Multidimensional Blind Deconvolution Method Based on Cross-Sparse Filtering for Weak Fault Diagnosis

SHAN WANG1, ZONGZHEN ZHANG1,2, JINRUI WANG1 and BAOKUN HAN1
1College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266000, China
2College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Corresponding author: Zongzhen Zhang (zhzz18@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 52005303, and in part by the Project of China Postdoctoral Science Foundation under Grant 2019M662399.

ABSTRACT Fault diagnosis and condition monitoring of rotating machinery has drawn considerable attention. The complex structure of rotating machinery and poor working conditions cause two challenges: weak signature detection (WSD) and weak compound fault separation (WCFS). A superior method should realize these two functions simultaneously. This paper proposed a multidimensional blind deconvolution method based on cross-sparse filtering (Cr-SF) for WSD and WCFS, which can enhance the weak signature and decompose the different components from compound fault adaptively without any preprocessing and priori knowledge. Cross kurtosis pursuit (CKP), a novel filter selection technology, is proposed for determining the final filters. The experimental and simulated signals verified the performance of the proposed algorithm. The robustness is also investigated using the success rate of repeated experiments. The results indicate that Cr-SF can handle different fault compounding modes under the strong noise environment and perform strong robustness and noise adaptability.

INDEX TERMS Unsupervised learning, cross-sparse filtering, blind deconvolution, weak signature detection, compound fault separation.

I. INTRODUCTION Rotating machinery is an important component of mechanical equipment. However, faults of rotating machinery reduce the efficiency of machinery and cause potential safety hazards. Impulsive signature and amplitude modulation (AM) indicate potential machine breakdowns in bearings and gears. However, the early fault signature is always corrupted by ambient noise [1], [2]. Inducing other faults is also easy when the rotating machinery has local faults. Therefore, weak compound fault detection of rotating machinery is important and the key is the separation of different fault features in the compound fault. Currently, weak signature detection (WSD) and weak compound fault separation (WCFS) have also been investigated widely [3]–[6].

Several techniques for WSD and WCFS have been proposed. An effective tool proposed by Antoni, Spectral Kurtosis (SK) is uses a dyadic filter bank to determine the frequency band with maximum kurtosis value [7]–[11]. However, the error of optimized bandwidth and calculation efficiency case the SK to be out of the field of online fault detection. In addition, SK encounters difficulties in the detection of harmonic features [12]. Wavelet transform determines a sparse representation of measured signal according to the proper wavelet basis [13], [14]. In [15], a discrete wavelet transform was employed to diagnose the multi-fault of ball bearings. However, prior experience is important for selecting the proper mother wavelet. Empirical mode decomposition (EMD) [16], [17] and variational mode decomposition (VMD) [18]–[20] can effectively extract harmonic features and are used widely in machinery fault diagnosis. Jiang et al. proposed a coarse-to-fine decomposing strategy for weak signature detection to address the dilemmas of VMD during the applications such as the determination of the input parameters [21]. Miao et al. proposed an improved parameter-adaptive VMD based on ensemble kurtosis to
identify the mechanical compound fault [22]. However, EMD and VMD methods perform poorly in the detection of impulsive signature.

Other methods focused on the sparse optimization of raw data. Minimum entropy deconvolution (MED) [23], [24] employs kurtosis to obtain the sparse features of the raw signal. MED and its variants have been widely used in the fault diagnosis of rolling bearings [25] and gears [26], [27]. However, the robustness of MED is poor when strong noise exists and always converges to the largest peak quickly in the application. Maximum correlated kurtosis deconvolution (MCKD), which was developed from MED, is an efficient tool for compound fault separation [28]. In [29], MCKD was used to diagnose the compound bearing fault. Miao et al. developed an improved MCKD, which estimates the iterative period by calculating the autocorrelation of the envelope signal [30]. However, MCKD requires a priori basis to determine input parameters and performs poorly in the detection of meshing components. In [31], morphological component analysis (MCA) was used to separate the periodic impact signal and meshing components in compound faults. However, the selection of proper dictionaries affects the separation performance of compound faults. MCA will also extract different impulsive signals as one fault.

Many studies have investigated the WSD and WCFS. However, most studies have limitations, such as performance gaps in applications in different fault modes and reliance on prior experience and knowledge.

A deep learning method can learn the key features of the samples automatically and eliminate the reliance on prior knowledge and human labor. In recent years, deep learning-based methods are applied gradually to rotating machinery fault diagnosis, such as intelligent fault classification [32], [33] and remaining useful life (RUL) prediction [34], [35]. The advantages of the machine learning method for reference are necessary to improve the calculation efficiency and intelligence of WSD and WCFS. However, using machine learning methods raises problems in WSD and WCFS. First, the training process is time-consuming. Second, feature learning requires a large number of samples. Third, the noise adaptability is poor. Fourth, feature learning process requires extensive hyperparameter tuning. Therefore, the machine learning method is seldom used in WSD and WCFS. Sparse filtering can obtain multiple filters [36], [37] and Jia proposed a novel method based on the convolutional sparse filter (CSF) and PCA for WSD [38]. However, WCFS using CSF has not been studied. The robustness of SF is also degraded in strong noise environment.

