Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary

Yi Qin, *Member, IEEE*, Jingqiang Zou, Baoping Tang, Yi Wang, Haizhou Chen

***Abstract*—To detect the incipient faults of rotating parts used in electromechanical systems widely, a novel transient feature extrac-tion method based on the improved orthogonal matching pursuit (OMP) and one-dimensional K-SVD algorithm is explored. Firstly, the stopping criterion of adaptive spark is developed, and then the corresponding OMP algorithm is used to re-move the modulated and harmonic signals adaptively. Secondly, the residual signal is reformulated as a signal matrix by period segmentation and circulating shift, and the initial transient dictionary is constructed via the time-domain average technique. Subsequently, a novel K-SVD algorithm is proposed to get the optimized transient dictionary for the one-dimensional signal. Finally, the repetitive transient signal is recovered by the optimized dictionary. The simulated and experimental results show that the proposed method can not only much faster extract the fault characteristics than the traditional K-SVD method, but also more accurately detect the repetitive transients than the infogram method and the traditional K-SVD method.**

***Index Terms*—dictionary learning, circulating shift, impulsive feature, bearing fault diagnosis, gear fault diagnosis.**

I. INTRODUCTION

As the important rotating parts, gears and bearings play an important role in electromechanical equipments, such as wind turbines, motors, aero-engines, et al. After the electromechanical systems run for a period of time, gears and bear-ings are often firstly subject to failure, and most faults occur in these rotating parts. For example, rolling bearing failures account for 40% of faults in induction motors [1]. With the gradual deterioration of the fault, it will seriously affect the operating per-formance of equipment, and even cause great economic losses and casualties. Therefore, in order to guarantee the reliability service of important equipments, it is necessary to diagnose the fault as soon as possible [2-5].

When gears and rolling bearings fail due to fatigue or other reasons, the running surface will suffer from various damages, such as crack, pitting and spalling. A sudden periodic impulsive force will arise when the moving part passes through the damage location. This force will produce a pseudocyclostationary transient signal [6]. However, the measured signal contains not only the repetitive transient signal, but also the interference signals such as harmonics, modulated signals and noise. Only by extract-ing this repetitive transient component from the measured signal can the fault type be accurately detected. Generally, the analysis approaches for fault feature extraction can be divided into three categories—time-domain analysis, frequency-domain analysis and time-frequency analysis. The time-domain analysis methods calculate the statistical parameters of a time-domain signal to detect the fault type. The commonly-used statistical parameters in time domain include peak value, root-mean-square value, kurtosis, et al [7]. The frequency-domain methods transform signals from time domain to frequency domain for detecting the fault characteristic frequency, including fast Fourier transform, envelope spectrum [8], et al. Time-frequency analysis methods include short time Fourier transform [9], wavelet transform [10-11], Hilbert-Huang transform (HHT) [12], which have been widely used to analyze the nonstationary signals of rotating parts. For example, Wang et al. proposed a method based on EEMD and tunable Q-factor wavelet transform to extract early fault characteristics of bearings [13]. Baydar et al. used the Wigner-Ville distribution to compare the acoustic and vibration signals of the faulty gears [14]. In the last ten years, the spectral kurtosis (SK) method is widely used to extract non-Gaussian transient signals from the vibration signals of faulty gears and bearings. Wang et al [15] found that spectral kurtosis could be decomposed into squared envelope and the square of the l_2 to l_1 ratio. This decomposition is interesting because l_2 to l_1 ratio is actually a modified sparse measure. The maximization of l_2 to l_1 ratio can result in repetitive transients, so it is clear to see that how important the decomposition of spectral kurtosis is for extrac-tion of repetitive transients and how the l_2 to l_1 ratio is connected to sparse representation. However, when the rolling bearing or gear has an incipient fault, the impulsive signal will be submerged by noise, harmonics and modulated signals, thus the SK method may lose efficacy in such case.

The above fault diagnosis methods usually detect the transient feature in different frequency bands, while sparse represen-ta-tion represents the original signal by linear representation of atoms in the dictionary, which can effectively avoid the

Manuscript received January 6, 2019; revised March 1, 2019; accepted March 29, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 51675065, in part by National Key R&D Program of China under grant 2018YFB2001300, in part by Chongqing Research Program of Basic Research and Frontier Technology under Grant cstc2017jcyjAX0459, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018CDQYJX0011.

Y. Qin, J. Zou, B. Tang and Y. Wang are with the State Key Laboratory of Mechanical Transmission, College of Mechanical Engineering, Chongqing University, Chongqing 400044, China (e-mail: qy_808@aliyun.com; binggehe@qq.com; bptang@cqu.edu.cn; wycqdx@cqu.edu.cn). H. Chen is with the Qingdao University of Science and Technology, Qingdao 266000, China (e-mail: chenhaizhou84@163.com)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2909305, IEEE Transactions on Industrial Informatics

influence of frequency aliasing. Therefore, sparse representation has been widely explored for transient feature detection in the past decade. Sparse reconstruction is the key part of sparse representation. It can be mainly categorized as two types: convex relaxation methods and greedy methods. The convex relaxation methods replace the non-convex l_0 norm term with a non-smooth but convex l_1 norm, and then a convex optimization problem is obtained that can be solved by an efficient numerical optimization method. The most representative convex relaxation methods are basis pursuit (BP) [16] and Lasso [17]. For the greedy method, its major characteristic is calculating the support set of the sparse signal in the dictionary and the value of the non-zero coefficient alternately. In each iteration, the local optimal search strategy is used to reduce the current reconstructed residual, so as to obtain a more accurate signal estimation. The typical greedy methods are matching pursuit (MP) [18] and orthogonal matching pursuit (OMP) [19]. Due to its simplicity and high computational efficiency, OMP has been paid extensive attention. For example, to represent harmonics and modulated signals, an appropriate dictionary can be constructed via their morphologies, and then the corresponding signal components can be effectively extracted by OMP. However, due to the complex spectral structure and time-varying morphological characteristics of pseudocyclostationary transient signals, it is difficult to represent them by a fixed dictionary. To solve this problem, a feasible method is to construct a data-driven dictionary (i.e. dictionary learning). The idea of dictionary learning was proposed by Olshausen, and the existing dictionary learning methods mainly contain MOD [20], RLS-DLA [21], K-SVD [22], where K-SVD is the most widely used dictionary algorithm. A number of improved K-SVD algorithms have been developed to solve various engineering problems. Shift-invariant K-SVD [23] could be used to learn a special dictionary to extract the transient signal in the case that the rotating speed was constantly changing. Discriminative K-SVD [24] and LC K-SVD [25] were used to classify the types of face images. Analysis K-SVD [26], i.e. the inverse problem of K-SVD, was effective for removing image noise. Unfortunately, K-SVD is mainly used for image processing. As for the vibration signal, it needs an initialization to meet the requirements of K-SVD. Zhou et al. converted two-dimensional sound signals into Hankel matrix, and then used analysis K-SVD for noise reduction and feature extraction [27]. Nevertheless, when the signal is too large, the size of the Hankel matrix constructed by the original signal is also large, thereupon the operation of SVD requires too much calculation time, and even it cannot be achieved. In addition, the initial dictionary of K-SVD is usually chosen as such a fixed dictionary as DCT dictionary, which may make the algorithm converge to the local optimal point in the calculation process. Therefore, it is necessary to study a new initialization method of signal matrix and an initial dictionary construction method to solve these problems.

As previously mentioned, the pseudocyclostationary (repetitive) transient signal is often disturbed by the strong harmonics and modulated signals, so it is hard to directly extract the transient signal by K-SVD. Thereupon OMP is firstly applied to remove harmonics and modulated components. However, the target spark should be set in advance. In this study, we will develop a stopping criterion via the adaptive spark to improve the adaptivity of OMP. Then an improved K-SVD algorithm for detecting the repetitive transients is explored. By the use of period segmentation and circulating shift, a new initialization method of signal matrix is proposed. Based on the signal matrix, the initial transient dictionary is constructed by the time-domain average of some specific columns in the signal matrix. The improved K-SVD algorithm can well extract the transient feature and greatly improve the computational efficiency. Finally, the proposed method based on the improved OMP and K-SVD algorithm with adaptive initial transient dictionary is applied to process the simulated signal and the actual fault vibration signals, meanwhile it is compared to the K-SVD method based on Hankel matrix and the infogram method [28], which is a development of the SK method. The experimental results indicate that proposed method has a lower computational cost and a higher accuracy of transient feature detection.

