

# 基于广义 $S$ 变换和奇异值分解的近红外光谱去噪

蔡剑华<sup>1</sup>, 肖永良<sup>2</sup>, 黎小琴<sup>1</sup>

<sup>1</sup>湖南文理学院物理与电子科学学院, 湖南 常德 415000;

<sup>2</sup>湖南财政经济学院信息技术与管理学院, 湖南 长沙 410205

**摘要** 针对近红外光谱物质含量检测过程中噪声影响模型精度和稳定性的问题, 引入广义  $S$  变换与奇异值分解(SVD)。利用广义  $S$  变换得到光谱数据的时频谱, 并将二维时频谱系数矩阵作为 SVD 的 Hankel 矩阵求解奇异值, 再采用  $k$ -均值聚类算法对奇异值序列进行分类计算, 确定重构奇异值个数, 对去噪后的数据矩阵进行广义  $S$  逆变换得到去噪后的光谱数据。给出组方法的基本理论和具体实现过程, 对仿真数据和谷朥粉导数光谱进行去噪, 并与传统的 9 点平滑法和小波软阈值法的去噪结果进行比较。结果表明: 所提方法克服了时域或频域单维滤波的局限性, 且无需参考噪声数据和选择基函数, 在谷朥粉导数光谱去噪中, 只需采用两个奇异值就能实现较好的去噪效果, 降低了滤波过程的复杂度。采用所提方法处理后, 近红外光谱的分析精度和模型的稳健性优于 9 点平滑处理法和小波软阈值法。相比 9 点平滑法, 所提方法的预测集的决定系数由 0.9436 增大为 0.9985, 预测均方根误差由 0.0843 减小为 0.0406, 明显提高了谷朥粉中水分含量定量检测的精度。

**关键词** 光谱学; 去噪; 广义  $S$  变换; 奇异值分解; 时频谱

中图分类号 O657.1

文献标识码 A

doi: 10.3788/AOS201838.0430005

## De-Noising of Near Infrared Spectra Based on Generalized $S$ Transform and Singular Value Decomposition

Cai Jianhua<sup>1</sup>, Xiao Yongliang<sup>2</sup>, Li Xiaoqin<sup>1</sup>

<sup>1</sup>College of Physics and Electronics Science, Hunan University of Arts and Science, Changde, Hunan 415000, China;

<sup>2</sup>School of Information Technology and Management, Hunan University of Finance and Economics, Changsha, Hunan 410205, China

**Abstract** In order to solve the problem of the influence of noise on model accuracy and stability in detecting materials content using near infrared spectra, we introduce the generalized  $S$  transform and singular value decomposition (SVD). Firstly, we use the generalized  $S$  transform to obtain time-frequency spectra of spectral data, and then use the two-dimensional time-frequency coefficient matrix as the Hankel matrix of SVD to solve singular values. Secondly, we use the  $k$ -means clustering algorithm to classify the singular value sequence and determine the reconstructed singular values. Finally, the de-noised coefficient matrix is transformed by the generalized  $S$  inversion to obtain de-noised spectral data. The basic theory and realization process of the combination method are given, and simulated data and the first derivative spectrum of wheat gluten are de-noised with the combination method. The results are compared with the traditional 9-point smoothing method and wavelet soft thresholding method. It is found that the proposed method overcomes the limitation of single dimension filtering (time domain or frequency domain), and does not need to reference noise data and select the base function. In the de-noising of wheat gluten derivative spectra, only 2 singular values are enough to achieve better de-noising effect, which reduces the complexity of the filtering process. The accuracy of the near infrared spectrum analysis and the robustness of the proposed model are better than those of the traditional 9-point smoothing method and the wavelet soft thresholding method. The predictive coefficient of the prediction set of the proposed method is 0.9985, which is larger than that of the 9-point smoothing method (0.9436). The root mean square error of the proposed method is 0.0406, which is smaller than that of the 9-point smoothing method (0.0843). The accuracy of quantitative detection of moisture content in wheat gluten is improved obviously.

收稿日期: 2017-07-10; 收到修改稿日期: 2017-11-06

基金项目: 国家自然科学基金(41304098)、湖南省自然科学基金(2017JJ2192, 2017JJ2015)、湖南省教育厅重点项目(16A146)

作者简介: 蔡剑华(1979—), 男, 博士, 副教授, 主要从事光电信息处理方面的研究。E-mail: cjh1021cjh@163.com

**Key words** spectroscopy; de-noising; generalized S transform; singular value decomposition; time-frequency spectrum  
**OCIS codes** 300.2530; 070.4790; 300.6390

## 1 引言

在采集近红外光谱时,受仪器和环境背景等因素的影响,光谱中往往会含有噪声,甚至出现基线漂移<sup>[1]</sup>。对光谱求导可有效消除背景噪声和基线漂移,突出波峰特征,但原始光谱经微分后噪声会被放大,而光谱分析中重点关注的两个特征量是尖峰信号的位置和幅度,噪声会严重影响它们的精确提取,因此需要进一步进行光谱去噪处理<sup>[1-2]</sup>。目前常用的光谱去噪方法包括基于系列平滑效应的时域方法<sup>[3]</sup>,以及基于 Fourier 变换和小波变换等的频域方法<sup>[4-6]</sup>。一些现代数字信号处理方法也被应用到光谱的预处理中,如源于小波变换的系列方法<sup>[6-7]</sup>、经验模态分解法<sup>[8]</sup>、自适应滤波法<sup>[9]</sup>、数学形态学滤波法<sup>[10]</sup>,以及它们的一些改进方法或它们相互结合的方法等<sup>[7,10]</sup>。这些方法在近红外光谱的去噪中各有优点,但也表现出一些不足。如平滑法在有效消除噪声的同时也会对有用数据进行平滑,从而改变光谱的形态<sup>[3]</sup>。傅里叶变换在时域、频域上对信号进行分析,可用于光谱数据的去噪,但无法对非平稳数据进行处理<sup>[6]</sup>。在小波变换中,小波基函数的选取和分解层次的人为确定使其自适应能力不强,不同的选择会带来截然不同的效果<sup>[7]</sup>。

Stockwell 等<sup>[11]</sup>从短时 Fourier 变换和连续小波变换的思想入手,提出了一种称为 S 变换的时频分析方法;Pinnegar 等<sup>[12]</sup>通过引入调节因子对高斯窗函数进行改进,发展了广义 S 变换。广义 S 变换吸取了前两者的优点,具有小波变换的多分辨特性,利用广义 S 变换的多分辨二维时频特性可以构建反 S 变换时频滤波方法。广义 S 变换的这些优点使其已广泛应用于生物医学<sup>[13]</sup>、地球物理<sup>[14]</sup>和故障诊断<sup>[15]</sup>等领域。本文将广义 S 变换与奇异值分解(SVD)滤波相结合,将其引入到近红外光谱的去噪中,对仿真信号和实测的谷朊粉近红外光谱数据进行处理,并与传统的 9 点平滑法和小波软阈值法的去噪结果进行比较,以评估广义 S 变换与 SVD 滤波结合方法的有效性,从而为光谱去噪寻求一种新的途径。

