Article

Weak Signal Detection Based on Combination of Sparse Representation and Singular Value Decomposition

Huijie Ma, Shunming Li *, Xianglian Li and Jiantao Lu

Abstract: Due to the inevitable acquisition system noise and strong background noise, it is often difficult to detect the features of weak signals. To solve this problem, sparse representation can effectively extract useful information according to the sparse characteristics of signals. However, sparse representation is less effective against non-Gaussian white noise. Therefore, a novel SRSVD method combining sparse representation and singular value decomposition is proposed to further improve the denoising performance of the algorithm. All the signal components highly matched with the dictionary are extracted by sparse representation, and then each component of singular value decomposition is weighted by the evaluation index, PMI, which can indicate the component with useful information in the signal, so that the signal denoising performance of the algorithm is greatly improved. The performance of the proposed method is verified by the processing of weak signals in the circuit and early fault signals of bearings. The results show that SRSVD can successfully suppress noise interference. Compared with other existing methods, SRSVD has better denoising performance.

Keywords: singular value decomposition; sparse representation; weak signal detection

1. Introduction

Extracting useful information from noise has always been a problem of extensive research in various fields. For example, in the field of image, image denoising and super-resolution analysis can enrich the information of the images [1]; in the field of condition monitoring, fault signal extraction can identify the existence of early fault [2,3]; in the field of electronics, the extraction of small changes in current can identify kinds of weak measurement changes [4]. Therefore, it is crucial and necessary to study effective signal denoising methods.

The evaluation criteria of a “weak signal” are not fixed, and there is no clear and widely agreed definition and interpretation of the weak signal. As weak signal detection is widely used in many fields, the detection objects in each field are completely different, and the detection methods vary greatly. Moreover, the degree of signal weakness in different detection objects is inconsistent, so it is difficult to define “weak”. Therefore, in order to distinguish it from weak signals in other fields, it should be emphasized that what is studied in this paper is weak sensor electrical signals, and weakness means that the energy of electrical signals is relatively weak compared with other electrical noises. Alternatively, weak signals can be interpreted from the perspective of signal detection; that is, weak signals refer to signals that are difficult to detect with conventional and traditional signal analysis methods.

The commonly used signal processing methods include wavelet transform (WT), short-time Fourier transform (STFT) [5], empirical mode decomposition (EMD) [6], variational mode decomposition (VMD) [7], fast kurtogram [8], and singular value decomposition (SVD) [9], etc. The sparse feature is a universal property in the signal [10]. In recent years, scholars have carried out extensive discussions and studied various methods by using the
In order to extract the sparse features of signals, the most representative method is the sparse representation (SR) algorithm [11]. The core idea of SR is that any signal can be represented as the superposition of some optimal basic functions from the basic waveform library, which matches the main structure of the signal best. The functions of the basic waveform library are called dictionaries, and dictionaries can be constructed based on prior knowledge of the signal, or based on paradigms of nonparametric signals. Since the dictionary matches some major features of the original signal, the coefficients of the signal are sparse under the decomposition of the dictionary. The theoretical system of sparse representation algorithm was first established on the matching tracking algorithm proposed by Mallat in 1993.

In 1998, Donoho proposed the sparse decomposition theory. In essence, it reasonably loosed the regularization term of $l_0$ in the mathematical model of matching tracking algorithm to $l_1$ regularization, which also marks the formal establishment of the two core algorithms of sparse representation, and the subsequent sparse representation algorithms are improved or extended on this basis. Greedy tracking algorithms based on $l_0$ regularization mainly include orthogonal matching pursuit (OMP) [12] in 1993, tree based tracking algorithm proposed by Jost et al. [13] in 2006, and phased orthogonal matching pursuit (StOMP) proposed by Donoho [14] in 2012. Algorithms based on $l_1$ or $l_p$ regularization mainly include the least absolute shrinkage selection operator (LASSO) [15] in 1996, focal underdetermined system solver (FOCUSS) [16] in 1997 and gradient projection-based sparse representation (GPSR) [17] in 2008. In the same year, Huggins [18] proposed greedy basis tracking, and Bobin [19] proposed morphological component analysis (MCA), also belonging to this category. In 2004, Daubechies et al. [20] proposed a new sparse atomic decomposition algorithm, called iterative shrinkage threshold algorithm (ISTA), and it gradually became one of the most widely used convex optimization algorithms.