The rest of this paper is organized as follows. In Section II, an OMP algorithm based on adaptive spark is proposed to remove the harmonics and modulated components of vibration signals. In Section III, the improved K-SVD algorithm based on adap-tive initial transient dictionary is elaborated. In Section IV, the proposed method is verified by a simulated signal. In Section V, experiments on a faulty gearbox and a faulty bearing of BLDC motor are performed to illustrate the effectiveness and ad-vantage of the proposed method. Conclusion and future work of the study are presented in Section VI.

II. OMP ALGORITHM BASED ON ADAPTIVE SPARK

Due to the influence of faults, manufacturing error and the fact that the perfect alignment is hardly achieved, the vibration signals of gears and bearings contain many harmonics and modulated components, which affect the extraction of impulsive characteristics. They are generated by the rotatory components, such as shaft rotation and gear meshing. Therefore, no matter what kind of fault it is, gear pitting, gear tooth break and damage in inner ring or outer ring, harmonics and modulated signals usually exist in the vibration signal. As a result, it is necessary to separate the harmonics and modulated components from the measured vibration signal.

The observed vibration signals of gears and bearings are commonly described as

$$\mathbf{s} = \mathbf{s}_m + \mathbf{s}_i + \mathbf{n} \tag{1}$$

where $\mathbf{s}$ denotes the observed vibration signal; $\mathbf{s}_m$ represents the harmonics and modulated components; $\mathbf{s}_i$ is the pseudo-cyclostationary transient signal; $\mathbf{n}$ is the noise. Since the three kinds of signals have different morphological characteristics, $\mathbf{s}_m$ can be separated from s by OMP with an appropriate dictionary. It is well-known that the modulated signals are generated from harmonics through the modulated effect, their morphological characteristics are similar. Thereby the Fourier dictionary can be used to linearly represent the modulated signals and harmonics. Assuming that $\mathbf{D}_1$ is the Fourier dictionary, $\boldsymbol{\alpha}$ is a coefficient matrix and T denotes the spark. To extract harmonics and modulated signals, the optimization problem is defined as

$$\min \ \|\mathbf{s}-\mathbf{D}_1\boldsymbol{\alpha}\|_2^2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_0 \le T \tag{2}$$

The atoms in the Fourier dictionary $\mathbf{D}_1$ are given by sine and cosine functions

$$縫\ \varphi(f,0)=\cos(2\pi ft), \quad \varphi(f,1)=\sin(2\pi ft) \tag{3}$$

where f is the frequency parameter. To oversample sine functions and cosine functions, the frequency parameter is defined as $f=k/(l\times N)$, where l is an integer and $l>1$. For cosine functions, $k=0,1,\cdots,l\times N/2$; For sine functions, $k=1,2,\cdots,l\times N/2-1$. Obviously, the over-complete Fourier dictionary consists of $l\times N$ Fourier atoms.

To solve the non-convex optimization problem (2), the orthogonal matching pursuit algorithm (OMP) is employed. The OMP algorithm is divided into the ergodic stage and the update stage of the temporary solution. In the ergodic stage of the k-th iteration, the algorithm searches for the atom that has the largest inner product with current residual.

$$\left|\left\langle \mathbf{r}^{k-1},\mathbf{d}_{j_k}\right\rangle\right| = \sup_{\mathbf{d}_i\in\mathbf{D}_1}\left|\left\langle \mathbf{r}^{k-1},\mathbf{d}_i\right\rangle\right| \tag{4}$$

where $\mathbf{r}^{k-1}$ is the residual after k-1-th iteration; $\mathbf{d}_i$ is the i-th atom in the dictionary; $\mathbf{d}_{j_k}$ is the k-th extracted atom. The index of $\mathbf{d}_j$ in the dictionary is stored into the support set S^k. In the update stage of the temporary solution, $\|\mathbf{s}-\mathbf{D}_1\boldsymbol{\alpha}\|_2^2$ is minimized under the support set S^k. $\mathbf{D}_{S^k}$ is defined as the matrix that consists of columns belonging to this support set in $\mathbf{D}_1$. It follows that the minimization problem is converted into minimizing $\left\|\mathbf{s}-\mathbf{D}_{S^k}\boldsymbol{\alpha}_{S^k}\right\|_2^2$, where $\boldsymbol{\alpha}_{S^k}$ is the non-zero elements. By setting the quadratic derivative as zero, the solution of the minimization problem can be obtained as

$$\mathbf{D}_{S^k}^T\left(\mathbf{D}_{S^k}^T\mathbf{x}_{S^k}-\mathbf{s}\right)=-\mathbf{D}_{S^k}^T\mathbf{r}^k=0 \tag{5}$$

where $\mathbf{r}^k=\mathbf{s}-\mathbf{D}_1\boldsymbol{\alpha}^k=\mathbf{s}-\mathbf{D}_{S^k}\boldsymbol{\alpha}_{S^k}$. Such an operation can make the residual $\mathbf{r}^k$ be orthogonal to the previously extracted atoms, so that these atoms are not reused in subsequent iterations. $\mathbf{d}_{S^i}$ is defined as the i-th atom ($i=1,2,\cdots,k$) in $\mathbf{D}_{S^k}$ and $\boldsymbol{\alpha}_{S^i}$ is defined as the coefficient corresponding to $\mathbf{d}_{S^i}$. As a result, the signal $\mathbf{s}$ can be written as

$$s=\sum_{i=1}^{k}\alpha_{S^i}\mathbf{d}_{S^i}+\mathbf{r}^k \tag{6}$$

It is easy to note that $\mathbf{r}^k$ is orthogonal to $\mathbf{D}_{S^k}$, i.e. $\mathbf{r}^k\mathbf{d}_{S^i}^T=0, k=1,2,\cdots,k$. As $\mathbf{D}_{S^k}$ is an over-complete Fourier dictionary, the atoms in $\mathbf{D}_{S^k}$, which are sine or cosine functions, are orthogonal to each other when their frequencies are different, i.e. $\mathbf{d}_{S^i}\mathbf{d}_{S^j}^T=0,\ i\ne j$ and $i,j=1,2,\cdots,k$. Moreover, $\mathbf{d}_{S^i}$ is a unit vector, i.e. $\left\|\mathbf{d}_{S^i}\right\|_2^2=1$. From the perspective of energy, the signal can be written as

$$\begin{aligned}\|s\|_2^2 &= \left\|\sum_{i=1}^{k}\alpha_{S^i}\mathbf{d}_{S^i}+\mathbf{r}^k\right\|_2^2\\ &=\sum_{i=1}^{k}\alpha_{S^i}^2\left\|\mathbf{d}_{S^i}\right\|_2^2+\left\|\mathbf{r}^k\right\|_2^2\\ &=\sum_{i=1}^{k}\alpha_{S^i}^2+\left\|\mathbf{r}^k\right\|_2^2\end{aligned} \tag{7}$$

It is easy to note that the harmonics and modulated components cluster in several certain frequencies, thus the coefficients corresponding to these frequencies are much bigger than other coefficients and there are only several big coefficients.

The traditional terminating condition of OMP is that a tar-get spark (the number of iteration) is reached or the residual is less than a certain value. From (7), we can note that the atom corresponding to the largest coefficient plays the most important role in representing the harmonics and modulated components. Therefore, in order to improve the adaptivity of the OMP algorithm, a stopping criterion based on the adaptive spark is proposed in Remark 1.

Remark 1: Suppose that the largest element in the coefficient matrix is a_1 after the k-th iteration, and the smallest element is a_k. If $a_1>ca_k$, then the iterative process is terminated.

Via a large number of experiments, we find that this stop-ping criterion is very effective for extracting the harmonics and modulated components from the complex signal when $c=4$, and the optimal spark can be got.

