

Article

Sparse Time–Frequency Representation for the Transient Signal Based on Low-Rank and Sparse Decomposition

Liang Yu ^{1,2}, Wei Dai ³, Shichun Huang ¹, Weikang Jiang ^{1,*}

¹ Institute of Vibration, Shock and Noise, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

² Key Laboratory of Aerodynamic Noise Control, China Aerodynamics Research and Development Center, Mianyang, 621000, Sichuan, China

³ Hagong Intelligent Robot Co., Ltd, Shanghai, 201100, China

* Correspondence: Weikang Jiang, Email: wkjiang@sjtu.edu.cn; Tel.: +86-21-34206332.

ABSTRACT

Rolling element bearings are important parts of rotating machinery, and they are also one of the most fault-prone parts in rotating machinery. Therefore, many new algorithms have been proposed to solve the vibration-based diagnosis problem of rolling bearings. The measured vibration signal is typically composed of a periodic transient signal severely contaminated by loud background noise when the faults occur. In this paper, a transient signal extraction algorithm is proposed which depends on spectrum matrix decomposition. The sparse time–frequency representation of the periodic transient signals is exploited, and, further, a low-rank and sparse model is established to extract transient signals from strong noise. First, the low-dimensional representation matrix of the measured signal is generated by the synchrosqueezing transform based on short-time Fourier transform. It is found that the low-rank of the transient signal will be approximately preserved in the transformed domain. Then, semi-soft go decomposition is used to decompose the spectrum matrix into a low-rank matrix and a sparse matrix. Finally, the transient signal can be recovered through the inverse transformation of the decomposed low-rank matrix. The proposed method is a data-driven approach, and it does not require prior training. The performance of the algorithm is investigated on both synthetic and real vibration signals, and the results demonstrate that the algorithm is effective and robust.

KEYWORDS: vibration-based diagnostics of rolling bearings; transient signals; low-dimensional representation; low-rank and sparse model

Open Access

Received: 23 April 2019

Accepted: 20 August 2019

Published: 22 August 2019

Copyright © 2019 by the author(s). Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

Rolling element bearings are one of the most widely used components in mechanical systems, and their failure is one of the most frequent reasons for a machine breakdown [1]. Because of its low cost, high

efficiency and convenient maintenance, it is widely used in the mechanical industry. It will cause a drop in the quality of the product when faults occur, or the bearings fail, and, in severe cases, it will lead to damage to the entire mechanical device. Therefore, considering the importance of bearings in mechanical equipment, fault diagnosis of bearings has caused widespread concern in the scientific community [2]. The ultimate goal of fault diagnosis is to determine the existence of the fault by analyzing the external fault information and determining whether the machine needs maintenance. Despite extensive research in the field, the diagnosis of these faults remains a challenging issue.

Vibration signal analysis has been extensively used in bearing fault diagnosis due to the simplicity of measurement and the relatively low cost of the sensor [3]. Generally, the presence of mechanical bearing faults can be detected by vibration signals analysis. From the literature research, diagnostic methods based on vibration signals analysis have proved their effectiveness in many situations, and researchers continually strive to develop improvements in the diagnostic algorithms based on vibration signals [4]. The defects of the rolling element bearings will produce a series of transient signals in the measured acceleration signal, as the fault repeatedly interacts with other rolling surfaces [5]. However, the measured vibration signals are always buried in background noise from other vibration sources on the machine, which results in the diversity and complexity of the vibration signal characteristics. Therefore, the detection and diagnosis of bearing faults can be considered as the problem of extracting periodic transient signals from the measured signals with strong noise.

In the past few decades, several methods for bearing fault diagnosis have been introduced and developed. These methods deal with vibration signals in a variety of ways and generally can be divided into the following categories [6].

Model-based methods: Models are used for enhancing the contrast between the fault signal and the normal vibration data. The most typical model-based approach is Autoregressive (AR) model, which is a time-sequence analysis method, and its parameters comprise essential information of the system condition [7]. AR model is widespread in machine fault detection and effective in extracting gear faults with little prior knowledge [8,9].

Spectral analysis methods: Spectral analysis methods aim to generate a sparse time–frequency representation of the signal. The spectrum illustrates the characteristics of a signal in the time–frequency representation and presents more information than in the time domain. There are many ways to obtain the sparse time–frequency representation of the signal, such as short-time Fourier transform (STFT) and wavelet transforms [10–12]. Generally, fault signals and noise will show different properties in the time–frequency domain. It means that a sparse transform yields a more distinguishable representation. However, time–

frequency transforms such as STFT and continuous wavelet transform spread the energy of the signal around the instantaneous frequencies of the original signal [13]. Hence, a post-processing method is proposed to sharpen the blurry signal information and to overcome the limitations of traditional time–frequency transforms. Synchrosqueezing transform (SST) utilizes frequency reassignment to shrink the spectrogram along the frequency axis, thus leading to a better resolution [14,15]. Therefore, SST can promote the sparsity of the representation matrix of the signal. Moreover, it is worth mentioning that SST is reversible.