To achieve the desired properties, a multidimensional blind deconvolution method based on Cr-SF is proposed for WSD and WCFS. The performance of the Cr-SF is validated by the simulated fault signals and the experimental data. The robustness and efficiency of Cr-SF under various noise environments are also discussed using the success rate and average time of 50 trails. The results confirmed the superior performance and stronger robustness of the proposed compared to the existing methods.

The rest of this paper is organized as follows. The motivation of this study and sparse optimization-based methods are discussed in section 2. In section 3, the details of the proposed Cr-SF are discussed and the cross-optimization process is introduced. Then, the blind deconvolution method based on Cr-SF and CKP is proposed. In sections 4 and 5 the proposed method is validated by the simulated and experimental data based on WSD and WCFS. Finally, the conclusions are drawn in section 6.

II. SPARSE OPTIMIZATION-BASED METHOD

Blind Deconvolution (BD) algorithms are effective methods for WSD such as MED, optimal MED, and generalized MED (GMED). MED filter is a non-parametric method first proposed in Geophysics. The kurtosis norm of the input signal is used to pursue an inverse filter that can represent the input signal with sparsity. The core idea of MED is to obtain the optimal filter to maximize the kurtosis of the measured signal, recover the raw signal, and highlight the impulse component.

The higher-order moment of the inputs is maximized instead of the kurtosis norm in the GMED. The results using planetary gear in [27] show that GMED can achieve sparser representation. The objective function for the GMED is presented as:

$$
\max_{x f_{inv}} \frac{\sum_i x_i^q}{(\sum_k x_k^2)^{p/q}} \quad S.T. \quad x = y \ast f_{inv} \quad (1)
$$

The cost function of the MED filter can be written as

$$
\max_{x f_{inv}} \frac{\sum_i x_i^4}{(\sum_k x_k^2)^2} \quad S.T. \quad x = y \ast f_{inv} \quad (2)
$$

where \( f_{inv} \) are the coefficients for the inverse filter, \( x \) is the input sample, and \( y \) is the extracted sparse signal.

In [39], OMED maximized the D-norm to pursue an inverse filter using a non-iterative algorithm. The cost function for the OMED filter can be described as:

$$
\max_{x f_{inv}} (D(x)) = \max_{x f_{inv}} \left( \max_k \frac{|x_k|}{\|X\|_{l_2}} \right) \quad S.T. \quad x = y \ast f_{inv} \quad (3)
$$

BD is mathematically similar to a sparse feature learning process. Sparsity is a core concept in machine learning in which high dimensional input samples can be represented using a desired small number of significant values. Sparse filtering achieves feature sparsity by normalizing the rows and columns of the feature matrix using \( l_2 \) norms. Jia et al.
proposed a novel sparse filter based on the generalized \( l_p/l_q \) norm to enhance impulsive signature [40]. The algorithm views the activation function as a convolution process. The sparse features trained by sparse filtering are considered as impulsive components.

\[
\min_f J_{pq}(f) = \min_f \sum_i^M \left( \frac{\sum_j^N |f_j|^p}{\sum_j^N |f_j|^q} \right)^{\frac{q}{p}} \quad \text{S.T.} \quad f = H \cdot W
\]

\[\|W\|_2 = 1, \quad p < q \quad (4)\]

where \( f \) is the sparse feature of the input data, \( H \) is the Hankel matrix derived from the input data, and \( W \in \mathbb{R}^L \) denotes the weight vector.

The sparse filter with generalized \( l_p/l_q \) norm is still unable to solve the problem of compound fault separation. Standard sparse filtering can obtain multiple filters simultaneously. In [38], convolutional sparse filter (CSF) is employed to enhance the impulsive signature. Dimension reduction technology such as PCA is employed to reduce the dimension of the feature matrix extracted by CSF. CSF can separate compound faults, but the research in this area has not been carried out.

\[
\min_f \text{CSF}(f) = \min_f \sum_i^M \left( \frac{\sum_j^N |f_j|}{\sum_j^N |f_j|^2} \right)^{\frac{1}{2}} \quad \text{S.T.} \quad \hat{f}_j = \frac{f_j}{\|f\|_2} = H \cdot W
\]

(5)

where \( W \) refers to the weight matrix of CSF, and \( f \) is the sparse feature extracted from the input data.

The advantage of CSF is simultaneous learning of multiple filters. However, CSF requires PCA to obtain the optimal filter, which shows poor robustness with low output dimension. In addition, the optimization time of CSF increases significantly with the increase of output dimension and the robustness is obviously degraded in the strong noise environment.