III. IMPROVED K-SVD ALGORITHM BASED ON ADAPTIVE TRANSIENT DICTIONARY

By the use of the OMP based on the stopping criterion, most of the harmonics and modulated components have been separated from the fault vibration signal, and the residual signal mainly contains the repetitive transient component, noise and other components. In order to obtain more obvious fault signature, it is important to further remove the remaining interference signals. Consequently, an improved K-SVD algorithm based on the construction of signal matrix and initial transient dictionary is explored to accurately and adaptively detect the pseudocyclostationary transients.

A. K-SVD algorithm

K-SVD is an over-complete dictionary learning algorithm proposed by Aharon et al. [22], which can effectively reduce the sparsity of the corresponding coefficient matrix of dic-tionary. The signal can be better represented by the atoms learnt by the K-SVD algorithm. The essence of K-SVD is to solve the following optimization problem

$$\min_{\mathbf{D},\mathbf{X}}\{\|\mathbf{Y}-\mathbf{DX}\|_F^2\} \quad \text{s.t.} \quad \forall i,\|\mathbf{x}_i\|_0<T_0 \tag{8}$$

where $\mathbf{Y}$ denotes the sample set; $\mathbf{D}$ is an over-complete dictionary; $\mathbf{X}$ is the coefficient matrix. T_0 is the supremum of spark.

The K-SVD algorithm includes two stages: the sparse coding stage and the dictionary learning stage. First, assume that

$$\begin{gathered}\mathbf{D} \in R^{n\times K}, \mathbf{y} \in R^{n}, \mathbf{x} \in R^{K} \\ \mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N}, \mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}\end{gathered} \tag{9}$$

where $\mathbf{D}$ is an over-complete dictionary; $\mathbf{y}$ is the training signal; $\mathbf{x}$ is the coefficient vector of training signal; $\mathbf{Y}$ is the set of N training signals; $\mathbf{X}$ is the set of solution vectors for $\mathbf{Y}$; and N is much larger than K. In the sparse coding stage, the following optimization problem need be solved

$$\min_{\mathbf{x}_i}\left\{\left\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\right\|_2^2\right\} \quad s.t. \quad \forall i, \left\|\mathbf{x}_i\right\|_0 \le T_0, i = 1,2,\cdots,N \tag{10}$$

The coefficient vectors corresponding to the current dictionary are obtained by sparse coding. The next is the dictionary learning stage. Suppose that $\mathbf{d}_k$ is the k-th column vector of the dictionary $\mathbf{D}$ which need be updated, then the decomposition of the sample set can be described as

$$\left\|\mathbf{Y} - \mathbf{D}\mathbf{X}\right\|_F^2 = \left\|\left(\mathbf{Y} - \sum_{j\ne k}\mathbf{d}_j\mathbf{x}_T^j\right) - \mathbf{d}_k\mathbf{x}_T^k\right\|_F^2 = \left\|\mathbf{E}_k - \mathbf{d}_k\mathbf{x}_T^k\right\|_F^2 \tag{11}$$

where $\mathbf{x}_T^k$ is the k-th row vector in the coefficient matrix $\mathbf{X}$. $\mathbf{E}_k$ is the decomposition error matrix. In order to perform singular value decomposition (SVD), the following definitions are introduced

$$\begin{aligned}\boldsymbol{\omega}_k &= \left\{i \mid 1 \le i \le K, \mathbf{x}_T^k(i) \ne 0\right\} \\ \mathbf{x}_R^k &= \mathbf{x}_T^k\boldsymbol{\Omega}_k \\ \mathbf{Y}_k^R &= \mathbf{Y}\boldsymbol{\Omega}_k \\ \mathbf{E}_k^R &= \mathbf{E}_k\boldsymbol{\Omega}_k\end{aligned} \tag{12}$$

where $\boldsymbol{\omega}_k$ is an index set of atoms used for sparse representation. $\boldsymbol{\Omega}_k$ is a matrix of the size $N \times l(\boldsymbol{\omega}_k)$ ($l(\boldsymbol{\omega}_k)$ is the length of $\boldsymbol{\omega}_k$), whose elements at $(\boldsymbol{\omega}_k(i), i)$ are 1 and others are 0. $\mathbf{x}_R^k$, $\mathbf{Y}_k^R$, $\mathbf{E}_k^R$ are the shrinkage results of $\mathbf{x}_T^k$, $\mathbf{Y}$, $\mathbf{E}_k$ by removing the zero input, respectively. It follows that the equation (9) can be rewritten as

$$\left\|\mathbf{E}_k\boldsymbol{\Omega}_k - \mathbf{d}_k\mathbf{x}_T^k\boldsymbol{\Omega}_k\right\|_F^2 = \left\|\mathbf{E}_k^R - \mathbf{d}_k\mathbf{x}_R^k\right\|_F^2 \tag{13}$$

Then $\mathbf{E}_k^R$ is processed by SVD so that $\mathbf{E}_k^R = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$. The first column of $\mathbf{U}$ is applied to replace the previous $\mathbf{d}_k$. In this way, all atoms $\mathbf{d}_k$ ($k = 1,2,\cdots,K$) in $\mathbf{D}$ are updated column by column, and a new dictionary can be obtained. By repeating sparse coding and dictionary learning, the optimal dictionary and its corresponding coefficient matrix of the sample set $\mathbf{Y}$ can be obtained.

B. Construction of signal matrix and initial transient dictionary

The requirement of K-SVD is $N > K$, and the dictionary used in K-SVD is an over-complete one, i.e. $K > n$. It then follows that $N > n$, i.e. the number of columns of the signal matrix must be larger than the number of rows of the signal matrix. K-SVD cannot process 2D signals, so the signals are usually converted to Hankel matrix [27]. After performing OMP, we obtain the residual signal $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$. The corresponding Hankel matrix (signal matrix) can be written as

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{n'} \\ x_2 & x_3 & \cdots & x_{n'+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_d & x_{d+1} & \cdots & x_n \end{bmatrix} \tag{14}$$

To perform K-SVD, $n' \gg d$ must be satisfied. Obviously, if the signal length is too large, the size of the matrix is also too large, which decrease the computational efficiency. More seriously, SVD cannot be achieved for a large matrix. Therefore, a new construction method of signal matrix is proposed. Firstly, the signal is segmented according to the main period (i.e. the average period of repetitive transients) of $\mathbf{x}$. Then, the segmented signals are circularly shifted to obtain the signal matrix that meets the requirement of K-SVD algorithm. Moreover, generating signal matrix via circulating shift is also significant for the construction of initial transient dictionary. The intervals between two different adjacent impulses in pseudocyclostationary impulsive signal are slightly different, as shown in Fig. 1. It follows that the atoms are generated via the circulating shift and time-domain averaging, and their impulsive positions are slightly different in such case.

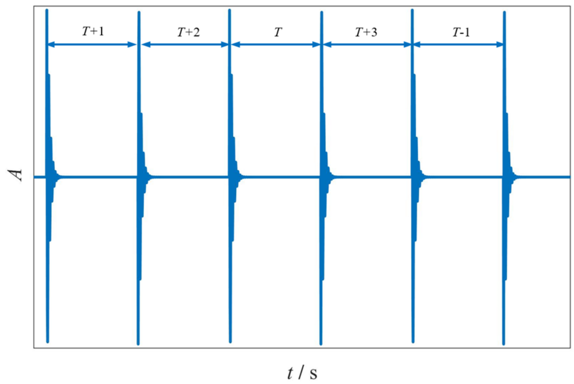

Fig. 1 Pseudocyclostationary impulsive signal, where T is the main period.