The singular value decomposition method is a widely used signal denoising method, but it has encountered many problems in practical application [9]. For example, the cut-off rule in the difference spectrum is difficult to determine, it is easily affected by noise energy and it does not have the ability to identify useful signals. Therefore, a novel signal denoising method called SRSVD is proposed to solve this problem. Through the powerful ability of SR to weaken Gaussian noise, we filter once, find out all components that best match the dictionary, and reconstruct the signal to be measured. Then the improved SVD algorithm is used to remove the irrelevant components to achieve the purpose of secondary denoising. Previously, most methods combining SR and SVD were used for pattern recognition. SVD was generally used for dimensionality reduction, and then SR is used for classification [21,22]. In our method, SR is first used to remove Gaussian noise, which requires the reconstruction of the original signal. In the second step, improved SVD is used for secondary filtering. As can be seen, the purpose, sequence, and flow of our method is fundamentally different from the previous methods.

The major contributions of this paper are as follows:

1. Combining SR and SVD for weak signal detection for the first time and achieved good results.
2. According to the characteristics of SR and SVD and combined with practical application, different dictionaries are adopted to achieve SRSVD denoising.
3. The PMI index is used for the reconstruction of the signal in SRSVD, and the SRSVD can filter the useful information in the signal.

This paper is organized as follows. In Section 2, SR and SVD are briefly described. Section 3 details the proposed method SRSVD. In Section 4, the SRSVD method will be analyzed using experimental test data. Finally, conclusions are drawn in Section 5.

2. Background

2.1. Sparse Representation

For a signal, if the coefficient vector decomposed by a suitable dictionary meets the following properties: if all the other elements are zero except that the value of a few
elements is non-zero, the signal can be sparse represented, and the coefficient vector obtained by decomposing is called the sparse representation vector of the original signal. Sparse representation is the most representative method in linear representation. It has also been proved to be a very powerful solution. It shows great potential in signal and image denoising and has been widely used in the fields of machine learning [23] and computer vision [24].

Assuming a given measurement signal \( y \in \mathbb{R}^M \), its sparse representation model [2] is defined as:

\[
y = Dx + \varepsilon = \sum_{k=1}^{K} d_k x_k + \varepsilon
\]  
(1)

where \( D \in \mathbb{R}^{M \times K} \) is the dictionary matrix, \( M \) is the number of rows of dictionary \( D \), \( K \) is the number of columns of dictionary \( D \); \( d_k \) is a column vector in a dictionary that can be called atom or basis function. \( x = [x_1, x_2, \ldots, x_k]^T \) is the sparse representation vector of input signal \( y \), which can be regarded as the weight of base function \( d_k \), \( \varepsilon \) is noise. In the sparse representation model, the number of columns in dictionary \( D \) is usually much greater than the number of rows (\( K \gg M \)), that is, the dictionary is over-complete or redundant, which leads to the fact that \( x \) can have infinite solutions. Adding sparse constraints to the model can solve the degradation problem caused by over-completeness. Therefore, the objective function of sparse representation can be expressed as

\[
\text{argmin}_x \frac{1}{2} \|y - Dx\|_2 + \lambda \|x\|_0
\]  
(2)

where \( \| \cdot \|_0 \) is the pseudo-norm of \( l_0 \) and the return value is the number of non-zero values of the vector. It is generally considered that \( \|y - Dx\|_2 \) is the data fidelity term, which ensures that the signal reconstructed by the product of the dictionary and coefficient vector is as consistent as possible with the original signal. \( \lambda \|x\|_0 \) is a regular term, and its design is generally designed with prior knowledge to constrain parameters in the model.