In addition, given that the multi-dimensional deconvolution method can learn multiple filters simultaneously, the optimal filter selecting method is necessary to determine the optimal filter. Based on the above discussion, this paper proposes a novel multidimensional blind deconvolution method based on Cr-SF and CKP.

**III. PROPOSED METHOD**

In this section, a novel multidimensional blind deconvolution method based on cross-sparse filtering is proposed for WSD and WCFS. Cross-kurtosis pursuit is proposed to reduce the filters dimension and determine the optimal filters.

**A. CROSS-SPARSE FILTERING**

As shown in Fig. 1, Cr-SF can be viewed as a single-layer neural network, which is mainly divided into three steps:

1. **Convolutional activation.** We used the convolution of the weights and the input signal as the feature activation process. Concretely, supposing that \( L \) and \( N_f \) are the input and output dimensions of Cr-SF, respectively. First, Hankel matrix \( H \in \mathbb{R}^{L \times (N-L+1)} \) is constructed using the input signal \( x \in \mathbb{R}^N \), where \( L \) denotes the filter length, and \( N_f \) means the number of training filters.

\[
H = \begin{bmatrix}
    x_1 & x_2 & x_3 & \cdots & x_{N-L+1} \\
    x_2 & x_3 & x_4 & \cdots & x_{N-L+2} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    x_{L} & x_{L+1} & x_{L+2} & \cdots & x_{N}
\end{bmatrix}
\]

For more intuitive description, the convolution process is replaced by the linear expression of \( H \) and \( W, f = WH \), where \( f \in \mathbb{R}^{N_f \times (N-L+1)} \) is the sparse features.

2. **Construction of objective function.** SF ensures the weights and the input signal as the feature activation process. However, equal activation will lead to feature consistency between features. Therefore, we approached the sparse feature extraction by using cross-normalization. Concretely, row normalization of feature matrix is used to guarantee the sparse distribution of each feature.

\[
\mathcal{L}_1 = \left\| \frac{F^i}{\|F^i\|_2} \right\|_1
\]

Then, the column normalization of the feature matrix is used to achieve the sparse distribution within features.

\[
\mathcal{L}_2 = \left\| \frac{F_j}{\|F_j\|_2} \right\|_1
\]

Each weight vector is constrained to unit vectors to eliminate the influence of redundancy in the optimization process. The objective function of Cr-SF can be expressed as follows:

\[
\mathcal{L}_{Cr-SF} = \sum_{i}^{(N-L+1)} \mathcal{L}_1 + \sum_{j}^{L} \mathcal{L}_2 \quad \text{S.T.} \quad \|W^k\|_2 = 1 \quad (9)
\]

3. **Optimization.** \( |F| \) is replaced by the soft-absolute function \( \sqrt{F^2 + \varepsilon} \) due to \( \mathcal{L}_{Cr-SF} \) is nonsmooth and nonconvex, where \( \varepsilon \) is a small positive number. In this paper, \( \varepsilon = 1 \times 10^{-8} \). Then, L-BFGS is used to optimize the objective.
function, the gradient function is expressed as follows:
\[
\frac{\partial L_{Cr-SF}}{\partial w} = \left[ \left( \frac{O}{\left( \sum_j^L |F_j|^2 \right)^{\frac{3}{2}}} \right) - \frac{P^2 \cdot \sum_j^L (F_j^t F_j)}{\left( \sum_j^L |F_j|^2 \right)^{\frac{3}{2}}} \right] \\
+ \left[ \left( \frac{O}{\left( \sum_i^M |F_i|^2 \right)^{\frac{3}{2}}} \right) - \frac{P^2 \cdot \sum_i^{M(L-1)} (F_i)}{\left( \sum_i^M |F_i|^2 \right)^{\frac{3}{2}}} \right] \\
\cdot \frac{f}{|f|} \cdot x'
\]

where \( o \in \mathbb{R}^{L \times (N-L+1)} \) is a matrix of all ones.

Fig. 2 presents a 2D explanation of the cross \( l_{1/2} \)-norm. Cr-SF is a cross-optimization process, and the column and row features are forced to move close to the coordinate axis simultaneously. In comparison, the optimization process of standard sparse filtering is divided into two steps. The column features are forced to move to the coordinate axis, and the row features will be differentiated due to the competition caused by normalized constraints.

![FIGURE 2. 2D explanation of the optimization of Cr-SF.](image)

**B. CROSS KURTOSIS PURSUIT**

The proposed Cr-SF will obtain multiple filters. Due to sparse constraints, not all the filters can extract the desired components. Therefore, we need to reduce the dimension of filters or filtered signals to extract the principal components. In [38], PCA and a diffusion map are used to pursue the optimal filters. However, as mentioned above, PCA has poor robustness when the dimension of filters is low, and the optimization time of Cr-SF will increase significantly when the dimension is high. To overcome this weakness, cross Kurtosis pursuit is proposed for achieving robustness filter selection under low-dimension.