If the impulsive frequency or the fault characteristic frequency is known, the segmented period can be directly calculated. If the fault type is unknown, noise resistant correlation (NRC) proposed by Fan et al. [29] can be used to calculate the segmented period of $\mathbf{x}$. Then the number of signal segmentation is calculated as $h = n / T$, where T is the number of points in one period. By the use of signal segmentation, the following matrix can be formulated as

$$\mathbf{X}_0 = \begin{bmatrix} x_1 & x_{T+1} & \cdots & x_{(h-1)\times T+1} \\ x_2 & x_{T+2} & \cdots & x_{(h-1)\times T+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_T & x_{2T} & \cdots & x_n \end{bmatrix} \in R^{T\times h} \tag{15}$$

For the gear or rolling bearing, its fault characteristic frequency is often not very large, which leads to a large T. In such case, the number of rows is larger than the number of columns for $\mathbf{X}_0$, i.e. $T > h$. Therefore, the matrix should be further processed to meet the requirement of K-SVD. Circulating shift is a way to move a whole vector circularly. This method can generate many new signal vectors whose morphological structures are approximate to that of the original signal vector, but the positions of transients are slightly different. Suppose that we circularly shift the signal vector by *p* points (*p* can be set as 1 or 2.) in one circulating shift, the signal matrix obtained by *k* times circulating shifts can be denoted as

$$\mathbf{X}_k = \begin{bmatrix} x_{T-k\times p+1} & x_{2T-k\times p+1} & \cdots & x_{hT-k\times p+1} \\ x_{T-k\times p+2} & x_{2T-k\times p+2} & \cdots & x_{hT-k\times p+2} \\ \vdots & \vdots & & \vdots \\ x_T & x_{2T} & \ddots & x_{hT} \\ x_1 & x_{T+1} & & x_{(h-1)\times T+1} \\ \vdots & \vdots & & \vdots \\ x_{T-k\times p} & x_{2T-k\times p} & \cdots & x_{hT-k\times p} \end{bmatrix} \in R^{T\times h} \quad (16)$$

where *k* is a positive integer.

If the total number of circulating shifts is *m*, these obtained matrices form the signal matrix, which is defined as

$$\mathbf{X} = \left[\mathbf{X}_0 \quad \mathbf{X}_1 \quad \cdots \quad \mathbf{X}_m\right] \in R^{T\times[(m+1)\times h]} \quad (17)$$

where *T* is the number of rows and $(m+1)\times h$ is the number of columns of **X**. If $(m+1)\times h > T$, **X** is a signal matrix that meets the requirement of K-SVD. Obviously, when *m* is smaller, the signal matrix has fewer elements. It then follows that the calculated amount of K-SVD can be effectively decreased. As a result, to satisfy the requirement of K-SVD, *m* can be just chosen as a small number.

In the K-SVD algorithm, the initial dictionary can affect the speed of the algorithm to reach the optimal solution and the accuracy of feature extraction. What is more, an inappropriate initial dictionary may make the algorithm converge to the local optimal solution. In order to improve the accuracy, speed and adaptivity of transient feature extraction, a new construction method for an adaptive initial transient dictionary is proposed according to the established signal matrix. Time-domain average algorithm [30] is a widely-used method, which can eliminate the influence of aperiodic components and random components effectively but retain the periodic components as much as possible. When processing the vibration signals of faulty gears and bearings, the time-domain average can exclude the interference of noise and non-synchronous modulated components. The matrices obtained by segmenting and circulating shift are composed of a series of periodic signal fragments. By averaging the *m*+1 matrices ($\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_m$) by column, *m*+1 atoms are obtained and the noise in each atom can be effectively reduced. The *k*-th atom can be written as

The number of the atoms obtained by time domain average generally cannot constitute an over-complete dictionary. Hence $\mathbf{d}_i^{\mathbf{X}}, i = 1, 2, \cdots, q$ are randomly chosen from the signal matrix **X** to generate the initial transient dictionary with the previous *m*+1 atoms, and the initial dictionary is defined as (19)

$$\mathbf{d}_k = \begin{bmatrix} \sum_{i=0}^{h} x_{(i+1)\times T-k\times p+1} \Big/ h \\ \sum_{i=0}^{h} x_{(i+1)\times T-k\times p+2} \Big/ h \\ \vdots \\ \sum_{i=0}^{h-1} x_{(i+1)\times T} \Big/ h \\ \sum_{i=0}^{h-1} x_{i\times T+1} \Big/ h \\ \vdots \\ \sum_{i=0}^{h-1} x_{(i+1)\times T-k\times p} \Big/ h \end{bmatrix}, \ k = 0, 1, \cdots, m \quad (18)$$

$$\mathbf{D}_2 = \left[\mathbf{d}_0 \quad \cdots \quad \mathbf{d}_m \quad \mathbf{d}_1^{\mathbf{X}} \quad \cdots \quad \mathbf{d}_q^{\mathbf{X}}\right] \in R^{T\times(m+1+q)} \quad (19)$$

By experience, the number of dictionary columns has little influence on the result of K-SVD so long as it is in the suitable range. In this study, the number of columns *m*+1+*q* is set as 0.8*h*, then *q* is determined.

C. Signal recovery

Since the improved K-SVD algorithm is applied to process one-dimensional signals, it is name as *one-dimensional K-SVD* in this study. With the one-dimensional K-SVD, the learnt dictionary $\mathbf{D}_L$ and its corresponding coefficient matrix $\mathbf{A}_L$ can be obtained. However, the atoms in $\mathbf{D}_L$ may be still disturbed by a small amount of noise. Therefore, the learnt dictionary $\mathbf{D}_L$ can be further optimized by solving the following optimization problem

$$\arg\min_{\mathbf{D}_2^*} \left\|\mathbf{D}_o - \mathbf{D}_L\right\|_2^2 + \lambda\left\|\mathbf{D}_o\right\|_0 \quad (20)$$

The problem (20) can be solved by hard thresholding, and then the optimized dictionary $\mathbf{D}_o$ is obtained, where the noise has been further removed. For a signal, the standard deviation represents the degree of data dispersion. Thus the threshold can be determined by the standard deviation. Suppose that the standard deviation of the optimized dictionary is defined as std, then the threshold can be set as $5\times std$ on the basis of a series of experimental results. In this study, the threshold is taken in the range of [0.15, 0.25]. With the optimized dictionary, the signal matrix is recovered by the following formulation

$$\mathbf{X}' = \mathbf{D}_o \times \mathbf{A}_L = \left[\mathbf{X}_0' \quad \mathbf{X}_1' \quad \cdots \quad \mathbf{X}_m'\right] \quad (21)$$

The submatrix $\mathbf{X}_0'$ in $\mathbf{X}'$ corresponds to $\mathbf{X}_0$ in the initial signal matrix **X**, thus we can obtain from Eq. (13) that

$$\mathbf{X}_0' = \begin{bmatrix} x_1' & x_{T+1}' & \cdots & x_{(h-1)\times T+1}' \\ x_2' & x_{T+2}' & \cdots & x_{(h-1)\times T+2}' \\ \vdots & \vdots & \ddots & \vdots \\ x_T' & x_{2T}' & \cdots & x_n' \end{bmatrix} \in R^{T\times h} \quad (22)$$

Then all columns of $\mathbf{X}_0'$ are combined to get the final pseudocyclostationary transient signal $\mathbf{x}'$

$$\mathbf{x}' = \left[x_1' \quad x_2' \quad \cdots \quad x_n'\right]^T \quad (23)$$

By combining the OMP algorithm based on adaptive spark and the improved K-SVD algorithm based on the adaptive transient dictionary, a new transient feature extraction method is proposed, and its explicit steps are described in Fig. 2.

The proposed transient feature extraction method

Input: the measured signal s with the length of n, the parameter of over-complete Fourier dictionary l, the number of points in one circulating shift p, the total number of circulating shift m, the number of randomly chosen atoms q, the error goal of K-SVD τ, the number of iterations of K-SVD L, threshold value ε which is taken in the range of [0.15, 0.25] in this study.

Initialize: $k=0$ and $\mathbf{R}^0\mathbf{s}=\mathbf{s}$

Procedure:

1. Construct the Fourier dictionary $\mathbf{D}_1 \in R^{n\times(l\times n)}$.

2. Separate modulated signal by OMP:

$$\mathbf{D}_{S^k}^T\left(\mathbf{D}_{S^k}^T\mathbf{x}_{S^k}-\mathbf{s}\right)=-\mathbf{D}_{S^k}^T\mathbf{r}^k=0$$

3. If $a_n^1>4a_n^{k+1}$, go to 5; otherwise, let $k\leftarrow k+1$, and repeat 2-3.

4. Separate the harmonics and modulated components from the original signal:

$$\mathbf{x}=\mathbf{s}-\mathbf{s}_{k+1}$$

5. Determine the segmented period according to the fault characteristic frequency or calculate the segmented period by the NRC method, then calculate the number of points in one period T according to the segmented period.