In the sparse representation model, it is a sparse constraint on the decomposed coefficient vector. The methods of solving Equation (2) are mainly divided into two categories: one is a greedy algorithm, such as match pursuit (MP), orthogonal match pursuit (OMP), and piecewise orthogonal match pursuit (SIOMP). Since the \( l_0 \) norm is non-convex, the solution strategy of this kind of algorithm is mainly to use an exhaustive search to solve the sparse representation problem, which is almost an NP-hard problem. With the increase of dimension, the complexity of the solution increases rapidly, resulting in the need to spend a lot of computational cost in the actual situation.

Another strategy is to smooth the \( l_0 \) norm into an \( l_1 \) norm convex optimization problem. Candes and Tao proved that under RIP conditions, the solution of Equation (2) in the \( l_0 \) norm is consistent with that in the \( l_1 \) norm, and the problem is easily solved by a gradient correlation algorithm. \( l_1 \) norm regularization problem is also called the Lasso regression problem and its objective function is:

\[
\text{argmin}_x \frac{1}{2} \|y - Dx\|_2 + \lambda \|x\|_1
\]  
(3)

where \( \|x\|_1 = \sum |x_n| \) is the \( l_1 \) norm. Typical methods for solving Equation (3) include an iterative shrinkage-thresholding algorithm (ISTA), gradient projection for sparse reconstruction (GPSR), and the interpolation point method, etc. Note that compared with Equation (2), Equation (3) only replaces the \( l_0 \) norm in the regularization term with the \( l_1 \) norm. Convert it to the following form:

\[
\min_x \frac{1}{2} \|y - Dx\|_2, \text{s.t.} \|x\|_1 \leq \lambda
\]  
(4)

It is assumed that \( x \) is two-dimensional to facilitate visual analysis, as shown in Figure 1a. Contour lines of Equation (4) can be drawn on the \((x_1, x_2)\) plane, and the constraint is a region with a size of \( \lambda \) on the plane (within the square contour). The place
where contour lines intersect with this region for the first time is the optimal solution. It can be seen that the region constrained by the $l_1$ norm has “angles” at the intersection of each coordinate axis, while the geodesic of the objective function will intersect at angles and generate sparsity in most cases unless the position is very well placed. In contrast, the $l_2$ norm does not have such properties, as shown in Figure 1b. Because it is a circle, the probability of the first intersection occurring at the position with sparsity is very small. Therefore, converting the $l_0$ norm to the $l_1$ norm can also produce sparsity, while the $l_2$ norm can produce smoothness but not sparsity.

Figure 1. Visualization of objective function under different norm constraints.

2.2. Singular Value Decomposition

For a discrete sampled signal $X \in \mathbb{R}^N$, its Hankel matrix is constructed:

$$ H = \begin{bmatrix} x(1) & x(2) & \cdots & x(n) \\ x(2) & x(3) & \cdots & x(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-n+1) & x(N-n+2) & \cdots & x(N) \end{bmatrix} \quad (5) $$

Let $m = N - n + 1$, in order to obtain the best denoising effect, when the signal length is even, we should take the number of columns $n = N/2$ and the number of rows $m = N/2 + 1$. While $N$ is odd, the number of columns $n = (N + 1)/2$ and the number of rows $m = (N + 1)/2$ should be selected.

The singular value decomposition of $H$ is:

$$ H = USV^T \quad (6) $$

The diagonal elements of $S$ are called singular values of $H$, denoted as $\sigma_1, \sigma_2, \ldots, \sigma_n$. These singular values are shown as the former $k$ values are large, and the later singular values are very close to 0. The difference spectrum can describe the change of the singular value sequence. As long as the superposition of the former $k$ components is selected and the singular value after $k$ is set to zero, the noise reduction signal can be obtained.