## 2 算法原理

### 2.1 S 变换及广义 S 变换

假设存在非平稳信号  $x(t)$ ,则其时频分析可表

示为

$$\text{FT}(t, f) = \int_{-\infty}^{\infty} x(\tau) \omega(t - \tau) e^{-i2\pi f\tau} d\tau, \quad (1)$$

式中  $t$  为时间,  $f$  为频率,  $\tau$  为时移因子,  $\omega(t)$  为窗函数。当窗函数  $\omega(t)$  为高斯窗函数时,即  $\omega(t) = \frac{1}{2\sqrt{\pi\alpha}} \exp\left(-\frac{t^2}{4\alpha}\right)$  ( $\alpha$  为尺度因子) 时, (1) 式即称为 Gabor 变换。Stockwell 等<sup>[11]</sup>提出当高斯窗口随频率  $f$  的增大而变窄时,即当  $\omega(t, f) = \frac{|f|}{\sqrt{2\pi}} \exp\left(-\frac{f^2 t^2}{2}\right)$  时, (1) 式称为 S 变换,即

$$S(t, f) = \int_{-\infty}^{\infty} x(\tau) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-i2\pi f\tau} d\tau. \quad (2)$$

Pinnegar 等<sup>[12]</sup>引入可调节参数  $\lambda, p$  进一步对高斯窗函数进行改进,用于窗函数时宽和衰减趋势的调节。此时,窗函数为  $\omega(t, f) = \frac{\lambda |f|^p}{\sqrt{2\pi\alpha}} \exp\left(-\frac{\lambda^2 f^2 p t^2}{2}\right)$ , 将其代入(1)式则称为广义 S 变换,信号  $x(t)$  的广义 S 变换<sup>[12]</sup>可写为

$$\text{GST}(t, f) = \int_{-\infty}^{+\infty} x(\tau) \frac{\lambda |f|^p}{\sqrt{2\pi}} \times \exp\left[\frac{-\lambda^2 f^2 p (\tau - t)^2}{2}\right] \exp(-i2\pi f\tau) d\tau. \quad (3)$$

进一步对窗函数进行能量归一化处理,即  $\int_{-\infty}^{\infty} |\omega(t)|^2 dt = 1$ , 则广义 S 变换是无损可逆的<sup>[16]</sup>。在应用中,常将广义 S 变换的反变换  $\text{IGST}(\cdot)$  直接利用傅里叶反变换  $\text{IFFT}(\cdot)$ <sup>[16-17]</sup> 来实现,即

$$x(t) = \text{IGST}(t, f) = \text{IFFT}\left[\sum_{i=1}^N \text{GST}(t_i, f)\right], \quad (4)$$

式中  $t_i$  为时间序列,  $i$  为时间点,  $N$  为序列长度。

相应地,一维离散广义 S 变换及其逆变换<sup>[18]</sup>分别定义为

$$\text{GST}\left(iT, \frac{n}{NT}\right) = \sum_{m=0}^{N-1} X\left(\frac{n+m}{NT}\right) \times \exp\left(-\frac{2n^2 \lambda^2 m^2}{N^2 p}\right) \exp\left(\frac{j2\pi mi}{N}\right), \quad (5)$$

$$x(kT) = \sum_{n=0}^{N-1} \left[\frac{1}{N} \sum_{i=0}^{N-1} \text{GST}\left(iT, \frac{n}{NT}\right)\right] \exp\left(\frac{j2\pi nk}{N}\right). \quad (6)$$

式中  $T$  为信号  $x(t)$  的采样间隔;  $N$  为采样点数的最

大值;  $i, k, n = 0, 1, 2, \dots, N - 1$ ;  $X \begin{bmatrix} n \\ NT \end{bmatrix}$  为信号  $x(t)$  的离散 Fourier 变换;  $m$  为进行 Fourier 变换的点数。根据(5)式对信号进行广义 S 变换, 可得到一个二维时频系数矩阵, 矩阵的列为时域参数, 行为频域参数, 反映了信号幅值对应于时间和频率的时频特征, 有助于对信号进行进一步分析。

## 2.2 广义 S 变换矩阵的全局 SVD 分解滤波

SVD 去噪方法可以有效抑制信号中的高斯白噪声<sup>[18]</sup>。本研究通过对导数光谱进行广义 S 变换, 获得导数光谱的二维时频系数矩阵, 再以此作为目标矩阵, 进行 SVD 去噪处理。

### 2.2.1 SVD 与重构

定理 设  $\mathbf{A}$  为  $m \times n$  型实矩阵, 则存在  $m$  阶正交阵  $\mathbf{U}$  和  $n$  阶正交阵  $\mathbf{V}$ , 使得

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T, \quad (7)$$

式中  $\mathbf{\Lambda}$  为对角矩阵,  $\mathbf{\Lambda} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ , 其对角元素  $\sigma_i > 0$  ( $i = 1, \dots, n$ ) 且为  $\mathbf{A}$  的奇异值, 并按降序排列 ( $\sigma_1 > \sigma_2 > \dots > \sigma_n > 0$ )。由于  $\mathbf{\Lambda}$  为对角阵, 因此 SVD 可将一个秩为  $k$  的  $m \times n$  矩阵表示为  $k$  个秩为 1 的  $m \times m$  子矩阵的和<sup>[18]</sup>。其中, 每个子矩阵由两个特征矢量(分别来自于  $\mathbf{U}$  和  $\mathbf{V}$ )和权值相乘得到<sup>[18-19]</sup>, 即

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^k \sigma_i \mathbf{A}_i, \quad (8)$$

式中  $k$  为  $\mathbf{A}$  的秩;  $\mathbf{u}_i$  和  $\mathbf{v}_i$  ( $i = 1, 2, 3, \dots$ ) 分别为  $\mathbf{U}$  和  $\mathbf{V}$  的第  $i$  列奇异值矢量;  $\sigma_i$  为的第  $i$  个奇异值;  $\mathbf{A}_i$  为包含  $\mathbf{u}_i$  和  $\mathbf{v}_i$  的子矩阵<sup>[18]</sup>。