Filtering methods: Filtering methods aim to reduce non-fault signals and enhance the visibility of the fault vibration. The fault information can be better revealed with the prior knowledge of the statistical characteristics of the fault signal. The most typical methods are presented, including Spectral kurtosis-based filtering and deconvolution. Spectral kurtosis (SK) is a method that selects a bandpass filter to maximize the kurtosis of the filtered resulting signal. SK has demonstrated its power in bearing fault diagnosis and has been receiving a lot of research attention within the last few years [16]. Minimum entropy deconvolution (MED) treats fault signal extraction as a deconvolution problem, in which a Finite Impulse Response (FIR) filter is selected to minimize the entropy of the filtered signal [17]. A limitation of MED is that the algorithm generally prefers to deconvolve only a single impulse, as opposed to the desired periodic impulses produced by bearing faults. Inspired by the MED deconvolution technique, an improved novel deconvolution norm was proposed, Correlated kurtosis (CK), which takes advantage of the periodicity of the faults [18]. The effectiveness of maximum correlated kurtosis deconvolution (MCKD) is proved by testing on the simulation and experimental data from a controlled gear tooth chip experiment.

Different from the previous work of building a matrix directly on the time domain and performing low-rank matrix decomposition [19], the present work depends on sparse time–frequency spectral analysis. The proposed method only requires that the signal has no information loss after the transformation, and the transformation is invertible. As mentioned above, the transform can capture the time–frequency structure of a periodic transient signal well. In this paper, the spectrum matrix of the measured signal is calculated using STFT-based SST due to its high resolution and high sparsity properties. It can be seen that the low-rank of transient signals will be approximately preserved during the transformation [20]. Then, the strategy is to decompose the matrix into low-rank and sparse components to obtain the transient signal corresponding matrix. Matrix factorization methods currently enjoy immense popularity in machine learning and signal processing [21]. There are many ways to obtain low-rank information from the noisy observation matrix with less error, such as robust principal component analysis (RPCA)[22,23] and go-decomposition (GoDec)[24].

The remainder of this paper is organized as follows. The signal model in bearing fault diagnosis and the STFT-based SST of signals are introduced in the Section “**LOW DIMENSIONAL REPRESENTATION OF THE SIGNAL**”. The selection of low-rank and sparse decomposition method is provided in Section “**LOW-RANK DECOMPOSITION**”. Then, Section “**VALIDATION ON SYNTHETIC SIGNALS**” presents the validation of the proposed methodology on synthetic signals. Finally, several actual vibration signals are used to verify the performance of the proposed approach in Section “**VALIDATION ON VIBRATION SIGNALS**”.

LOW DIMENSIONAL REPRESENTATION OF THE SIGNAL

Typically, the measured signal $y(t)$ can be described as the additive model,

$$y(t) = x(t) + n(t) \quad (1)$$

where “informative signal” $x(t)$ and the stationary background noise $n(t)$ are assumed mutually independent. In the case of bearing fault diagnosis, the bearing components repeatedly strike the defect, resulting in periodic transient signals $x(t)$.

The transient signal can be well captured in the time–frequency representation, on the contrary to the noise $n(t)$, which is spread all over the time–frequency representation matrix [25]. It is found that different time–frequency transforms do not affect the extraction performance of transient signals through experiments. In this paper, based on the STFT of the signal, a sparser representation matrix is generated using SST.

The STFT expression of a function $s \in L^2(R)$ is defined as:

$$V_f(\eta, t) = \int s(\tau)g(\tau - t)e^{-2i\pi\eta(\tau - t)}d\tau \quad (2)$$

where $g(t)$ is the window function with a unit norm $\|g\| = 1$.

Although the energy of the time–frequency representation $V_f(\eta, t)$ diverges in the ridge direction, the phase of $V_f(\eta, t)$ is not affected by the length of the time window, so the instantaneous frequency $\hat{\omega}_f$ can be estimated by the phase of $V_f(\eta, t)$ [26,27].

$$\hat{\omega}_f(\eta, t) = -i \left(V_f(\eta, t) \right)^{-1} \frac{\partial}{\partial t} V_f(\eta, t) \quad (3)$$

In the actual calculation, η, t and ω_f are discretized, i.e., $V_f(\eta, t)$ is computed only at a discrete a_k , with $a_k - a_{k-1} = (\Delta a)_k$, and its synchrosqueezed transform $T_f(\omega, t)$ is determined only at the center frequency ω_l of successive bins $[\omega_l - \frac{1}{2}\Delta\omega, \omega_l + \frac{1}{2}\Delta\omega]$, with $\Delta\omega = \omega_l - \omega_{l-1}$, by summing different contributions:

$$T_f(\omega_l, t) = \sum_{a_k: |\hat{\omega}_f(a_k, b) - \omega_l| \leq \frac{\Delta\omega}{2}} V_f(a_k, b) a_k^{-\frac{3}{2}} (\Delta a)_k \quad (4)$$

Finally, the signal $x(t)$ can be reconstructed by the inverse transformation of $T_f(\omega_l, t)$:

$$x(t) \approx \text{Re}[C_\psi^{-1} \sum_l T_f(\omega_l, t) (\Delta\omega)] \quad (5)$$

where $C_\psi = \frac{1}{2} \int_0^\infty \overline{\hat{g}(\xi)} \frac{d\xi}{\xi}$, $\overline{\hat{g}(\xi)}$ is the complex conjugate of the Fourier transform of $\hat{g}(\xi)$.