3. Proposed SRSVD Method

The basic flowchart of the proposed method is shown in Figure 2, and the steps are explained in detail below.

SR from the perspective of probability interpretation, the fidelity term of the SR model is given by the maximum likelihood estimation of $y|x$ to be inferred likelihood estimator, and the fidelity term follows white Gaussian distribution. Therefore, the SR model has a strong ability to remove white Gaussian noise. Because a lot of noise in nature is caused by the irregular movement of some particles. Especially in the circuit, the irregular motion of electrons, such as thermal motion, produces strong white Gaussian noise. Therefore, the first purpose of using SR is to remove Gaussian white noise.
Next, the reconstructed signal is decomposed by SVD, and the decomposed signal components are given proper weights by appropriate evaluation indexes to extract the favorable parts of the signal components. The specific evaluation index and weight selection will be given in Section 4 combined with the actual situation.

Therefore, the SRSVD noise reduction algorithm steps can be summarized as follows:

Step 1. Construct a dictionary matrix $D$ by prior knowledge or dictionary learning. For dictionary learning, its objective function is:

$$\text{argmin}_{D,A} \frac{1}{2} \| Y - DA \|_F^2, \quad \text{s.t.} \quad \|A_i\|_0 \leq T, \quad i = 1, 2, \ldots, m$$

(7)

where $Y = (y_j)_{j=1}^m \in \mathbb{R}^{m \times n}$ is the sample matrix; it may consist of signals of nonparametric signal observation. $A = (a_i)_{i=1}^m$ is the sparse coefficient matrix and $T$ is sparsity. $\|X\|_F = \sqrt{\sum \sum |x_{ij}|^2}$ is the Frobenius norm of the matrix $X$. Dictionary learning algorithm can construct an appropriate dictionary without prior knowledge of the sample, so it has a wider range of application fields than an analytic dictionary.

Step 2. The sparse coefficients are solved by ISTA [20]. The objective function is shown in Equation (3) and the ISTA procedure is shown in Algorithm 1.

Step 3. Reconstructed signal $y = D x^{(i+1)}$.

Step 4. Construct Hankel matrix $H$ of new $y$, the construction method is shown in Equation (5).

Step 5. Decompose $H$ as shown in Equation (6). The decompose components are denoted as $f_1, f_2, \ldots, f_n$.

Step 6. Each component is given weight according to the evaluation index. In this paper, we choose the periodic modulation intensity (PMI) [25] as the component evaluation index, which can be used as an important indicator to measure the information in the signal. The weight assigned to each component by PMI is denoted as $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_n$, and finally, after denoising $\tilde{y} = \sum_{i} \bar{w}_i f_1$. 

---

**Figure 2. Flowchart of SRSVD.**

SR

- Input signal
- Dictionary construction
- Solving the sparse coefficient
- Signal reconstruction

SVD

- Hankel matrix construction
- Singular value decomposition
- Reconstruction based on the weighted sum according to evaluation indexes
- Output signal

**Flowchart of SRSVD.**

SR from the perspective of probability interpretation, the fidelity term of the SR
Algorithm 1. Iterative shrinkage-thresholding algorithm.

| Initialization: $x^{(0)}, \mu < \|D\|_2^2$, $\text{Iter}$; |
| --- |
| For $i = 0$ to $\text{Iter}$ do |
| $z^{(i)} = x^{(i)} - \mu D^T (Dx^{(i)} - y)$; |
| $x^{(i+1)} = \text{soft}(z^{(i)}; \mu \lambda)$; |
| end for |
| return $x^{(i+1)}$ |

4. Validation of SRSVD
4.1. Case Study I

In this section, the feasibility of the SRSVD algorithm is tested by building a signal acquisition system. The test system is shown in Figure 3, and consists of five parts. The signal generator generates useful signals for testing; the amplifier analog circuit board attenuates the useful signal output of the signal generator with large capacitance and outputs the amplification result of the amplifier chip; the signal acquisition instrument (DN8509N) is used to collect the output signal of the amplifier chip; and the shielding box blocks external interference and the PC is used to record the chip output signal.