First, we obtained the spectrum of each filter using the Fourier transform. Then, we derived the kurtosis of the spectrum distribution of the filter and obtain threshold \( K \). The number of filters \( N_k \) is the number of principal components to be extracted with the kurtosis values larger than the threshold. Then the kurtosis of each spectrum is calculated and the position of the maximum peak value of the spectrum is recorded. The filter is selected according to the kurtosis value. When the peak position is the same as any one that has been selected, this filter is skipped. Stop searching when the number of filters equals \( N_k \), or when the kurtosis value is less than the threshold value \( K \).

As the kurtosis is scale invariant, the threshold \( K \) can be obtained by experiment. We studied and summarized the kurtosis and found the best threshold through repeated experiments. Another advantage of CKP is the early warning in practical application. Once the calculation fails without warning, judging the fault information from the time-domain waveform and spectrum is a waste of time. If all the kurtosis values retrieved are less than the threshold value, then the algorithm has not succeeded in extracting the features.

**C. DIAGNOSI METHOD BASED ON CR-SF AND CKP**

The implementation of the diagnosis method based on Cr-SF and CKP is displayed on the flow chart Fig. 3.

![FIGURE 3. Flow chart of the proposed method.](image)
FIGURE 4. (a) Simulated bearing fault signal, (b) simulated gear fault signal.

FIGURE 5. (a) Time waveform of simulated signal with SNR = -13dB, (b) Hilbert spectrum of the simulated gear fault, (c) time waveform of the exacted component using Cr-SF, (d) Hilbert spectrum of the exacted component using Cr-SF, (e) time waveform of the exacted component using CSF, (f) Hilbert spectrum of the exacted component using CSF, (g) time waveform of the exacted component using MED, (h) Hilbert spectrum of the exacted component using MED.

IV. WEAK SIGNATURE DETECTION USING THE PROPOSED METHOD

In this section, the effectiveness of Cr-SF is demonstrated using the simulated signal and experimental data.

A. SIMULATION ANALYSIS

1) DATA DESCRIPTION

As shown in Fig. 4 (a), simulated vibration series for bearing outer-race failure takes the form

\[ x(t) = \sum_i A(t_i)S_b(t - iT_b - \delta T) + n(t) \]  \hspace{1cm} (11)

where

\[ B(t) = \left[ 1 - \cos(2\pi f_r t) \right] / 2 \]  \hspace{1cm} (12)

\[ S_b(t) = e^{-\alpha t}\sin(2\pi f_r t) \]  \hspace{1cm} (13)

\[ 0.5 \leq T_b \leq 1 \]  \hspace{1cm} (14)

where \( A \) and \( B(t) \) simulate the amplitude coefficient and amplitude modulation respectively. \( S_b(t) \) denotes the impulse response that can be expressed as (13). \( T_b \) and \( f_r = 1/T_b \) refer to the time interval of two impulse component and the frequency of the impact signal. \( \delta T \) denotes a random time variable caused by the slippage of rolling elements, which usually accounts for 1–2% of \( T_b \). \( f_r \) specifies the resonant frequency and \( \alpha \) is the coefficient of resonance damping.

The vibration signal of the gearbox with the local fault contains amplitude and phase modulations. The simulated gear fault signal is shown in Fig. 4 (b). The vibration signal of gear fault collected by a sensor can be simulated as

\[ h(t) = \sum_{m=0}^{M} X_m(1 + a_m(t))\cos(2\pi mzf_n + \phi_m) + b_m(t) + n(t) \]  \hspace{1cm} (14)

where \( X_m \) denotes the amplitude coefficient and \( \phi_m \) is the phase of the \( m \)th meshing harmonic. \( M \) is the number of tooth-meshing harmonics, \( f_n \) denotes the frequency of rotation rate, \( z \) is the tooth number of the gear, and \( a_m \) and \( b_m \) are the periodic functions. \( n(t) \) in Eq. 11 and 14 refer to the random interference simulated using Gaussian noise. Signature to
noise ratio ($SNR_{dB}$) can be given as follows

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right)$$

where $P_{signal}$ is the power of the signal and $P_{noise}$ the power of the noise.

2) RESULTS ANALYSIS

Although CSF can compare the performance of MED in the case of one-dimensional output, the amplitude modulation features cannot be extracted. We investigated the case of multi-dimension, $N_f = 10$, $\lambda = 1$, $L = 100$. The simulated signals are set as follows: $f_s = 20$ kHz, $f_n = 40$ Hz, $z = 45$, $X_m = 10$, $f_s = 2000$ Hz, $T_b = 0.025$ s and $\alpha = 500$. Given that only one fault is found in the input signal. The maximum kurtosis is taken as the threshold, and the number of dominant components is set as 1. The detection results are shown Fig. 5 when $SNR_{dB} = -13$dB. It can be seen that only the proposed Cr-SF can extract the fault components. The proposed method shows better reconstruction performance in the time domain than other methods. Hence, better performance can be seen from the results of MED, the time domain waveform recovered poorly. Clearly, the time interval between the two impacts can be observed. The results indicated that the proposed Cr-SF showed stronger noise adaptability than CSF.