6. Construct the following matrix by signal segmentation:

$$\mathbf{X}_0=\begin{bmatrix} x_1 & x_{T+1} & \cdots & x_{(h-1)\times T+1} \\ x_2 & x_{T+2} & \cdots & x_{(h-1)\times T+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_T & x_{2T} & \cdots & x_n \end{bmatrix}$$

7. Circularly shift k times of $\mathbf{X}_0$ with p, and obtain $\mathbf{X}_k$; after m times of circulating shifts, the signal matrix is formulated as:

$$\mathbf{X}=\left[\mathbf{X}_0 \quad \mathbf{X}_1 \quad \cdots \quad \mathbf{X}_m\right]\in R^{T\times[(m+1)\times h]}$$

8. Generate m+1 atoms $\mathbf{d}_k$ $(k=0,1,\cdots,m)$ by the column average of m+1 submatrices, and then construct the initial transient dictionary $\mathbf{D}_2$ by combining q atoms that are randomly chosen from $\mathbf{X}$:

$$\mathbf{D}_2=\left[\mathbf{d}_0 \quad \cdots \quad \mathbf{d}_m \quad \mathbf{d}_1^{\mathbf{X}} \quad \cdots \quad \mathbf{d}_q^{\mathbf{X}}\right]\in R^{T\times(m+1+q)}$$

9. Process the signal matrix by the K-SVD algorithm based on τ, l and $\mathbf{D}_2$, then get the learnt dictionary $\mathbf{D}_L$ and its corresponding coefficient matrix $\mathbf{A}_L$.

10. Process $\mathbf{D}_L$ by hard thresholding ε, and get the optimized dictionary $\mathbf{D}_o$.

11. Calculate the denoised signal matrix:

$$\mathbf{X}'=\mathbf{D}_o\times\mathbf{A}_L=\left[\mathbf{X}'_0 \quad \mathbf{X}'_1 \quad \cdots \quad \mathbf{X}'_m\right]$$

12. Recover the signal from $\mathbf{X}'_0$:

$$\mathbf{x}'=\left[x'_1 \quad x'_2 \quad \cdots \quad x'_n\right]^T$$

Fig. 2 Explicit steps of the proposed method.

IV. SIMULATION ANALYSIS

In order to demonstrate the effectiveness of the proposed method, it is applied to analyze a simulated signal and compared to the K-SVD method based on Hankel matrix and the SES infogram method based on filter bank which is sensitive to discrete frequencies in a certain frequency band. The K-SVD based on Hankel matrix just takes the place of the improved K-SVD in the procedure of the proposed transient feature extraction method, and the other steps keep the same. Moreover, after performing the K-SVD based on Hankel ma-trix on the residual signal x, the signal can be recovered by taking the average value of each inverse diagonal line in the output matrix of K-SVD.

As mentioned in the Section II, the vibration signals of faulty gears and bearings are usually composed of modulated signals and harmonics, a pseudocyclostationary impulsive signal and noise. Therefore, the simulated signal is defined as

$$\begin{aligned} s=&0.4\cos(2\pi\times3.5t)+0.8\cos(2\pi\times35\,\mathrm{t})+0.4\cos(2\pi\times5\,\mathrm{t})\cos(2\pi\times35t) \\ &+0.4\cos(2\pi\times25\,\mathrm{t}+2\sin(2\pi\,\mathrm{t}))+\sum_k A_k z\left(t-kT_k\right)+n(t) \end{aligned} \tag{24}$$

where A_k is the amplitude of the k-th impulse; T_k is the random occurence of the k-th impulse; and the difference $T_{k+1}-T_k$ is a nonnegative random variable. $z(t)$ can be written as

$$z(t)=\exp\left(-\sigma\omega_n t\right)\cos\left(\omega_n\sqrt{1-\sigma^2}t\right) \tag{25}$$

where ω_n is the natural frequency and ω_n=3240; σ is the damping coefficient and σ=0.111; A_k is the random value in [1,1.5]; n is the white noise whose variance is 1 and the mean value is 0. The mean period of impulses is 0.05 s. The sampling frequency is 2000 Hz, and the data length is 2000. The time-domain waveform of simulated signal is shown in Fig. 3.

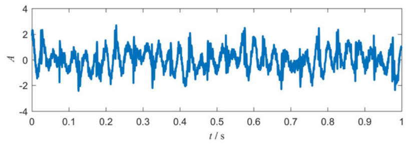

Fig. 3 Time-domain waveform of the simulated signal.

Firstly, the squared envelope spectrum (SES) infogram method was applied to the simulated signal. The obtained infogram is illustrated in Fig.4. With it, the extracted transient component and its envelope spectrum are respectively shown in Fig. 5(a) and (b). It can be seen from Fig. 5 that the infogram method cannot remove the harmonics and modulated components. Moreover, the spectral peaks appear only at the frequencies of the harmonics. Thereupon, the OMP based on adaptive spark are used to process the simulated signal. The obtained signal is shown in Fig. 6, and then it was processed by the traditional K-SVD based on Hankel matrix. As this K-SVD method needs a long time to process the whole signal, the original signal is divided into 5 segments of the length 400. These segments were transformed into Hankel matrices, which can be processed by K-SVD respectively. The obtained transient signal and its envelope spectrum are respectively illustrated in Fig. 7(a) and (b). There are some clear impulses in Fig. 7(a), and the spectral peaks at the fault characteristic frequency and its frequency multiplication in Fig. 7(b) are obvious, whereas the extracted signal is still interfered by some noise obviously. The running time of this method is 13.4 s on a core-i5-7500@ 3.40G-Hz computer. Finally, the proposed K-SVD method is used to extract the transient component. For the proposed method, let $m=20$, $p=1$, $L=20$, $\tau=0.33$ and $\varepsilon=0.25$, $h=400$. Firstly, we consider the effect of column number on transient feature extraction. In the neighborhood of $0.8h$, the column number is respectively set as 310, 330, 320, and then the obtained results are illustrated in Figs. 8-10. The pseudocyclostationary impulses can be clearly seen in the three figures, and it proves that the number of dictionary columns

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2909305, IEEE Transactions on Industrial Informatics

less influences the result of K-SVD. However, when the column number is 320, the best result can be obtained. Secondly, we study the influence of error goal on the detection result. The error goal τ has a great impact on the processed result. Unfortunately, it is difficult to determine this parameter by a mathematical equation or algorithm, so τ is often empirically selected by some trial tests. In this simulation, with the optimal column number, the tests are performed when $\tau = 0.33, 0.30, 0.36$, and the corresponding results are respectively illustrated in Figs. 10-12. Comparing these figures, we can see that the most obvious impulsive feature is extracted when $\tau = 0.33$. Thirdly, the effect of threshold value is investigated under the optimal column number and error goal. When ε is set as 0.25, 0.2 and 0.3 respectively, the obtained results are respectively illustrated in Figs. 10, 13 and 14. It then immediately follows that the most obvious pseudocyclostationary impulsive signal can be detected when $\varepsilon = 0.25$. The standard derivation of the optimized dictionary is 0.0504, which is almost one fifth of 0.25. Therefore, the proposed rule for determining the threshold value ($5 \times std$) is reasonable. Finally, comparing Fig. 7 with Fig. 10, we can see that the impulses detected by the proposed K-SVD are more distinct and the spectral lines at the characteristic frequencies in Fig. 10(b) are much larger than those in Fig. 7(b), especially at 20 Hz. Moreover, the running time of the proposed method is just 5.1 s on the same computing platform. Therefore, the proposed method is superior to the two conventional methods.