在实际应用中, 对光谱数据的二维时频矩阵进行分解, 得到与时频子空间相对应的一系列奇异值和奇异值向量。矩阵  $\mathbf{A}$  表示数据的二维时频系数矩阵, 对应的  $\mathbf{u}_i$  和  $\mathbf{v}_i$  分别表示频率和时间信息, 并且奇异值的大小就表示该时频段信号能量的大小。本研究要处理的光谱数据在时频谱图上对应有用光谱信号的能量比较大, 且分布相对集中; 而噪声能量相对较小, 在时频谱中的分布也相对比较分散。因此, 可以对光谱数据进行 SVD 处理, 提取较大的奇异值, 剔除较小的奇异值, 然后进行数据重构, 从而达到抑制噪声的目的。数据重构公式<sup>[20]</sup>为

$$\mathbf{A}_{\text{re}} = \sum_{i=1}^j \sigma_i \mathbf{A}_i, \quad (9)$$

式中  $\mathbf{A}_{\text{re}}$  为去噪后数据的二维时频矩阵;  $j$  为选取较大奇异值的数量, 且  $j < k$ 。对  $\mathbf{A}_{\text{re}}$  进行广义 S 逆变换就得到去除噪声后的光谱。

### 2.2.2 去噪方法与步骤

根据 2.2.1 节的分析可知, SVD 去噪算法的关键在于选择较大奇异值的数量  $j$ 。如果选择的重构奇异值数量较多, 则 SVD 去噪效果不显著; 如果选择的重构奇异值数量较少, 则易丢失光谱数据的部分特征量。由于对应有用光谱的奇异值比较大且分布比较分散, 而对应白噪声的奇异值相对较小且分布比较集中, 因此, 可以通过分析奇异值的大小和奇异值序列子集的离散性来选择用于重构的奇异值。本研究选取标准差作为子集离散度的评价参数, 选取奇异值和奇异值子集的标准差作为奇异值序列的特征量, 采用  $k$ -均值聚类算法<sup>[21]</sup>对奇异值序列进行分类并选择重构奇异值, 具体步骤<sup>[18, 20]</sup>如下:

1) 对含噪的光谱数据进行广义 S 变换, 得到光谱数据的时频谱, 以该二维时频矩阵作为目标矩阵进行 SVD 处理, 得到奇异值序列  $\sigma_1, \sigma_2, \dots, \sigma_n$ 。

2) 构建奇异值序列的特征量  $Z_1, Z_1 = \sigma_1, \sigma_2, \dots, \sigma_n$ 。

3) 建立基于奇异值序列的子集群, 各子集  $L_i$  由集合中第  $n$  个到第  $i$  个奇异值元素组成, 即  $L_1 = \{\sigma_n, \sigma_{n-1}, \dots, \sigma_1\}, L_2 = \{\sigma_n, \sigma_{n-1}, \dots, \sigma_2\}, \dots, L_i = \{\sigma_n, \sigma_{n-1}, \dots, \sigma_i\}, \dots, L_n = \{\sigma_n\}$ 。

4) 从  $L_1$  开始计算各子集的标准偏差  $\text{STD}_i$ , 并以标准偏差集合作为奇异值序列的另一个特征量  $Z_2, Z_2 = \text{STD}_1, \text{STD}_2, \dots, \text{STD}_n$ 。

5) 采用  $k$ -均值聚类算法对奇异值序列特征量  $Z_1$  和  $Z_2$  进行分类计算, 分出对应光谱数据和对应噪声的两个类, 选择奇异值较大的分类为重构奇异值, 得出重构奇异值个数  $j$ 。

6) 由(9)式重构光谱数据的二维时频矩阵, 再通过广义 S 逆变换恢复抑制噪声后的近红外光谱。

## 3 仿真分析

根据导数光谱的波峰特征, 选择具有多峰、峰值不等特征的“Bump”信号作为仿真信号, 对其进行处理来评估本研究所提方法的有效性。无噪“Bump”信号的时域波形及其经广义 S 变换后的时频谱如图 1 所示。图 1(b)清楚地呈现了“Bump”信号能量随时频的变换情况以及能量的聚集区域, 具有很高的时频分辨率, 信号能量与时域波形对应, 呈 5 段分布在 100 Hz 以内的低频区域。给“Bump”信号加入信噪比(SNR)  $r_{\text{SNR}} = 20$  dB 的随机噪声, 模拟导数光谱中被放大的背景噪声。加噪后, “Bump”信号的时域波形及其经广义 S 变换后的时频谱如图 2 所示。

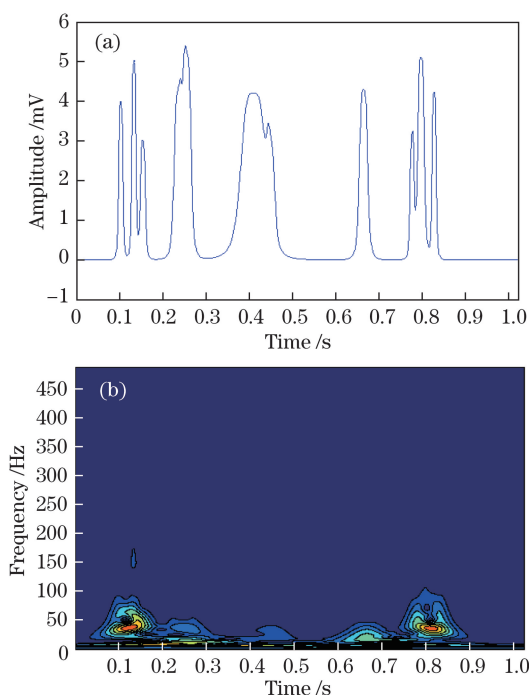


图 1 (a) 无噪“Bump”信号及(b)其广义 S 变换时频谱  
Fig. 1 (a) Original “Bump” signal and (b) its time-frequency spectrum from generalized S transform

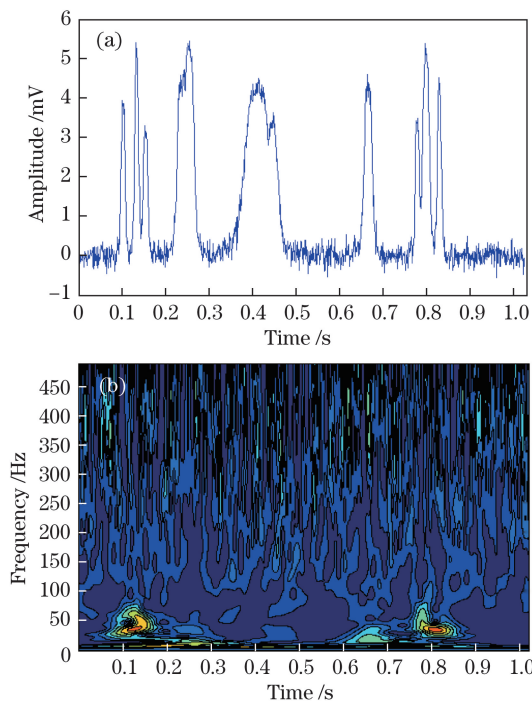


图 2 (a) 加入信噪比  $r_{SNR} = 20$  dB 的噪声后的“Bump”信号及(b)其广义 S 变换时频谱  
Fig. 2 (a) Noised “Bump” signal (signal noise ratio  $r_{SNR} = 20$  dB) and (b) its time-frequency spectrum from generalized S transform