3) ROBUSTNESS AND OPTIMIZATION EFFICIENCY

Increasing the training sample number can improve training accuracy; therefore, in practical applications, we can increase the number of training samples to improve robustness.
However, this step will reduce efficiency. For Cr-SF, the convolution step size has considerable influence on the number of training samples. In order to study the robustness and computational efficiency of the proposed algorithm, 50 experiments were repeated to evaluate the robustness of the algorithm. The robustness and the average computing time of CSF and Cr-SF for 1s signals with different convolution steps are shown in Fig. 6.

Fig. 6(a) shows the robustness and optimization efficiency for simulated bearing fault signals. The two algorithms have the same computational time with the same convolution step. However, when the convolution step is equal to 12, the success rate of Cr-SF is still 100% and, CSF is only 63%. Hence, the robustness of the Cr-SF algorithm can still be guaranteed in the case of low-dimensional input. In addition, under the boundary conditions of this experiment \( (L = 100, N_f = 10, \) iteration step is 100), CSF needs to calculate 0.58s to ensure 100% success rate, while Cr-SF needs 0.49s.

Fig. 6(b) shows the robustness and optimization efficiency for simulated gear fault signals. The proposed Cr-SF showed significant improvement in robustness and optimization efficiency. Cr-SF took only 0.51 s to guarantee a 100% success rate. However, CSF can only guarantee 75% accuracy at most and needs 2.65 s for optimizing. When the convolution step is 11, the success rate of Cr-SF is 100%. However, in this case, CSF cannot extract the characteristics of gear fault.

### 4) KURTOSIS THRESHOLD

Kurtosis is a scale-invariant and threshold \( K \) is determined through experiments with different output dimensions. Fig. 7 shows the kurtosis distribution under 50 tests. Most of the \( K \) of the two methods are concentrated below 6 and increased sharply after exceeding 8. By analyzing the filters with different \( K \), the kurtosis values can be roughly divided into three regions. Firstly, when the \( K < 6 \), no obvious frequency component was in the filter, and the filter cannot extract meaningful features in this case. Secondly, when \( 6 \leq K \leq 8 \), the frequency component of the filter was not obvious and the characteristic frequency is seriously disturbed. At this time, several filters can extract meaningful features.
features under the interference of noise. When \( K > 8 \), the frequency characteristic is obvious and the filter can successfully extract fault components. Therefore, in practical applications, the kurtosis threshold range is \( 6 - 8 \). When we take 6 as the threshold, this conservative value can ensure that all filters related to fault characteristics are selected. When the kurtosis threshold is 8, we can ensure the high quality of all the selected filters. When \( K > 8 \), the kurtosis values of Cr-Sf are mainly distributed in about 41 and 21. However, the kurtosis of CSF is small and scattered. This result further proved the better robustness of Cr-SF. In this study, we take 6 as the kurtosis threshold.

B. EXPERIMENTAL ANALYSIS

We used experimental data to demonstrate the performance of Cr-SF. The experimental device, as shown in Fig. 8, mainly consists of a motor, planetary gear box, vibration sensor, and data acquisition system. In the outer ring of the bearing, a groove with a width of 0.8 mm and a depth of about 0.5 mm is cut to simulate the rolling bearing with outer-race fault, as shown in Fig. 8(c). The vibration sensor is fixed on the bearing seat of the driving end, as shown in Fig. 8(a). The driving motor ran at 1000 rpm during the data acquisition process. The sampling frequency was set as 25.6 kHz. Each vibration sample contained 6400 data points.
Although the test signal was obtained in the laboratory with noise interference, the noise was still smaller than that in the real working environment. To demonstrate the effectiveness of the proposed method under different noise environments, we added various Gauss white noise to the test signal, and the SNR was set as $-5 \text{ dB}$ and $-10 \text{ dB}$. $N_f = 20$, $\lambda = 1$, $L = 100$.

The compared results of the outer race faults are shown in Fig. 9. When $SNR = -5\text{ dB}$, Cr-SF, CSF and MED can extract the bearing fault component in the vibration signal. The obvious peak values were at the characteristic frequency and its harmonics. The recovery ability of Cr-SF in the time domain was superior to CSF and MED. When $SNR = -10\text{ dB}$, we cannot observe the fault information from the extracted components using CSF and MED. Only Cr-SF successfully separated the bearing fault components, which showed that the noise adaptability of Cr-SF was stronger than that of CSF and MED.