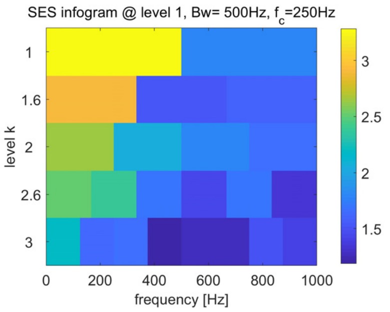

Fig. 4 The SES infogram of the simulated signal.

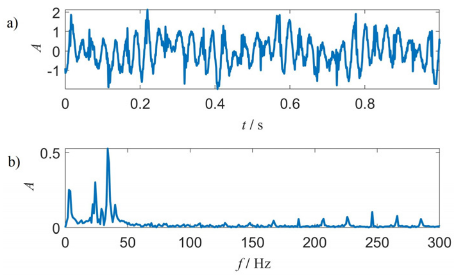
Fig. 5 The result of the simulated signal obtained by the SES infogram method. a) time-domain waveform. b) envelope spectrum.

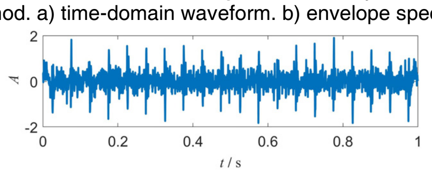
Fig. 6 The signal obtained by OMP based on adaptive spark for the simulated signal.

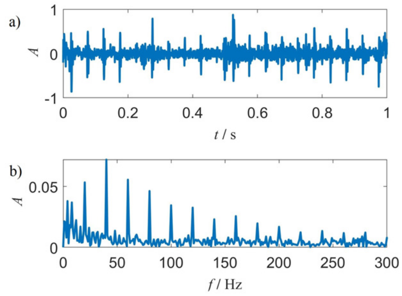
Fig. 7 The result obtained by K-SVD based on Hankel matrix for the simulated signal. a) time-domain waveform b) envelope spectrum.

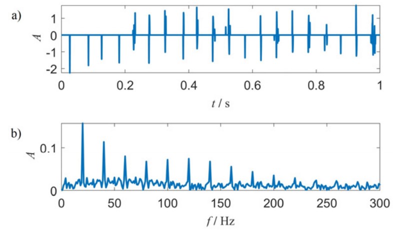
Fig. 8 The result obtained by the proposed method for the simulated signal (the column number is 310, $\tau = 0.33$, $\varepsilon = 0.25$). a) time-domain waveform. b) envelope spectrum.

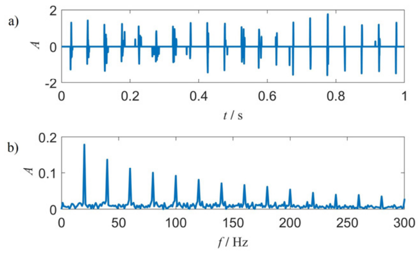
Fig. 9 The result obtained by proposed method for the simulated signal (the column number is 330, $\tau = 0.33$, $\varepsilon = 0.25$). a) time-domain waveform. b) envelope spectrum.

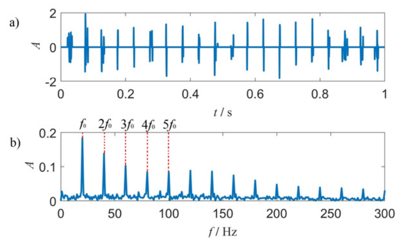
Fig. 10 The result obtained by proposed method for the simulated signal (the column number is 320, $\tau = 0.33$, $\varepsilon = 0.25$). a) time-domain waveform. b) envelope spectrum.

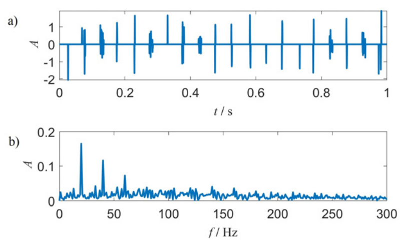
Fig. 11 The result obtained by proposed method for the simulated signal (the column number is 320, $\tau = 0.30$, $\varepsilon = 0.25$). a) time-domain waveform. b) envelope spectrum.

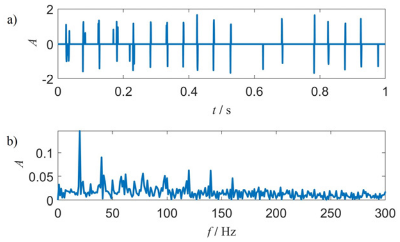

Fig. 12 The result obtained by proposed method for the simulated signal (the column number is 320, $\tau = 0.36$, $\varepsilon = 0.25$). a) time-domain waveform. b) envelope spectrum.

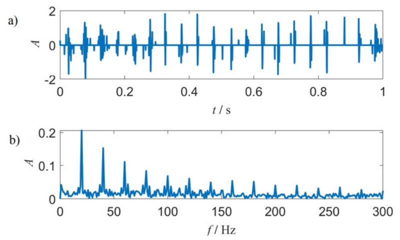

Fig. 13 The result obtained by proposed method for the simulated signal(the column number is 320, $\tau = 0.33$, $\varepsilon = 0.2$). a) time-domain waveform. b) envelope spectrum.

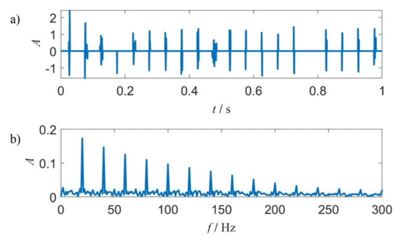

Fig. 14 The result obtained by proposed method for the simulated signal(the column number is 320, $\tau = 0.33$, $\varepsilon = 0.3$). a) time-domain waveform. b) envelope spectrum.

V. APPLICATIONS AND DISCUSSIONS

In order to verify the effectiveness of the proposed method in real application, the proposed method was applied to detect the fault of a secondary gearbox. The structure of this gearbox and its measurement point are illustrated in Fig. 15. The teeth numbers of gears 1, 2, 3, 4 are respectively 30, 69, 18 and 81. The gear 1 in the input shaft has spalling. Measured by the photoelectric tachometric transducer, the rotating speed of the motor is 900 rpm, i.e. the fault characteristic frequency fo is 15 Hz. The accelerometer is placed at the posi-tion of bearing 5. The sampling frequency is 12000 Hz, and the data length is 12800. The sampled signal is shown in Fig. 16.

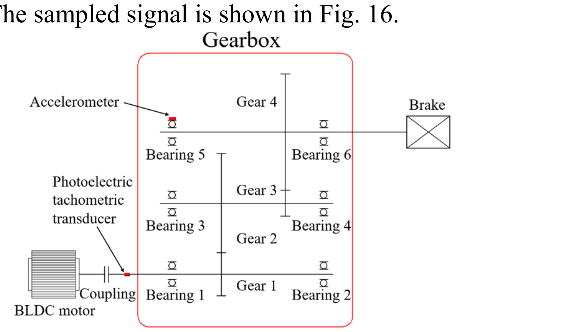

Fig. 15 Structure of the secondary gearbox and its measurement point.

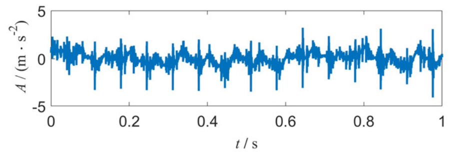

Fig. 16 The sampled vibration signal of a faulty gearbox.

Via the SES infogram method, the obtained transient signal and its envelope spectrum are illustrated in Fig. 17. It can be seen that the infogram method cannot successfully extract the repetitive transient signal. Thereupon, the OMP based on adaptive spark are employed to analyze the vibration signal of the faulty gearbox, and the obtained signal is shown in Fig. 18. Then the K-SVD based on Hankel matrix is applied to extract the transient feature. The obtained transient signal and its envelope spectrum are respectively illustrated in Fig. 19(a) and (b). We can see from this figure that the faulty feature of the gearbox has been well detected and there are clear spectral peaks at f_o and its frequency multiplication. However, the amplitude at f_o is much smaller than that at $2f_o$, and the computation time of the method is up to 1018.1 s.