噪声使得“Bump”信号的时域波形不再平滑,在图 2(b)中噪声的能量遍布了整个时频区域,不仅严

重干扰了低频段 100 Hz 以内原“Bump”信号的时频分布,而且导致原本洁净的高频区域也呈现出杂乱的噪声能量。

根据所提方法,选择图 2(b)所示时频谱图对应的二维时频系数矩阵作为目标矩阵进行 SVD 处理,得到奇异值序列,并建立基于奇异值序列的子集群,将各子集的标准偏差构成另一序列。基于这两个特征向量,使用  $k$ -均值聚类算法进行分类计算,选择奇异值数值较大的分类为重构奇异值,得出重构奇异值个数。计算出的奇异值序列及其聚类结果如图 3 所示,计算得到用于重构的奇异值。去噪后的“Bump”信号的时域波形及其广义 S 域时频谱如图 4 所示。由图 4 可知,附在“Bump”信号中的噪声得到了较好的抑制,时域波形重显光滑,时频谱图中高频区域的噪声能量被抹净,低频段 100 Hz 以内的能量分布及其形态特征得到了恢复。

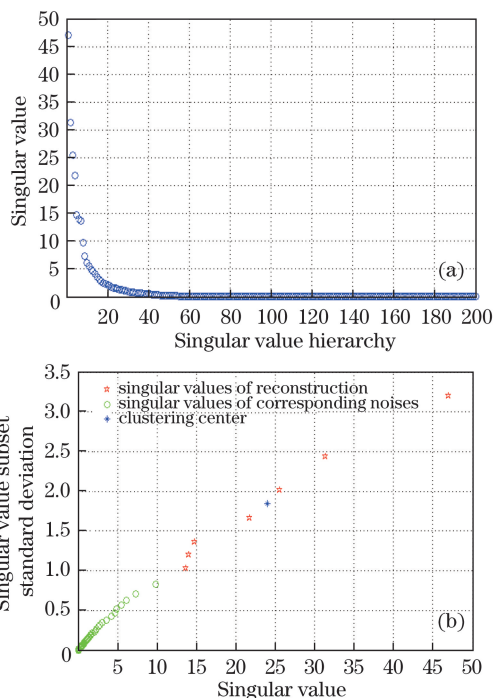


图 3 (a) 加噪“Bump”信号经广义 S 变换后时频矩阵的奇异值序列和(b)特征向量的  $k$ -均值聚类结果  
Fig. 3 (a) Singular value sequence of time-frequency matrix of noised “Bump” signal and (b)  $k$ -means clustering results of eigenvector

为了进一步评估所提方法的应用效果,给“Bump”信号加入信噪比分别为 10 dB、20 dB、30 dB 的噪声,分别采用光谱预处理中常用的 9 点平滑法、小波软阈值法(小波基的类型为 db2 小波,分解层数均为 7)和所提方法进行去噪处理。引入信噪比  $r_{SNR}$ 、均方根误差  $x_{RMSE}$ 、平滑度指数  $r_{SI}$ <sup>[4,7,21-22]</sup> 共



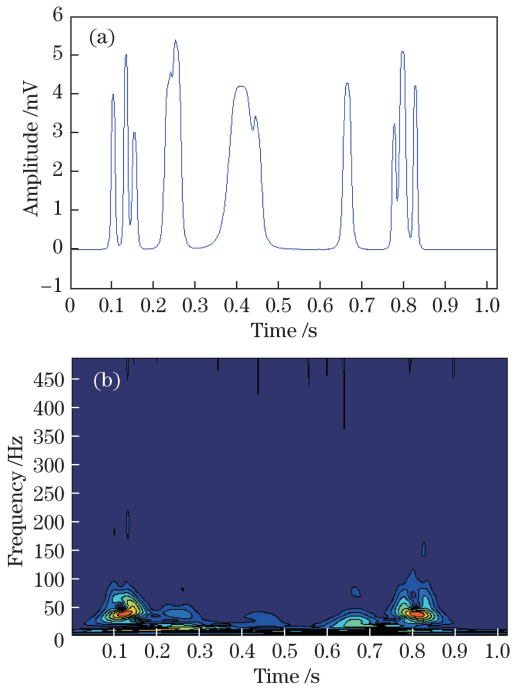


图 4 (a)去噪后的“Bump”信号及(b)其广义 S 变换时频谱  
 Fig. 4 (a) De-noised “Bump” signal and  
 (b) its time-frequency spectrum from generalized S transform  
 3 个去噪效果评价参数对各方法的去噪效果进行评价,即

$$r_{\text{SNR}} = 10 \lg \frac{\sum_{i=1}^N f^2(m'_i)}{\sum_{i=1}^N [\hat{f}(m'_i) - f(m'_i)]^2}, \quad (10)$$

$$x_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^N [\hat{f}(m'_i) - f(m'_i)]^2}{N}}, \quad (11)$$

$$r_{\text{SI}} = \frac{\sum_{i=1}^{N-1} [\hat{f}(m'_{i+1}) - \hat{f}(m'_i)]}{\sum_{i=1}^{N-1} [f(m'_{i+1}) - f(m'_i)]}, \quad (12)$$

式中  $f(m'_i)$  和  $\hat{f}(m'_i)$  分别为去噪前后的信号,  $m'$  为波段。信噪比  $r_{\text{SNR}}$  反映了信号与噪声的能量比,其值越大,去噪能力越强;均方根误差  $x_{\text{RMSE}}$  反映了去噪后光谱与原始光谱的幅值差异,其值越小,去噪效果越好;平滑度指数  $r_{\text{SI}}$  的值越接近 1,去噪后光谱的平滑度越接近原始光谱的平滑度。表 1 给出 3 种方法去噪效果评价参数的对比。由表 1 可知:3 种方法均能实现对噪声的抑制;9 点平滑法去噪结果的信噪比较小,小波软阈值法和所提方法均优于 9 点平滑方法;相比之下,本研究所提方法对噪声的抑制效果更好,去噪效果评价参数优于小波软阈值法。