![Figure 3. Experiment platform system and analog amplifier circuit board. (a) The test system and (b) the analog amplifier circuit.](image)

We set the sampling frequency to 128 kHz and the signal frequency to be measured to 123 Hz. We randomly intercept a segment of signal with 6400 sampling points for analysis. Figure 4 shows the signal waveform and its frequency spectrum. As can be seen from Figure 4b, the signal generated by the signal generator cannot be effectively identified due to the adverse effects of noise in the circuit.

![Figure 4. (a) The original signal waveform, (b) its frequency spectrum, and (c) partial enlargement.](image)
We used SRSVD to denoise this signal and try to extract the signal from the signal generator. First, we used the dictionary learning method to train a dictionary. It should be noted here that although the data in this experiment are generated by the signal generator, we often do not know the types and features of the collected signals in practical situations. Therefore, the prior of the signal generator was ignored in this experiment. K-SVD algorithm was used for training, and the algorithm procedure is shown in Algorithm 2. K-SVD algorithm is a dictionary learning algorithm proposed by Aharon et al. [1] in 2006 and has gradually become the most widely used dictionary learning algorithm. K-SVD updates the dictionary columns one by one, reducing the computational complexity of updating the dictionary. The initial dictionary is set to a Gaussian random dictionary,

\[
\text{Initialization:} \\
\text{Construct a Gaussian random dictionary, } D(0) \in \mathbb{R}^{n \times m}, \text{using ISTA algorithm to solve Equation (3) and obtain the sparse coefficient } \alpha_j. \\
\text{Sparse representation stage:} \\
\text{Iterate:} \\
\text{For } j = 1, 2, \ldots, m: \\
\text{For } k = 1, 2, \ldots, n: \\
\text{Update the residual:} \\
\text{Compute:} \\
\text{Break rule: If the change of } R_j \text{ is small enough, the iteration is terminated.} \\
\text{Return: } D(k) 
\]

K-SVD updates the dictionary columns one by one, reducing the computational complexity of updating the dictionary. The initial dictionary is set to a Gaussian random dictionary, \( n = 6400, m = 10, \) and the number of samples is set to 50. Some dictionary atoms obtained by K-SVD are shown in Figure 5. It can be seen that the learned atom has a certain harmonic performance, which has a certain degree of matching with the signal generated by the signal generator in the internal structure. The original signal is denoised by SR through the learned dictionary, and the result is shown in Figure 6.

![Learned dictionary atoms](image1)

**Figure 5.** Learned dictionary atoms.

![Denoised signal waveform and frequency spectrum](image2)

**Figure 6.** (a) The denoised signal waveform by SR and (b) its frequency spectrum.

SVD is used to perform secondary filtering on the denoised signal. After being filtrated by PMI, the final result is shown in Figure 7. It can be seen that by SRSVD denoising, the dynamic range of the frequency of the irrelevant components is smaller. Moreover, the frequency of 123 Hz is more obvious.
As shown in Figure 8, the test bench [26] consists of an AC induction motor, a motor value, and a hydraulic loading system. Two accelerometers are mounted vertically and horizontally on the housing of the test bearing. Set the sampling speed controller, a support shaft, and a hydraulic loading system. Two accelerometers are often be used as prior knowledge in the process of fault signal extraction. The fault signals can be estimated to be about 109 Hz through the fault bearing physical mode.