As a summary of the validation in this section, Cr-SF is found superior to the CSF due to better robustness and feature extraction ability in the strong noise environment.

V. COMPOUND FAULT SEPARATION USING THE PROPOSED METHOD

A. SIMULATION ANALYSIS

First, simulation signals are used to study the compound fault separation performance of CSF. The details of the parameter setting of the simulated signal can be found in Table 1. The performance of Cr-SF was compared using three fault compound modes, including the compound fault of bearings ($S_1 + S_2$), gears ($S_3 + S_4$) and bearing and gear ($S_1 + S_3$).
The threshold of kurtosis $K$ was set as $6$, $N_f = 20$, $\lambda = 1$, $L = 100$.

The Kurtogram of the compound fault in Fig. 10 shows two frequency-band signals at the fifth decomposition level indicated large kurtosis. Fig. 10(b) shows the frequency-band signals and their envelope spectrum. The fault characteristic frequency and its harmonics of two bearing faults were extracted. Comparatively, the separation of $S_2$ was effective. Fig. 11 shows the separation results using Cr-SF and CSF. Obvious impulse components were observed in the time domain of the extracted components. The characteristic frequencies and harmonics of two different faults were observed clearly in the corresponding Hilbert spectra. The experiment proved that Cr-SF and CSF can realize separation of compound bearing faults. The separation performance and filter learning ability of Cr-SF were superior to those of CSF.
FIGURE 17. Decomposition results of the experimental compound fault using Kurtogram.

FIGURE 18. Results of the separation of experimental compound fault. (a) The first extracted component; (b) Hilbert spectrum of the first extracted component, (c) the second extracted component, (d) Hilbert spectrum of the second extracted component.

Figures 12 shows the separation results of the compound fault of two gears using Kurtogram, Cr-SF and SF. The highest kurtosis frequency band was in Level 4.6. However, no-fault information was obtained from the corresponding waveform and spectrum.

The separation results using Cr-SF and CSF are displayed in Fig. 13. When \( K > 6 \), Cr-SF extracted three components, but CSF extracted only one component. Obvious periodic components in the time domain signal were extracted by Cr-SF. The corresponding Hilbert spectrum had an obvious peak value at rotation frequency. However, the component extracted by CSF had no fault information.

Fig. 14 shows the separation results of the compound fault of bearing and gear using the Kurtogram. The highest kurtosis frequency band was in Level 4.6 and the frequency was consistent with a resonance frequency \( f_r \). However, no remarkable increase in kurtosis exists in the meshing frequency was observed. Kurtogram only extracts the impulsive components. Fig. 15 indicates the separation results using Cr-SF and CSF. The first two components extracted by Cr-SF had impulsive and harmonic components in the time domain signal. The corresponding Hilbert spectra exhibited gear and bearing faults in the signal. In this case, CSF could only extract bearing faults, but not identify gear faults.

As a summary, to compare the performance of the two methods, \( SNR \) in this section was set as \(-10\,\text{dB}\). When the value of \( SNR \) was high, CSF identified gear faults with poor robustness. Therefore, from the above experimental results, Cr-SF was an effective method for weak compound fault separation performed better noise adaptability than CSF, and can be applied to different fault combination modes.

B. EXPERIMENTAL ANALYSIS

We simulated the compound fault of the bearing outer race and gear wear on the test bench to further verify the performance of Cr-SF. The planetary gear set contains mainly three components: a sun gear, three planet gears, and a ring gear. The tooth numbers of the sun and planet gears are 38 and 18, respectively. The tooth surface of the planet gear is ground to simulate the wear gear fault as shown in Fig. 8(d). The driving motor worked on 1000rpm in the testing process. As shown in Fig. 16, the sampling frequency was 25.6 kHz. Approximately 6400 data points were in each input sample. \( N_f = 30, \lambda = 1, L = 100 \).

The diagnostic results of the experimental compound faults using Kurtogram are shown Fig. 17. The band with the highest kurtosis was located at Level 5. The envelope spectrum obtained from the extracted signal components is plotted in Fig. 17(b). Three peaks could lead to misdiagnosis. The gear and bearing fault were not separated.

Fig. 18 shows the separation results using Cr-SF and CSF. The impulse component and AM component can be judged...
clearly from the fault component separated from Cr-SF in the time domain. From the Hilbert spectra, component 1 had an obvious peak value at the rotating frequency and component 2 had obvious bearing fault characteristics. The extracted components of Cr-SF recovered better than that of CSF whether in the time or frequency domain. The comparison of the filters trained by the two methods in Fig. 19 shows that the filters learned by Cr-SF are obviously superior to CSF.