表 1 9 点平滑法、小波软阈值法和本研究所提方法去噪效果评价参数的对比

Table 1 Evaluation parameters comparison of de-noising effect of 9-point smoothing method, wavelet soft threshold method, and proposed method in this paper

Parameters before de-noising		De-noising method	Parameters after de-noising		
$r_{\text{SNR}}$	$x_{\text{RMSE}}$		$r_{\text{SNR}}$	$x_{\text{RMSE}}$	$r_{\text{SI}}$
10	0.0463	9 point smoothing method	23.784	0.01826	0.967
		Wavelet soft thresholding method	30.932	0.01423	0.972
		Proposed method in this paper	37.098	0.01185	0.978
20	0.0313	9 point smoothing method	45.671	0.01022	0.983
		Wavelet soft thresholding method	57.983	0.00938	0.987
		Proposed method in this paper	61.327	0.00800	0.990
30	0.0165	9 point smoothing method	75.024	0.00654	0.991
		Wavelet soft thresholding method	84.921	0.00534	0.993
		Proposed method in this paper	88.673	0.00409	0.997

## 4 光谱数据处理

### 4.1 实验

实验仪器为 Luminar 5030 型 Mini-AOTF 便携式近红外光谱仪(NIR),波长范围为 1100 ~ 2300 nm,波长增量为 2 nm,扫描次数为 600,采用砷化镓(InGaAs 检测器)。实验样本为谷朥粉样本,共 65 个,其中校正集 35 个,验证集 30 个,数据

如表 2 所示。65 个谷朥粉样本均采用漫反射的测样方式采集光谱。对谷朥粉近红外光谱的处理都在 MATLAB 7.0 软件平台上完成,采用广义 S 变换时,通过反复实验后取  $\lambda = 1.2, p = 0.7$ ,此时可获得更精确的时频精度。

### 4.2 光谱去噪预处理

对采集到的原始光谱分别求导,35 个谷朥粉样本的原始光谱和一阶导数光谱如图 5 所示。由图 5(a)

表 2 谷朊粉样本数据  
Table 2 Data of wheat gluten samples

Sample set	Total sample number	Moisture content range / %	Average of moisture content / %
Calibration set	35	4.50-5.90	5.39
Validation set	30	4.50-5.85	5.41

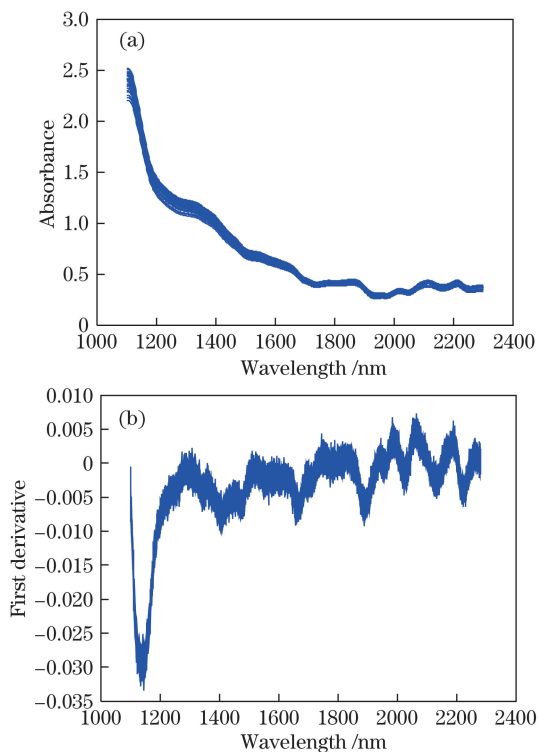


图 5 谷朊粉样本的近红外光谱。(a)原始光谱；(b)一阶导数光谱  
Fig. 5 Near infrared spectra of wheat gluten samples.

(a) Raw spectra; (b) first derivative spectra

可知,原始光谱比较光滑,但波峰特征不明显,且背景光谱非常强,基线漂移较严重,峰值点不明显,会给进一步的分析带来误差。由图 5(b)可知,导数光谱消除了基线漂移,波峰特征得以显现,但求导使随机误差也被放大,需要进一步进行光谱的去噪处理。

采用本研究提出的去噪方法对谷朊粉近红外导数光谱进行去噪处理,以 9 号样本为例进行详细阐述。图 6 所示为 9 号谷朊粉样本的一阶导数光谱及其广义 S 变换时频谱(横轴上的量为波数域的采样点,纵轴上的量为类比的归一化频率)。由图 6(a)可知,光谱峰值特征得以显现,但同时随机误差也被放大,光谱信噪比减小。由图 6(b)可知,噪声的存在使得能量的分布杂乱无章。

采用本研究提出的去噪方法对图 6(b)所示时频谱图的时频矩阵进行 SVD 处理,计算出的奇异值序列及其聚类结果如图 7 所示。由图 7 可知,重构

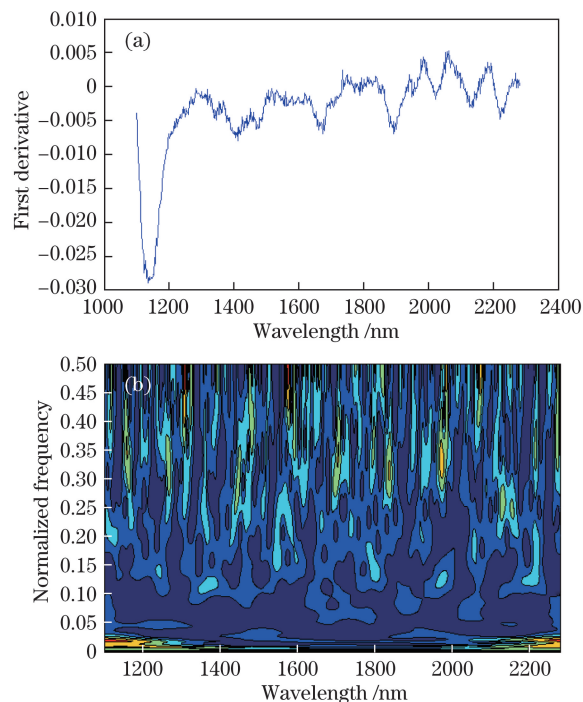


图 6 (a) 9 号谷朊粉样本的一阶导数光谱及 (b) 其广义 S 变换时频谱

Fig. 6 (a) First derivative spectrum of No. 9 wheat gluten sample and (b) its time-frequency spectrum from generalized S transform

