](image)

**Figure 1.** Wavelet frequency domain space.

It can be seen that the standard wavelet transform performs multiple resolutions on the low frequency part of the signal to achieve multi-resolution analysis. This is not in line with the need to extract the spectral absorption band. The best method for dividing the wavelet space should be selected according to the characteristics of the analyzed signal, that is, the optimal basis is selected. The specific step is to determine the search method, and search for the wavelet bases that constitute the optimal basis in the entire wavelet library according to the rule that minimizes the cost function.

3. **Baseline correction algorithm**

This paper proposes a baseline correction algorithm based on the optimal wavelet basis. The main idea is to use the best wavelet base before the continuous spectrum fitting, so that the correlation peak energy in the wavelet coefficients is concentrated in a few scales, and then the threshold method is used to eliminate the strong.

3.1. **Algorithm implementation steps**

The basic steps of the baseline correction algorithm for the SF6 insulating gas spectral signal based on the optimal wavelet basis are as follows.
(1) Generate a wavelet library
When the scale is small enough, the spectral data to be processed can be assumed, and the scaling function is approximated as a function, so the sampling sequence \( f(n), n = 1, 2, \ldots, N \), which is considered to be \( f(t) \), is the wavelet packet function.
Let \( f_{m,l}(n) \) denote the decomposition coefficient of \( f(n) \) in the \( l \)th subspace of dimension \( m \), where \( m = 1, 2, \ldots, M \), \( l = 1, 2, \ldots, 2m \), then from the above assumptions, there are \( f_0, 0(n) = f(n) \), and the wavelet library of \( M = 4 \) is calculated according to the calculation formula of the wavelet packet.

(2) Calculating the cost function
Construct the cost function \( V(m, l) \) to reflect the concentration of signal energy in the wavelet coefficients. The cost function is solved for the decomposition coefficient of each base of the signal in the wavelet library.

(3) Search for the optimal basis
The optimal base is searched in the wavelet library in the order of \( e \)-tops. The principle is to minimize the total cost function of the wave base. At this time, the wavelet coefficients of the signal decomposition have the highest energy concentration.

(4) Threshold method to remove peak components
The threshold operation is used to search for the strong band on the wavelet decomposition coefficient \( f_{m,l}(n) \) of each optimal basis.

\[
H(f_{m,l}(n), h) = \begin{cases} 
|f_{m,l}(t)| & |f_{m,l}(t)| \geq h \\
0 & |f_{m,l}(t)| < h 
\end{cases}
\]

(1)

where the threshold is taken.
The above operation considers the wavelet coefficients larger than the threshold to be dominated by the strong band components, and the others are retained.

(5) Inverse transform to obtain a signal that removes peak components
After the threshold processing, the interference of the strong band to the fitted continuum is substantially eliminated. The wavelet coefficients are reconstructed on a scale-by-scale basis, and finally the spatial wavelet coefficients are obtained after removing the strong bands.

(6) Fitting the continuum and then obtaining the line spectrum
The continuum of the fit is obtained by a least squares fitting algorithm. The continuum is subtracted from the original signal \( f(n) \), which is the baseline corrected line spectrum.

3.2. Sequence length problem of wavelet packet transform
Conventional wavelet packet transform algorithms usually use the definition of inner product. In the calculation process, the data will become shorter and shorter, which often makes the subsequent processing algorithms difficult to obtain better results due to insufficient quantity. The traditional convolution type wavelet packet transform algorithm keeps the length of the decomposed data consistent with the original spectral data. The wavelet packet decomposition formula is as follows.

\[
f_{m+1, 2l}(n) = \frac{1}{2} \sum_{k \in Z} h(k) \cdot f_{m, l}(n - 2^m k) \\
\]

\[
f_{m+1, 2l+1}(n) = \frac{1}{2} \sum_{k \in Z} g(k) \cdot f_{m, l}(n - 2^m k) 
\]

(2)

where \( m \) is the scale of the decomposition, \( l \) is the subspace in the entire function space, \( h(k) \) is the low-pass filter coefficient, and \( g(k) \) is the high-pass filter coefficient corresponding to the wavelet packet reconstruction formula.
where \( h(k) \) and \( g(k) \) represent the conjugate of \( h(k) \) and \( g(k) \), respectively.

### 3.3. Selection of cost function

Since the information to be extracted mainly exists in the form of a peak, which appears as a modulus maxima in the wavelet space, a cost function is constructed to count the number of modulus maxima exceeding a certain threshold value, in all wavelet coefficients constituting the wavelet base. When the number of modulus maxima exceeding the threshold is the smallest, it is considered that the energy of the spectral signal is most concentrated at this time, and the selected wavelet base is the optimal wavelet base.