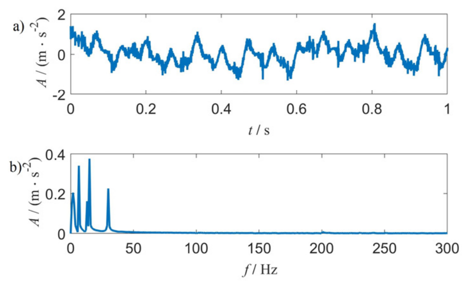

Fig. 17 The result obtained by SES infogram method for the vibration signal of the faulty gearbox. a) time-domain waveform. b) envelope spectrum.

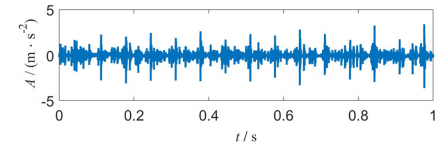

Fig. 18 The signal obtained by OMP based on adaptive spark for the vibration signal of the faulty gearbox.

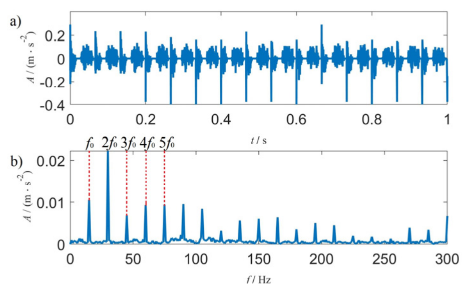

Fig. 19 The result obtained by K-SVD based on Hankel matrix for the vibration signal of the faulty gearbox. a) time-domain waveform b) envelope spectrum.

Similarly, the proposed method was used to detect the fault of gearbox. For the proposed method, let $m = 60$, $p = 1$, $L = 10$, $\tau = 0.25$ and $\varepsilon = 0.17$. The obtained transient signal and its envelope spectrum are respectively illustrated in Fig. 20(a) and (b). It can be seen from Fig. 20(a) that the repetitive transients have been effectively detected, which are less disturbed by the noise than those in Fig. 19(a). And the spectral

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2909305, IEEE Transactions on Industrial Informatics

peaks occur at the fault characteristic frequency more regularly in Fig. 20(b) than those obtained by the K-SVD based on Hankel matrix in Fig. 19(b). Moreover, its running time is only 88.3 s, much less than 1018.1 s. From the above experimental results, we can conclude that the proposed method has the higher accuracy of transient feature extraction than the infogram method and the traditional K-SVD method, and its computation speed is far faster than the traditional K-SVD method.

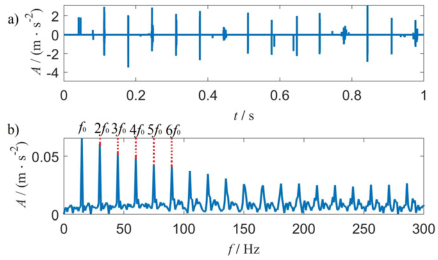

Fig. 20 The result obtained by the proposed method for the vibration signal of the faulty gearbox. a) time-domain waveform. b) envelope spectrum.

To further demonstrate the advantages of the proposed method, the proposed method was applied to detect the fault of a bearing in a brushless direct-current (BLDC) motor which is widely used in industry. As shown in Fig. 21, the test rig consists of generator, accelerometer, BLDC motor, data acquisition and motor controller. There is a small defect in the outer ring of the experimental bearing. The accelerometer is used to sample the vibration signal of BLDC motor. When the rotating speed of the BLDC motor is 1810 rpm and the sampling frequency is 25600 Hz, the acquired vibration signal of the length 16284 is illustrated in Fig. 22. According to the parameters of the bearing [6], the fault characteristic frequency of the outer ring f_o is calculated as 108.3 Hz.

With the SES infogram method, the detection result is shown in Fig. 23. It can be seen that the infogram method is still influenced by the harmonics and modulated components existing in the measured signal. In Fig. 23 (b), spectral peaks only occur in the rotating frequency and its double. It immediately follows that the infogram method has a poor performance on detecting such bearing fault of the BLDC motor. Then the OMP based on adaptive spark and the K-SVD algorithm based on Hankel matrix are applied to process the vibration signal, and the obtained results are respectively illustrated in Figs. 24 and 25. From Fig. 25(a), we can see that there seems to be some repetitive impulses, but the transient feature in the time range of 0.09-0.15 s is not particularly obvious. In Fig. 25(b), although there are clear spectral peaks at f_o, $2f_o$, $3f_o$ and $4f_o$, some large spectral peaks still occur in the low frequency band. Moreover, its calculation time is 236.6 s. Finally, the proposed method is used. For the proposed method, we let $m=10$, $p=2$, $L=10$, $\tau=1.16$ and $\varepsilon=0.2$. The obtained transient signal and its envelope spectrum are respectively illustrated in Fig. 26(a) and (b). It can be noted from this figure that the impulses with obvious period are extracted, and there are very clear spectral peaks at f_o, $2f_o$, $3f_o$, $4f_o$ and $5f_o$. Comparing Fig. 25 with Fig. 26, the signal in Fig. 26(a) hardly has noise and the spectral peaks in Fig. 26(b) are much larger. Better yet, the running time is just 40.7 s.

To better show the advantages of the improved K-SVD algorithm in terms of computational efficiency, the calculation time of the proposed method and the K-SVD method based on Hankel matrix for the simulated signal and real signals are respectively listed in TABLE I. Particularly, the longer the impulsive period of the signal is, the higher the efficiency of the improved K-SVD algorithm is. The above two experiments show that the proposed method has better performance in weak transient feature extraction than the two conventional methods.

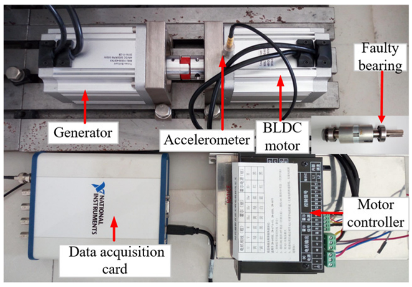

Fig. 21 Test rig for a BLDC motor with bearing fault.

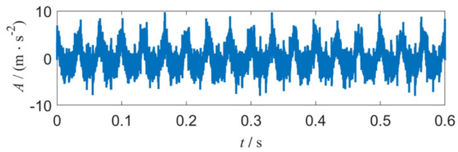

Fig. 22 Vibration signal of a bearing with an outer ring fault in a BLDC motor.

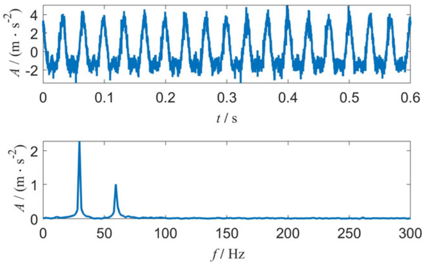

Fig. 23 The result obtained by SES infogram method for the vibration signal of the faulty bearing. a) time-domain waveform. b) envelope spectrum.

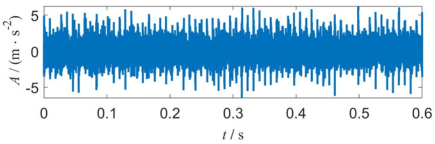

Fig. 24 The signal obtained by OMP based on adaptive spark for the vibration signal of the the faulty bearing.

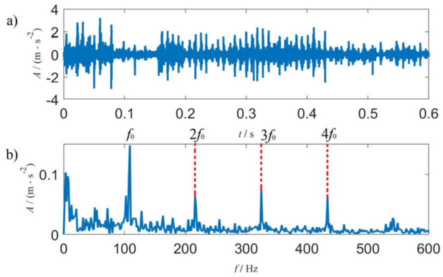

Fig. 25 The result obtained by K-SVD based on Hankel matrix for the vibration signal of the faulty bearing. a) time-domain waveform b) envelope spectrum.