奇异值数量  $j = 2$ 。去噪后 9 号谷朊粉样本导数光谱及其广义 S 域时频谱如图 8 所示。由图 8 可知,时频谱图很干净,仅在低频段有能量分布,导数光谱的噪声得到了有效抑制,吸收峰光滑无毛刺,有利于确定吸收峰的准确位置。

由于无法测得不含噪声的 9 号谷朊粉样本的导数光谱,不能使用去噪评价参数对光谱的去噪效果进行评价,进而引入噪声抑制比  $\eta_{\text{pnr}}$  和幅值衰减比  $\eta_{\text{par}}$  [18,23] 来评价光谱的去噪效果。噪声抑制比和幅值衰减比的表达式为

$$\eta_{\text{pnr}} = 10(\lg \delta_1^2 - \lg \delta_2^2), \quad (13)$$

$$\eta_{\text{par}} = \frac{A_{m1} - A_{m2}}{A_{m1}} \times 100\%, \quad (14)$$

式中  $\delta_1$  和  $\delta_2$  分别为去噪前后光谱的标准偏差,  $A_{m1}$  和  $A_{m2}$  分别为去噪前后光谱的最大幅值。 $\eta_{\text{pnr}}$  反映了去噪前后光谱波峰的凸显程度,  $\eta_{\text{par}}$  则反映了去噪前后导数光谱的衰减程度。同样,采用光谱预处理中常用的 9 点平滑法、小波软阈值法(小波基为 db2 小波,分解层数均为 7)进行对比分析,并引入耗时参数来评估 9 点平滑法、小波软阈值法和本研究所提方法的计算效率,3 种方法去噪效果的评价参数如表 3 所示。由表 3 可知:与传统 9 点平滑法相比,

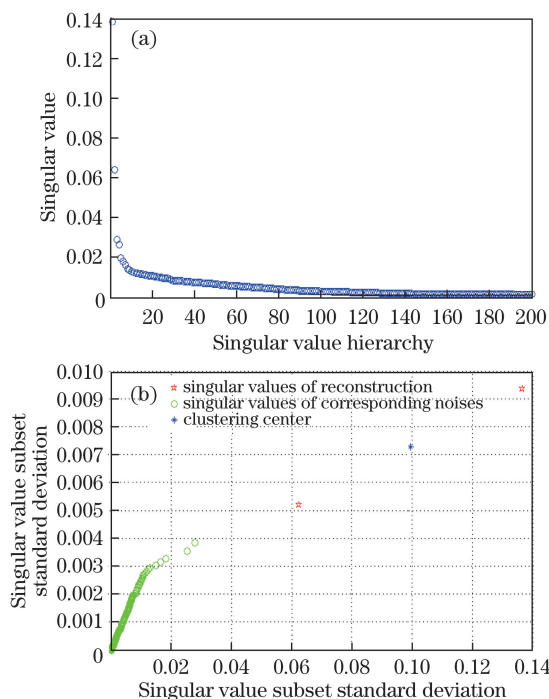


图 7 (a) 导数光谱广义 S 变换时频矩阵的奇异值序列和 (b) 特征向量的 k-均值聚类结果  
Fig. 7 (a) Singular value sequence of time-frequency matrix of derivative spectra and (b) k-means clustering results of eigenvector

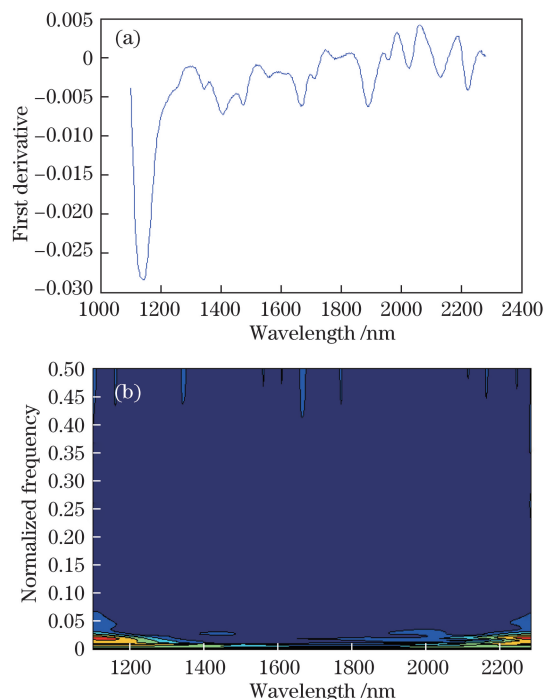


图 8 (a) 去噪后 9 号样本的一阶导数光谱及 (b) 其广义 S 变换时频谱  
Fig. 8 (a) De-noised first derivative spectrum of No. 9 sample and (b) its time-frequency spectrum from generalized S transform

表 3 9 点平滑法、小波软阈值法和本研究所提方法对 9 号光谱样本来噪效果评价参数的对比

Table 3 Evaluation parameters comparison of de-noising effect of No. 9 spectrum sample using 9-point smoothing method, wavelet soft thresholding method, and proposed method in this paper

De-noising method	Noise reduction ratio	Magnitude attenuation ratio /%	Time consuming /s
9-point smoothing method	13.18	28.66	0.031
Wavelet soft thresholding method	15.69	24.66	0.910
Proposed method in this paper	21.54	19.76	5.750

本研究所提算法的  $\eta_{par}$  由 28.66% 减至 19.76%，光谱的幅值衰减程度较低； $\eta_{pnr}$  由 13.18 增至 21.54，有利于微小变化波峰的检测，很好地保留了光谱信号的特征尖峰点；所提方法对导数光谱的噪声抑制具有较好的效果，且优于小波软阈值法；然而相对于其他两种方法，本研究提出的组合去噪方法的计算量较大，计算时间较长。

采用上述方法，对其他 64 个样本的光谱进行滤波处理。图 9 所示为去噪后 65 个谷朥粉样本的一阶导数光谱。由图 9 可知，在滤除噪声的同时，光谱的峰形保持得很好，且反映原始光谱的特征峰值点得以显现，为下一步的物质成分检测奠定了基础。

### 4.3 水分含量检测

随机选取 30 个样本，分别采用 9 点平滑法、小

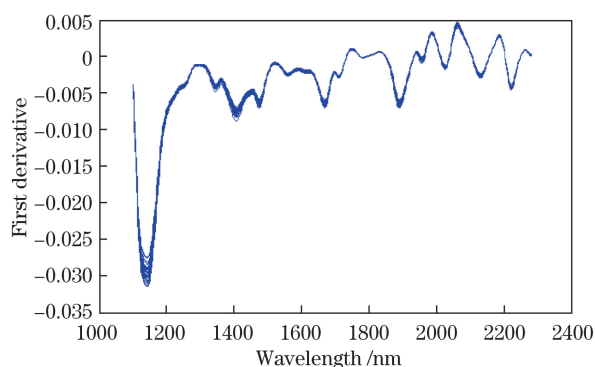


图 9 去噪后 65 个谷朥粉样本的一阶导数光谱  
Fig. 9 First derivative spectra of 65 wheat gluten samples after de-noising