It can be seen that the key to solving is to determine the threshold. A more reasonable consideration is to weight the variance \( \sigma \) of the noise contained in the wavelet coefficients at each node in the wavelet library. As a threshold, it can generally take 3–5.

### 3.4. Optimal basis search method

Search in order from bottom to top.

First, mark the lowest layer (\( m = 4 \) each node, initialize the optimal base.

Then, for each layer 2\( m \) nodes, two or two groups, compare the sum of the two node cost function values of each group from the bottom to the top and the corresponding node cost function value of the previous layer; if the former is greater than the latter, then remove the labels of the two nodes in the lower layer and replace them with the nodes of the higher layer. Otherwise, the sum of the costs of the two nodes in the next layer replaces the cost function value of the corresponding node in the previous layer, but the label is not processed.

Compare from bottom to top layer by layer (scale) until all scales are processed. The wavelet base corresponding to all the nodes marked is the optimal wavelet base.

This method can quickly and efficiently search for the optimal basis.

### 4. Experimental results and discussion

This section verifies the performance of the proposed SF6 insulating gas spectral signal baseline correction algorithm through experiments. For comparison, the effect of using the iterative wavelet for baseline correction is first given, as shown in Fig. 2.

**Figure 2.** Baseline image of interactive wavelet.
The solid line in Fig. 2(a) is the absorption spectrum curve of the insulating gas SF6 after use in the wavenumber of 3240-1265 cm\(^{-1}\), and the broken line is the continuum extracted by the iterative wavelet method. It can be seen that the intensity of the absorption peaks at 1716 and 1585 cm\(^{-1}\) in the spectral signal is large, which affects the accuracy of the baseline correction, resulting in the continuum of the region near the strong absorption peak deviating significantly from the true baseline, and the error is large. Fig. 2(b) is a line spectrum diagram after removing the continuum spectrum. In the part outside the absorption peak region, the spectrum is relatively flat, and the change basically reflects the spectral absorption intensity. In the strong absorption band, the peak of the peak signal is significant drift, peak intensity no longer truly reflects the absorption of the mixed gas to be tested, can not be applied to the application of quantitative analysis of components in the subject.

Figure 2. Absorption spectrum.

Figure 3. Best wavelet baseline image.

Figure 3 shows the experimental effect of continuous spectrum removal using the optimal wavelet basis method. The solid line in Fig. 3(a) is the same spectral curve as described above, and the broken line is the continuous spectrum extracted by the optimal wavelet basis method. It can be seen that the fitting operation basically eliminates the interference of the strong band and accurately estimates the trend of the baseline fluctuation. Figure 3(b) is a spectrum of the SR mixed gas with the effect of removing the continuum. The baseline is relatively stable, and the peaks can be clearly identified. The spectral width and intensity accurately reflect the true absorption spectrum of the gas to be tested. Definitive and quantitative applications. In addition, the optimal wavelet base concentrates the signal energy on a few wavelet scales. The threshold method can be used to remove the peak interference with high energy. This overall threshold operation makes the algorithm have higher computational efficiency.

5. Summary
In this paper, the best wavelet basis is used to correct the baseline signal of SF6 insulating gas. Compared with the existing baseline correction method, the algorithm removes the strong spectrum information before the continuous spectrum fitting, avoids the influence on the fitting, and makes the algorithm accurate to the baseline; according to the SF6 insulating gas spectrum signal itself Features, adaptively select the best base for time-frequency signal analysis, high computational efficiency, and good robustness. The experimental results show that the optimal wavelet base correction algorithm accurately fits the baseline of the power insulation gas spectral signal, and extracts the spectral line information effectively from the original spectrum signal. The algorithm is accurate and reliable, and it is the online performance monitoring of the power insulating gas SF6. And spectral techniques for power equipment fault diagnosis provide algorithm support.

6. References
[1] Xiong Chengji. Image threshold denoising algorithm based on wavelet transform [J]. Science and Technology, 2019 (20): 62.
[2] Wang Zhen. Research on background deduction and quantitative algorithm of partial discharge decomposition components in SF₆ [D]. Chongqing University, 2011.

[3] Yao. Study on chromatographic peak background deduction and overlapping signal separation of partial discharge decomposition products in SF₆ [D]. Chongqing University, 2009.

[4] Dong Yue. On-line monitoring and vibration signal analysis of SF₆ high voltage circuit breaker [D]. Shanghai Jiaotong University, 2008.

[5] A High-performance Small Signal Amplifier [J]. Journal of Northeast Agricultural University (English Edition), 2005(02): 141-145.