VI. CONCLUSION
Sparse optimization is an effective method in weak signature detection. We summarized the weaknesses of the traditional methods and introduced the desired properties of WSD and WCFS, including strong noise adaptability and robustness, fast optimization process, no requirement for prior experience and knowledge, and applicable to various fault compound modes. A multidimensional BD method based on Cr-SF was proposed for WSD and WCFS to achieve the desired properties. The CKP method was recommended to determine the final filters. The main conclusions are summarized as follows:

First, the proposed method exhibited stronger noise adaptability that can identify weak fault information under a strong noise environment. Second, the robustness of Cr-SF is improved obviously with the same parameter setting and the computational efficiency was higher with the same success rate. Third, the experimental results confirmed that Cr-SF is a superior tool that can handle different fault complicating modes under strong noise environment without any preprocessing and priori knowledge.

Future research will focus on the following topics. First, Cr-SF can be combined with the transfer learning to realize the feature transform from laboratory to real world industry. Second, this method can be further modified further to enhance noise adaptability and robustness.

REFERENCES
[1] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, “A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines,” Mech. Syst. Signal Process., vol. 116, pp. 668–692, Feb. 2019.
[2] Z. Zhang, S. Li, J. Lu, J. Wang, and X. Jiang, “A novel intelligent fault diagnosis method based on fast intrinsic component filtering and pseudo-normalization,” Mech. Syst. Signal Process., vol. 145, Nov. 2020, Art. no. 106923.
[3] M. Kedadouche, M. Thomas, and A. Tahar, “A comparative study between empirical wavelet transforms and empirical mode decomposition methods: Application to bearing defect diagnosis,” Mech. Syst. Signal Process., vol. 81, pp. 88–107, Dec. 2016.
[4] Y. Gao, M. Karimi, A. A. Kudreyko, and W. Song, “Sparse optimized based on improved ADMM and the minimum entropy de-convolution for the early weak fault diagnosis of bearings in marine systems,” ISA Trans., vol. 78, pp. 98–104, Jul. 2018.
[5] B. Pang, G. Tang, and T. Tian, “Enhanced singular spectrum decomposition and its application to rolling bearing fault diagnosis,” IEEE Access, vol. 7, pp. 87769–87782, 2019.
[6] B. Pang, G. Tang, and T. Tian, “Rolling bearing fault diagnosis based on SVD-based kurtagram and iterative autocorrelation of teaser energy operator,” IEEE Access, vol. 7, pp. 77222–77237, 2019.
[7] Z. Liu and L. Zhang, “Naturally damaged wind turbine blade bearing fault detection using novel iterative nonlinear filter and morphological analysis,” IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8713–8722, Oct. 2020.
[8] J. Antoni, “The spectral kurtosis: A useful tool for characterising non-stationary signals,” Mech. Syst. Signal Process., vol. 20, no. 2, pp. 282–307, Feb. 2006.
[9] F. Combet and L. Gelman, “Optimal filtering of gear signals for early damage detection based on the spectral kurtosis,” Mech. Syst. Signal Process., vol. 23, no. 3, pp. 652–668, Apr. 2009.
[10] T. Barszcz and R. B. Randall, “Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine,” Mech. Syst. Signal Process., vol. 23, no. 4, pp. 1352–1365, May 2009.
[11] T. Barszcz and A. Jabłoński, “A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtagram,” Mech. Syst. Signal Process., vol. 25, no. 1, pp. 431–451, Jan. 2011.
[12] Y. Wang, J. Xiang, R. Markert, and M. Liang, “Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications,” Mech. Syst. Signal Process., vol. 66–67, pp. 679–698, Jan. 2016.
[13] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zhi, J. Yuan, B. Chen, and Z. He, “Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review,” Mech. Syst. Signal Process., vol. 70–71, pp. 1–35, Mar. 2016.
[14] X. Zhang, Z. Liu, J. Wang, and J. Wang, “Time–frequency analysis for bearing fault diagnosis using multiple Q-factor Gabor wavelets,” ISA Trans., vol. 87, pp. 225–234, Apr. 2019.
[15] V. Purushotham, S. Narayanan, and S. A. N. Prasad, “Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition,” NDT E Int., vol. 38, no. 8, pp. 654–664, Dec. 2005.
[16] Z. Zhang, S. Li, J. Wang, Y. Xin, Z. An, and X. Jiang, “Enhanced sparse filtering with strong noise adaptability and its application on rotating machinery fault diagnosis,” Neurocomputing, vol. 398, pp. 31–44, Jul. 2020.
[17] M. Pan and W. C. Tsao, “Using appropriate IMFs for envelope analysis in multiple fault diagnosis of ball bearings,” Int. J. Mech. Sci., vol. 69, pp. 24–114, Apr. 2013.
[18] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.
[19] Y. Wang, R. Markert, and J. Xiang, “Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system,” Mech. Syst. Signal Process., vol. 60, pp. 243–251, Aug. 2015.
[20] Y. Wang and R. Markert, “Filter bank property of variational mode decomposition and its applications,” Signal Process., vol. 120, pp. 509–521, Mar. 2016.