波软阈值法和本研究所提方法对一阶导数光谱进行去噪处理。将去噪后的光谱数据与谷朥粉中的水分

等化学成分的基础数据进行关联,采用偏最小二乘法和交叉验证,用定量分析软件 The Unscrambler 建立模型,分析谷朊粉中的水分含量。在波段选择、回归算法等因素一致的情况下,比较不同预处理技术对谷朊粉水分含量检测的影响,并用决定系数  $r^2$  和预测均方根误差  $x_{\text{RMSE,p}}$  进行评估,即

$$r^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (15)$$

$$x_{\text{RMSE,p}} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N - 1}}, \quad (16)$$

式中  $N$  为预测集样本数,  $\hat{y}_i (i=1, 2, \dots, N)$  为样本的预测值,  $\bar{y}$  为  $\hat{y}_i$  的均值,  $y_i$  为对应的参考值。决定系数  $r^2$  的取值范围为  $0 \sim 1$ , 越接近 1, 模型的拟合优度越高。预测均方根误差  $x_{\text{RMSE,p}}$  是评估所有样本预测值与参考值之间差异的标准偏差, 代表模型的整体残余误差, 其值越小越好。图 10 所示为 3 种去噪方法对谷朊粉近红外光谱水分含量预测结果与化学分析结果的比较。由图 10 可知, 所提方法的去噪效果最好, 相比 9 点平滑法, 预测集的决定系数由 0.9436 增大为 0.9985, 预测均方根误差由 0.0843 减小为 0.0406, 有效地提高了光谱的分析精度和模型的稳健性。

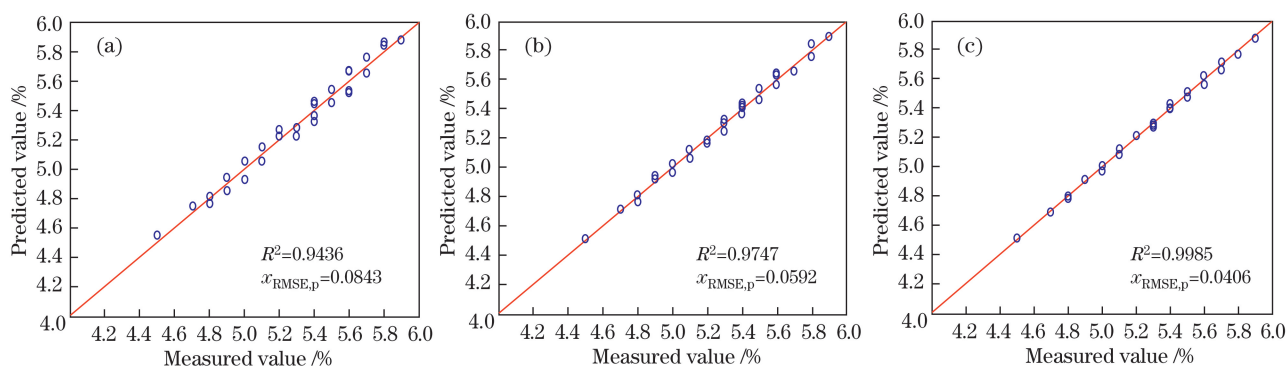


图 10 3 种去噪方法对谷朊粉样本近红外光谱水分含量预测结果与化学分析结果的比较。

(a) 9 点平滑法; (b) 小波软阈值法; (c) 本研究所提方法

Fig. 10 Comparison of near infrared moisture content prediction results obtained from three de-noising methods and measured values of wheat gluten samples. (a) 9-point smoothing method; (b) wavelet soft thresholding method; (c) proposed method in this paper

## 5 结 论

基于广义 S 变换的时频滤波方法克服了单维度(时域或频域)滤波的局限性, 无需选择小波基函数和参考噪声数据; 对谷朊粉导数光谱的二维时频矩阵进行 SVD 处理, 光谱对应的奇异值较大, 噪声对应的奇异值较小, 采用  $k$ -均值聚类算法, 选择较大的奇异值重构光谱可以有效提高滤波效果(实例中只采用了 2 个奇异值), 降低了滤波过程的复杂度。

采用所提方法处理后, 近红外光谱的分析精度和模型的稳健性明显优于传统的 9 点平滑法和小波软阈值法; 相比 9 点平滑法, 预测集的决定系数由 0.9436 增大为 0.9985, 预测均方根误差由 0.0843 减小为 0.0406, 提高了谷朊粉中水分含量定量检测的精度, 为谷朊粉近红外光谱预处理提供了新的途径。

值得一提的是, 虽然广义 S 变换已在众多学科

中得到了广泛应用, 但 2 个窗函数调节参数的选取问题还有待进一步完善; 另外, 相对于小波变换等去噪方法, 本研究提出的去噪方法的计算效率不高, 需进一步寻求简化运算, 以提高其在近红外光谱现场检测或在线检测中的适用性; 所提方法也可应用到其他光谱预处理中, 具有广阔的应用前景。

## 参 考 文 献

[1] Gao G M, Li X, Qin Z D, *et al.* New method for eliminating background fluorescence of Raman spectrum and its application[J]. Acta Optica Sinica, 2013, 33(2): 0230002.  
高国明, 李雪, 覃宗定, 等. 消除拉曼光谱荧光背景的新方法及其应用[J]. 光学学报, 2013, 33(2): 0230002.

[2] Wu C X, Yang M W, Yang Y H, *et al.* Effect of fiber bend on interferometric fiber optical gyroscopes performance[J]. Acta Optica Sinica, 2014, 34(3): 0306005.