It should be noted that different from case study I, the bearing fault signal model can be expressed as the sum of a series of impulses:

\[ x(t) = \sum_{i=1}^{M} A_i s(t - iT - \tau_i) + n(t) \]  

where \( s(t) = A \sin(\omega_1 t) e^{-\xi_1 t} \), \( A_i \) is the magnitude of the \( i \)th impulse, \( T \) is the period of the impulse, \( \tau_i \) is the rolling mass slip time of the \( i \)th impulse, which is usually a random small value, \( M \) is the number of impulses, and \( n(t) \) is random noise. Therefore, according to this model, we adopt an analytic dictionary composed of tunable Q-factor wavelet transform...
This work makes a lot of sense because if early fault characteristics can be detected during the operation of rotating equipment, the loss of economy and life can be greatly reduced. We use SRSVD to analyze the early stage signal and try to extract the fault characteristics. This work makes a lot of sense because if early fault characteristics can be detected during the operation of rotating equipment, the loss of economy and life can be greatly reduced.

The severity of bearing fault can be roughly judged by the value of the root mean square (RMS), and the development stage of the fault can be divided according to this value, as shown in Figure 9. Figure 9c and 9d respectively show the signal waveforms in the early and medium stages of the fault. It can be seen that the dynamic range of the signal in the medium stage is larger and presents obvious periodic pulse characteristics, while it is difficult to find fault characteristics in the early stage of the signal waveform. We use SRSVD to analyze the early stage signal and try to extract the fault characteristics. This work makes a lot of sense because if early fault characteristics can be detected during the operation of rotating equipment, the loss of economy and life can be greatly reduced.

We chose the 60th minute signal for analysis, with data points of 32,768. The parameters of TQWT are as follows: $Q = 2$, $r = 5$, $J = 10$. In the process of SR, the ISTA algorithm is also used to solve the sparse coefficients, and the evaluation index for SVD is also PMI. The denoised signal obtained by SRSVD is shown in Figure 10g, and the fault signal is...
demodulated through Hilbert envelope spectrum analysis, as illustrated in Figure 10h. As can be seen from Figure 10h, the fault characteristic frequency (109 Hz) and its 2nd, 3rd, and 4th harmonics can be clearly found.

Figure 10. Denoising results of four different methods and their envelope spectra. From top to bottom: L1-Kurtosis (a,b), RSVD (c,d), VMD (e,f), and SRSVD (g,h).

We used three other widely used denoise algorithms to compare with SRSVD, these algorithms were L1-Kurtosis [27], RSVD [25], and variational mode decomposition (VMD) [7]. L1-kurtosis is a kurtosis-based weighted sparse representation denoising model. RSVD algorithm is the secondary filtering algorithm used in SRSVD, i.e., the PMI evaluation index is embedded in SVD so that the restructuring process has the ability to select components according to the information content. VMD is an adaptive, fully non-recursive method for modal variation and signal processing.

In this paper, the number of decomposition levels was selected as six, and the first two components with the largest kurtosis were selected for reconstruction. The denoising results of these three methods are shown in Figure 10a–f. It can be seen that there was no obvious fault feature in their envelope spectra. Only in the VMD method denoised signal envelope spectra, fault frequency (109 Hz) and its third, fifth harmonic (327 Hz, 545 Hz) could be found obviously, but the noise reduction effect was still not as obvious as that of SRSVD. In addition, kurtosis is used to evaluate the signal after noise reduction and a larger kurtosis indicates more fault information. Kurtosis is an important index to evaluate the number of fault information in signal and is widely used in fault diagnosis. Its calculation formula is:

\[
K = \frac{1}{n\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^4 \left( \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \right)^2
\]