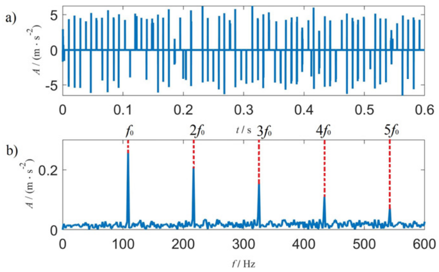

Fig. 26 The result obtained by the proposed method for the vibration signal of the faulty bearing. a) time-domain waveform. b) envelope spectrum.

TABLE I
CALCULATION TIME FOR DIFFERENT SIGNALS

method \ signal	Simulated signal	Gearbox signal	Bearing signal
K-SVD based on Hankel matrix	13.4	1018.1	236.6
The improved K-SVD method	5.1	88.3	40.7

VI. Conclusions

This paper presents an impulsive fault feature extraction method based on the adaptive OMP algorithm and improved K-SVD algorithm. The two main contributions of this study lies in that 1) a novel construction method of the signal matrix and initial transient dictionary for K-SVD is proposed, which can speed up the calculation and improve the accuracy of transient feature detection and 2) an OMP algorithm based on a new stopping criterion of adaptive spark is proposed to remove harmonics and modulated components. It also lays a foundation for the application of dictionary learning in long one-dimensional signal processing. Firstly, according to the morphological characteristics of harmonics and modulated components, OMP can adaptively and accurately separate the harmonics and modulated components from the observed signals by using an over-complete Fourier dictionary and the adaptive spark. Then, the signal matrix is built by the circulating shift and main period segmentation according to the residual signal. Owing to its small dimension, the constructed signal matrix can significantly improve the computational efficiency and reduce the calculation time. With the signal matrix, a construction method of initial transient dictionary based on time-domain average and data-driven idea is proposed. Since this transient dictionary has less noise, it can effectively improve the accuracy of transient feature extraction and avoid one-dimensional K-SVD to converge to a local optimal solution. In the process of recovering the signal, the hard thresholding is performed on the learnt dictionary, and then the repetitive transient signal with higher SNR can be obtained. Compared with the infogram method and the K-SVD method based on Hankel matrix, the simulated and experimental results show that the proposed method is more suitable for weak transient feature extraction, and its computation speed is far faster than that of the traditional K-SVD algorithm. Therefore, the proposed method has great application potential in fault diagnosis.

Future work will pay more attention to the adaptive optimization of some parameters in one-dimensional K-SVD, such as the error goal, the number of atoms and the size of the learnt atoms.

References

[1] W. Sun, R. Zhao, R. Yan, S. Shao, and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1350-1359, 2017.

[2] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814-3824, 2019.

[3] D. Liu, X. Yin, Y. Song, W. Liu, and Y. Peng, "An On-Line State of Health Estimation of Lithium-Ion Battery Using Unscented Particle Filter," *IEEE Access,* vol. 6, pp. 40990-41001, 2018.

[4] L. Cui, B. Li, J. Ma, Z. Jin, "Quantitative trend fault diagnosis of a rolling bearing based on Sparsogram and Lempel-Ziv," *Measurement*, vol. 128, pp. 410-418, 2018.

[5] Y. Wang, Z. Wei, and J. Yang, "Feature Trend Extraction and Adaptive Density Peaks Search for Intelligent Fault Diagnosis of Machines," *IEEE Trans. Ind. Inform.*, 2018.

[6] Y. Qin, "A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2716-2726, 2018.

[7] T. W. Chow and H.-Z. Tan, "HOS-based nonparametric and parametric methodologies for machine fault detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1051-1059, 2000.

[8] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. Signal Proc.*, vol. 25, no. 2, pp. 485-520, 2011.

[9] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," *IEEE Trans. Ind. Inform.*, vol. 9, no. 2, pp. 760-771, 2013.

[10] J. Chen, J. Pan, Z. Li, Y. Zi, and X. Chen, "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals," *Renew. Energy*, vol. 89, pp. 80-92, 2016.

[11] Y. Qin, Y. Mao, and B. Tang, "Multicomponent decomposition by wavelet modulus maxima and synchronous detection," *Mech. Syst. Signal Proc.*, vol. 91, pp. 57-80, 2017.

[12] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proc. R. Soc. A-Math. Phys. Eng. Sci.*, 1998, vol. 454, no. 1971, pp. 903-995: The Royal Society.

[13] H. Wang, J. Chen, and G. Dong, "Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform," *Mech. Syst. Signal Proc.*, vol. 48, no. 1-2, pp. 103-119, 2014.

[14] N. Baydar and A. Ball, "A comparative study of acoustic and vibration signals in detection of gear failures using Wigner–Ville distribution," *Mech. Syst. Signal Proc.*, vol. 15, no. 6, pp. 1091-1107, 2001.

[15] D. Wang, "Spectral L2/L1 norm: A new perspective for spectral kurtosis for characterizing non-stationary signals," *Mech. Syst. Signal Proc.*, vol. 104, pp. 290-293, 2018.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129-159, 2001.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, pp. 267-288, 1996.

[18] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374-377, 2002.

[19] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655-4666, 2007.

[20] I. Tosic and P. Frossard, "Dictionary Learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27-38, 2011.

[21] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2121-2130, 2010.

[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, p. 4311, 2006.

[23] B. Yang, R. Liu, and X. Chen, "Fault diagnosis for a wind turbine generator bearing via sparse representation and shift-invariant K-SVD," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1321-1331, 2017.

[24] Z. Wang, J. Liu, and J.-H. Xue, "Joint sparse model-based discriminative K-SVD for hyperspectral image classification," *Signal Process.*, vol. 133, pp. 144-155, 2017.

[25] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651-2664, 2013.

[26] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661-677, 2013.

[27] W. R. Gong, H. Y. Li, and D. Zhao, "An Improved Denoising Model Based on the Analysis K-SVD Algorithm," *Circuits Syst. Signal Process.*, vol. 36, no. 10, pp. 4006-4021, Oct 2017.

[28] J. Antoni, "The infogram: Entropic evidence of the signature of repetitive transients," *Mech. Syst. Signal Proc.,* vol. 74, pp. 73-94, 2016.

[29] W. Fan, Y. Li, K. L. Tsui, and Q. Zhou, "A Noise Resistant Correlation Method for Period Detection of Noisy Signals," IEEE Trans. Signal Process., 2018.

[30] P. McFadden, "Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration," *Mech. Syst. Signal Proc.*, vol. 1, no. 2, pp. 173-183, 1987.

Yi Qin received the B.Eng. and Ph.D. degrees in mechanical engineering from Chongqing University, Chongqing, China, in 2004 and 2008 respectively.

Since January 2009, he has been with the Chongqing University, Chongqing, China, where he is currently a Professor in the College of Mechanical Engineering. His current research interests include signal processing, fault prognosis, mechanical dynamics and smart structure.

Jingqiang Zou received the B.Eng. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2016.

He is currently working toward the M.S. degree in College of Mechanical Engineering of Chongqing University, Chongqing, China.

His research interests mainly include signal processing and mechanical fault diagnosis.

Baoping Tang received the M.Sc. and Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 1996 and 2003 respectively.

Since April 1996, he has been with the Chongqing University, Chongqing, China, where he is currently a Professor in the College of Mechanical Engineering. His current research interests include mechanical equipment security service and life prediction, measurement technology and instruments, wireless sensor networks.

Yi Wang received his B.Eng. degree from Southwest Jiaotong University, Chengdu, China, in 2011, Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2017, respectively. During 2016.8-2017.2, he was a visiting scholar in City University of Hong Kong, Hong Kong, China.

Since July 2017, he has been a lecture in Chongqing University, Chongqing, China. His current research interests include mechanical signal processing, weak signal detection, rotating machinery fault diagnosis under speed variation conditions, manifold learning and deep learning.

Haizhou Chen received the MA. Eng and Ph.D. degrees in mechanical engineering from Chongqing University, Chongqing, China, in 2010 and 2017 respectively.

Since July 2017, he has been with the Qingdao University of Science and Technology, Qingdao, China, where he is currently a Lecturer in the College of Electromechanical Engineering. His current research interests include failure mechanism analysis, fault prognosis and tribology.