- 吴长莘, 杨明伟, 杨远洪, 等. 光纤弯曲对干涉式光纤陀螺性能的影响[J]. 光学学报, 2014, 34(3): 0306005.
- [3] Xie J, Pan T, Chen J M, *et al.* Joint optimization of Savitzky-Golay smoothing models and partial least squares factors for near-infrared spectroscopic analysis of serum glucose[J]. Chinese Journal of Analytical Chemistry, 2010, 38(3): 342-346.
- 谢军, 潘涛, 陈洁梅, 等. 血糖近红外光谱分析的 Savitzky-Golay 平滑模式与偏最小二乘法因子数的联合优选[J]. 分析化学, 2010, 38(3): 342-346.
- [4] Tian G Y, Yuan H F, Chu X L, *et al.* Near infrared spectra (NIR) analysis of octane number by wavelet denoising-derivative method[J]. Spectroscopy and Spectral Analysis, 2005, 25(4): 516-520.
- 田高友, 袁洪福, 褚小立, 等. 结合小波变换与微分法改善近红外光谱分析精度[J]. 光谱学与光谱分析, 2005, 25(4): 516-520.
- [5] Li S W, Xie P H, Li Y J, *et al.* Wavelet transform based differential optical absorption spectroscopy data processing[J]. Acta Optica Sinica, 2006, 26(11): 1601-1604.
- 李素文, 谢品华, 李玉金, 等. 基于小波变换的差分吸收光谱数据处理方法[J]. 光学学报, 2006, 26(11): 1601-1604.
- [6] Hao Y, Chen B, Zhu R. Analysis of several methods for wavelet denoising used in near infrared spectrum pretreatment[J]. Spectroscopy and Spectral Analysis, 2006, 26(10): 1838-1842.
- 郝勇, 陈斌, 朱锐. 近红外光谱预处理中几种小波消噪方法的分析[J]. 光谱学与光谱分析, 2006, 26(10): 1838-1842.
- [7] He F. The application of wavelet transform in spectrum data processing[D]. Tianjin: Tianjin University of Technology, 2012.
- 贺芳. 小波去噪方法在光谱数据处理中的应用[D]. 天津: 天津理工大学, 2012.
- [8] Cai J H, Wang X C. Near-infrared spectrum pretreatment based on empirical mode decomposition[J]. Acta Optica Sinica, 2010, 30(1): 267-271.
- 蔡剑华, 王先春. 基于经验模态分解的近红外光谱预处理方法[J]. 光学学报, 2010, 30(1): 267-271.
- [9] Chen C, Lu Q P, Peng Z Q. Preprocessing methods of near-infrared spectrum based on NLMS adaptive filtering[J]. Acta Optica Sinica, 2012, 32(5): 0530001.
- 陈丛, 卢启鹏, 彭忠琦. 基于 NLMS 自适应滤波的近红外光谱去噪处理方法研究[J]. 光学学报, 2012, 32(5): 0530001.
- [10] Cai J H, Wang X C, Hu W W. Near-infrared spectrum detection of tobacco nicotine content based on morphological wavelet[J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(15): 281-286.
- 蔡剑华, 王先春, 胡惟文. 基于形态小波的烟草尼古丁含量近红外光谱检测[J]. 农业工程学报, 2012, 28(15): 281-286.
- [11] Stockwell R G, Mansinha L, Lowe R P. Localization of the complex spectrum: The S transform[J]. IEEE Transactions on Signal Processing, 1996, 44(4): 998-1001.
- [12] Pinnegar C R, Mansinha L. The S-transform with windows of arbitrary and varying shape[J]. Geophysics, 2003, 68(1): 381-385.
- [13] Yin B Q, He Y G, Wu X M. A method for magnetocardiograms filtering based on singular value decomposition and S-transform[J]. Acta Physica Sinica, 2013, 62(14): 148702.
- 尹柏强, 何怡刚, 吴先明. 心磁信号广义 S 变换域奇异值分解滤波方法[J]. 物理学报, 2013, 62(14): 148702.
- [14] Jing J E, Wei W B, Chen H Y, *et al.* Magnetotelluric sounding data processing based on generalized S transformation[J]. Chinese Journal of Geophysics, 2012, 55(12): 4015-4022.
- 景建恩, 魏文博, 陈海燕, 等. 基于广义 S 变换的大地电磁测深数据处理[J]. 地球物理学报, 2012, 55(12): 4015-4022.
- [15] Guo Y J, Wei Y D, Zhou X J, *et al.* Impact feature extraction from rolling bearing fault signal by S transform[J]. Journal of Vibration, Measurement & Diagnosis, 2014, 34(5): 818-822.
- 郭远晶, 魏燕定, 周晓军, 等. S 变换用于滚动轴承故障信号冲击特征提取[J]. 振动、测试与诊断, 2014, 34(5): 818-822.
- [16] Simon C, Ventosa S, Schimmel M, *et al.* The S-transform and its inverses: Side effects of discretizing and filtering[J]. IEEE Transactions on Signal Processing, 2007, 55(10): 4928-4937.
- [17] Schimmel M, Gallart J. The inverse S transform in filters with time-frequency localization[J]. IEEE Transactions on Signal Processing, 2005, 53(11): 4417-4422.
- [18] Liu Y S, Zhou W J, Li P F, *et al.* Partial discharge ultrahigh frequency signal denoising method based on generalized S-transform modular time-frequency matrix[J]. Transactions of China Electrotechnical Society, 2017, 32(9): 2011-2020.
- 刘宇舜, 周文俊, 李鹏飞, 等. 基于广义 S 变换模时频矩阵的局部放电特高频信号去噪方法[J]. 电工技术学报, 2017, 32(9): 2011-2020.
- [19] Cong F Y, Chen J, Dong G M, *et al.* Short-time

- matrix series based singular value decomposition for rolling bearing fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2013, 34(1/2): 218-230.
- [20] Zhu Y, Jiang S Y. Rolling bearing default impact feature extracting method based on generalized S transform time-frequency spectrum denoised by SVD[J]. *Bearing*, 2016(11): 53-57.  
朱怡, 蒋思源. 广义 S 变换时频谱 SVD 降噪的滚动轴承故障冲击特征提取方法 [J]. *轴承*, 2016(11): 53-57.
- [21] Kanungo T, Mount D M, Netanyahu N S, *et al.* An efficient  $k$ -means clustering algorithm: Analysis and implementation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 881-892.
- [22] Yang Z, Wang Y T, Pan Z. Fluorescence spectrum denoising method for low concentration petroleum pollutants based on EMD-LWT[J]. *Acta Optica Sinica*, 2016, 36(5): 0530001.  
杨哲, 王玉田, 潘钊. 基于 EMD-LWT 的低浓度石油类污染物荧光光谱去噪法 [J]. *光学学报*, 2016, 36(5): 0530001.
- [23] Chen Z W, Zhang F, Zhou Y, *et al.* Denoising method of spectral signal with multiplicative and additive mixed random noises[J]. *Acta Optica Sinica*, 2017, 37(7): 0730001.  
陈正伟, 张方, 周扬, 等. 光谱信号乘性加性混合随机噪声去除方法 [J]. *光学学报*, 2017, 37(7): 0730001.