where \( n \) is the number of data points, and \( \mu \) is the mean value. After calculation, the kurtosis of the original signal was 4.673, and the kurtosis of the signal by SRSVD, L1-kurtosis, RSVD, and VMD denoised were 5.486, 4.759, 5.146, and 5.464, respectively. Therefore, after SRSVD noise reduction, the fault features in the signal were the most obvious, and the fault information was the most abundant.
In addition, the same analysis and comparison experiments were performed for the signals at the 50th minute and 100th minute. These two signals can be regarded as the stronger noisy case and the less noise case, respectively. Firstly, the 100th minute signal was processed. The envelope spectra of the signals after denoising by the four methods are shown in Figure 11. It can be seen that when in the middle fault stage, all the four methods can find the fault frequency and its high harmonic frequency, among which VMD and SRSVD were particularly excellent. Their corresponding kurtosis values were $K_{\text{L1-Kurtosis}} = 5.663$, $K_{\text{RSVD}} = 5.176$, $K_{\text{VMD}} = 5.862$, and $K_{\text{SRSVD}} = 5.937$, respectively. Secondly, the signal at the 50th minute was tested, and the obtained envelope spectra comparison results are shown in Figure 12. It can be seen that at the 50th minute, the performance of the four methods was not ideal; only SRSVD could find the third harmonic frequency of the fault. At this time, their corresponding kurtosis values were $K_{\text{L1-Kurtosis}} = 3.765$, $K_{\text{RSVD}} = 3.567$, $K_{\text{VMD}} = 3.649$, and $K_{\text{SRSVD}} = 3.956$, respectively.

![Figure 11. Denoising the 100th minute signals envelope spectra of 4 different methods.](image1)

![Figure 12. Denoising the 50th minute signals envelope spectra of 4 different methods.](image2)

5. Conclusions

This paper presents a signal processing method for weak signal detection. By combining sparse representation with singular value decomposition, Gaussian white noise is removed by sparse representation for primary filtering, and useful information in the signal is extracted by weighted singular value decomposition for secondary filtering to achieve the purpose of denoising. This multi-layer filtering method is mainly taken into account that sparse representation is sensitive to Gaussian noise, while the improved SVD method can select useful components in signals by using the appropriate indicator. After two layers of filtering, the useful components in the signal can be extracted. Two experiments verify the effectiveness of the proposed method and the experimental results show that SRSVD...
has advantages in fault signal extraction compared with the L1-kurtosis method, RSVD, and VMD. Future research can focus on useful information indexes to extract weak signals more effectively.