[21] X. Jiang, J. Wang, and J. Shi, “A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines,” Mech. Syst. Signal Process., vol. 116, pp. 668–692, Feb. 2019.

[22] Y. Miao, M. Zhao, and J. Lin, “Identifications of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition,” ISA Trans., vol. 84, pp. 82–95, Jan. 2019.

[23] R. A. Wiggins, “Minimum entropy deconvolution,” Geoexploration, vol. 16, nos. 1–2, pp. 21–35, Apr. 1978.

[24] C. A. Cabelli, “Minimum entropy deconvolution and simplicity: A non-iterative algorithm,” Geophysics, vol. 50, no. 3, pp. 394–413, Mar. 1985.

[25] N. Sawalhi, R. B. Randall, and H. Endo, “The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis,” Mech. Syst. Signal Proc., vol. 21, no. 6, pp. 2616–2633, Aug. 2007.

[26] H. Endo and R. B. Randall, “Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter,” Mech. Syst. Signal Process., vol. 21, no. 2, pp. 906–919, Feb. 2007.

[27] J. Obuchowski, A. Wylomanska, and R. Zimroz, “New criteria for adaptive blind deconvolution of vibration signals from planetary gearbox,” in Advances in Condition Monitoring of Machinery in Non-Stationary Operations, Dec. 2016, pp. 111–125.

[28] G. McDonald, Q. Zhao, and M. Zuo, “Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection,” Mech. Syst. Signal Proc., vol. 33, pp. 237–255, Nov. 2012.

[29] G. Tang, X. Wang, and Y. He, “Diagnosis of compound faults of rolling bearings through adaptive maximum correlated kurtosis deconvolution,” J. Mech. Sci. Technol., vol. 30, no. 1, pp. 43–54, Jan. 2016.

[30] Y. Miao, M. Zhao, J. Lin, and Y. Lei, “Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings,” Mech. Syst. Signal Proc., vol. 92, pp. 173–195, Aug. 2017.

[31] D. Yu, M. Wang, and X. Cheng, “A method for the compound fault diagnosis of gearboxes based on morphological component analysis,” Measurement, vol. 91, pp. 519–531, Sep. 2016.

[32] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, May 2015.

[33] Y. Lei, F. Jia, J. Lin, and S. Xing, “An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data,” IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3137–3147, Jan. 2016.

[34] S. S. Udmale and S. K. Singh, “Application of spectral kurtosis and improved extreme learning machine for bearing fault classification,” IEEE Trans. Instrum. Meas., vol. 68, no. 11, pp. 4222–4233, Nov. 2019.

[35] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, “Machinery health prognostics: A systematic review from data acquisition to RUL prediction,” Mech. Syst. Signal Proc., vol. 104, pp. 799–834, May 2018.

[36] J. Ngiam, Z. Chen, and S. A. Bhaskar, “Sparse filtering,” in Proc. Adv. Neural Inf. Process. Syst., Dec. 2011, pp. 1125–1233.

[37] Z. Zhang, S. Li, J. Wang, “General normalized sparse filtering: A novel unsupervised learning method for rotating machinery fault diagnosis,” Mech. Syst. Signal Proc., vol. 124, pp. 596–612, Jun. 2018.

[38] X. Jia, M. Zhao, Y. Di, C. Jin, and J. Lee, “Investigation on the kurtosis filter and the derivation of convolutional sparse filter for impulsive signature enhancement,” J. Sound Vibrat., vol. 386, pp. 433–448, Jan. 2017.

[39] R. Peled, S. Braun, and M. Zacksenhouse, “A blind deconvolution separation of multiple sources, with application to bearing diagnostics,” Mech. Syst. Signal Proc., vol. 19, pp. 1181–1195, Nov. 2005.

[40] X. Jia, M. Zhao, Y. Di, P. Li, and J. Lee, “Sparse filtering with the generalized $l_p$/$l_q$ norm and its applications to the condition monitoring of rotating machinery,” Mech. Syst. Signal Proc., vol. 102, pp. 198–213, Mar. 2018.

SHAN WANG received the B.S. and M.S. degrees from the Nanjing University of Aeronautics and Astronautics, China, in 2020. He is currently a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interests include intelligent fault diagnosis of rotating machinery, noise and vibration analysis and control, intelligent fault diagnosis, signal processing, and machine learning.

ZONGZHEN ZHANG received his Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2019. He is currently a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interest includes intelligent fault diagnosis of machines.

JINRUI WANG received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2019. He is currently a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interest includes intelligent fault diagnosis of machines.

BAOKUN HAN received the Ph.D. degree from the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, in 2003. He is currently a Professor with the Shandong University of Science and Technology (SDUST). His current research interests include vehicle dynamics, noise and vibration analysis and control, signal processing, and fault diagnosis.

* * *

VOLUME 8, 2020