**Author Contributions:** Conceptualization, H.M.; methodology, H.M.; software, H.M.; validation, H.M.; formal analysis, H.M. and S.L.; investigation, H.M.; resources, S.L.; data curation, X.L.; writing—original draft preparation, H.M.; writing—review and editing, S.L.; visualization, H.M.; supervision, S.L.; project administration, S.L.; funding acquisition, S.L. and J.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Key Research and Development Program of China (No. 2018YFB2003300), National Science and Technology Major Project, China (No. 2017-IV-0008-0045), and the National Natural Science Foundation of China (No. 51975276).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets in Section 4.2 can be found on this website: https://drive.google.com/open?id=1_ycmG46PARIykt82ShfnFyQsaXv3_VK (accessed on 18 April 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Elad, M.; Aharon, M. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Trans. Image Process.* 2006, 15, 3736–3745. [CrossRef]
2. Ma, H.; Li, S.; Lu, J.; Zhang, Z.; Gong, S. Structured sparsity assisted online convolution sparse coding and its application on weak signature detection. *Chin. J. Aeronaut.* 2021, 35, 266–276. [CrossRef]
3. Zhao, X.; Ye, B. Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock. *Mech. Syst. Sig. Process.* 2011, 25, 1617–1631. [CrossRef]
4. Dong, Y.; Xu, S. A New Directional Weighted Median Filter for Removal of Random-Valued Impulse Noise. *IEEE Signal Process Lett.* 2007, 14, 193–196. [CrossRef]
5. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* 2006, 52, 1289–1306. [CrossRef]
6. Delechelle, E.; Lemoine, J.; Niang, O. Empirical mode decomposition: An analytical approach for sifting process. *IEEE Signal Process Lett.* 2005, 12, 764–767. [CrossRef]
7. Zosso, D.; Dragomiretskiy, K. Variational Mode Decomposition. *IEEE Trans. Signal Process.* 2013, 62, 531–544.
8. Wang, Y.; Liang, M. An adaptive SK technique and its application for fault detection of rolling element bearings. *Mech. Syst. Signal Process.* 2011, 25, 1750–1764. [CrossRef]
9. Ma, Z.; Zhang, J.; Zhang, A.; Ruan, W. Fault feature extraction of rolling bearings based on VMD-SVD joint de-noising and FSWT. *Zhendong Yu Changji/J. Vib. Shock.* 2018, 37, 210–217.
10. Feng, Z.; Zhou, Y.; Zuo, M.J.; Chu, F.; Chen, X. Atomic decomposition and sparse representation for complex signal analysis in machinery fault diagnosis: A review with examples. *Measurement* 2017, 103, 106–132. [CrossRef]
11. Sun, R.B.; Yang, Z.B.; Zhai, Z.; Chen, X.F. Sparse representation based on parametric impulsive dictionary design for bearing fault diagnosis. *Mech. Syst. Signal Process.* 2019, 122, 737–753. [CrossRef]
12. Mallat, S.; Zhang, Z. Matching pursuit with time-frequency dictionary. *IEEE Trans. Signal Process.* 1993, 41, 3397–3415. [CrossRef]
13. Jost, P.; Vanderheynst, P.; Frossard, P. Tree-based pursuit: Algorithm and properties. *IEEE Trans. Signal Process.* 2006, 54, 4685–4697. [CrossRef]
14. Donoho, D.L.; Tsag, Y.; Drori, I.; Starck, J.L. Sparse Solution of Underdetermined Systems of Linear Equations by stagewise Orthogonal Matching Pursuit. *IEEE Trans. Inf. Theory* 2012, 58, 1094–1121. [CrossRef]
15. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* 1996, 58, 267–288. [CrossRef]
16. Gorodnitsky, I.F.; Rao, B.D. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.* 2002, 45, 600–616. [CrossRef]
17. Figueiredo, M.; Nowak, R.D.; Wright, S.J. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE J. Sel. Top. Signal Process.* 2008, 1, 586–597. [CrossRef]
18. Huggins, P.S.; Zucker, S.W. Greedy Basis Pursuit. *IEEE Trans. Signal Process.* 2007, 55, 3760–3772. [CrossRef]
19. Bobin, J.; Starck, J.L.; Fadili, J.M.; Moudden, Y.; Donoho, D.L. Morphological Component Analysis: An Adaptive Thresholding Strategy. *IEEE Trans. Image Process.* 2007, 16, 2675–2681. [CrossRef]
20. Daubechies, I.; Defrise, M.; Mol, C.D. Iteratively solving linear inverse problems under general convex constraints. *Commun. Pure Appl. Math.* 2010, 57, 1413–1457. [CrossRef]
21. Zhang, S.; Zhang, C.; Wang, Z.; Kong, W. Combining Sparse Representation and Singular Value Decomposition for Plant Recognition. *Appl. Soft Comput.* 2018, 67, 164–171. [CrossRef]
22. Zhang, C.; Liu, H.; Qiang, Z. Face recognition based on image sparse representation and singular value decomposition. *J. Comput. Appl.* 2013, 33, 233–235.

23. Zhong, M.; Tang, H.; Chen, H.; Tang, Y. An EM algorithm for learning sparse and overcomplete representations. *Neurocomputing* 2004, 57, 469–476. [CrossRef]

24. Bruckstein, A.M.; Donoho, D.L.; Elad, M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Rev.* 2009, 51, 34–81. [CrossRef]

25. Zhao, M.; Jia, X. A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery. *Mech. Syst. Sig. Process.* 2017, 94, 129–147. [CrossRef]

26. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* 2018, 69, 401–412. [CrossRef]

27. Zhang, H.; Chen, X.; Du, Z.; Yan, R. Kurtosis based weighted sparse model with convex optimization technique for bearing fault diagnosis. *Mech. Syst. Signal Process.* 2016, 80, 349–376. [CrossRef]