LOW-RANK DECOMPOSITION

Let matrix $\mathbf{X} \in \mathbb{C}^{M \times N}$ denote the STFT-based SST representation of transient signal $x(t)$. The low-rank prior of the transient signal will be approximately preserved in the transformed domain, thus the spectrogram $\mathbf{L} = |\mathbf{X}|^2$ is a low-rank matrix. Matrix $\mathbf{Y} \in \mathbb{C}^{M \times N}$ is defined as the sparse time-frequency transform of the measured signal $y(t)$ and the spectrogram $\mathbf{M} = |\mathbf{Y}|^2$ can be represented as $\mathbf{M} = \mathbf{L} + \mathbf{S}$. $\mathbf{S}$ corresponds to a stationary noise signal that is sparse. Therefore, by using the low-rank and sparse decomposition method, the noisy observation matrix can be decomposed to obtain the low-rank matrix, which containing transient signal information.

Without a constraint of specific rank, Robust Principal Component Analysis (RPCA) defines the low-rank component unclearly. Moreover, due to its convexity relaxation, the decomposed results sometimes are insufficiently promoted for low-rank or sparsity. As an improvement, in [24], the authors considered GoDec with a different signal model, $\mathbf{M} = \mathbf{L} + \mathbf{S} + \mathbf{R}$, with additive Gaussian noise $\mathbf{R}$ and penalizing constraints of rank and non-zero entries directly. Therefore, the problem of low-rank matrix and sparse decomposition can be formally expressed as,

$$\mathbf{M} = \mathbf{L} + \mathbf{S} + \mathbf{R}, \text{rank}(\mathbf{L}) \leq r, \text{card}(\mathbf{S}) \leq k \quad (6)$$

where $\text{rank}(\mathbf{L})$ is the rank of $\mathbf{L}$, and $\text{card}(\mathbf{S})$ is the cardinality of $\mathbf{S}$. The problem in Equation (6) is approximately equivalent to solving the minimization of the following decomposition error:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \quad & \|\mathbf{M} - \mathbf{L} - \mathbf{S}\|_F^2 \\ \text{Subject to} \quad & \text{rank}(\mathbf{L}) \leq r \\ & \text{card}(\mathbf{S}) \leq k \end{aligned} \quad (7)$$

where the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a^2(i,j)}$. The optimization problem in Equation (7) can be solved by alternatively assigning the low-rank approximation of $\mathbf{M} - \mathbf{S}$ to $\mathbf{L}$ and assigning the sparse approximation of $\mathbf{M} - \mathbf{L}$ to $\mathbf{S}$. Moreover, the singular value decomposition (SVD) is replaced with a Bilateral Random Projection (BRP)-based low-rank approximation, to significantly reduce the time cost [28,29](see Appendix A for details).

Generally, the cardinality of $\mathbf{S}$ is hard to estimate. Thus, Semi-Soft Go Dec (SSGD) adopts soft thresholding to the entries of $\mathbf{S}$, where parameter k can be automatically determined by soft threshold λ . Different from

GoDec which imposes hard threshold to the entries of the sparse part $\mathbf{S}$, $\text{card}(\mathbf{S})$ can be relaxed by replacing it with the L1 norm $\|\mathbf{S}\|_{L1} = \sum_{i,j} |S(i,j)|$. The value of regularization λ is the trade-off between the error term and the sparsity of $\mathbf{S}$. Moreover, the computation cost of SSGD is substantially smaller than the original GoDec while the error rate is kept the same or even smaller. SSGD is formulated as the optimization problem in Equation (8), can be solved by the alternative optimization method, and the corresponding algorithm is presented in Appendix A.

$$\begin{aligned} \min_{L, S} \quad & \|\mathbf{M} - \mathbf{L} - \mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_{L1} \\ \text{Subject to} \quad & \text{rank}(\mathbf{L}) \leq r \end{aligned} \quad (8)$$

Here, the convergence of SSGD needs to be discussed. Following the original GoDec [24], we can prove that SSGD can also converge to local optimum. The optimization problem in Equation (8) is equivalent to the following form.

$$\begin{aligned} \min_{L, S} \quad & \|\mathbf{M} - \mathbf{L} - \mathbf{S}\|_F^2 \\ \text{Subject to} \quad & \text{rank}(\mathbf{L}) \leq r \\ & \|\mathbf{S}\|_{L1} \leq k \end{aligned} \quad (9)$$

Then, the optimization problem in Equation (9) can be solved by alternatively solving the following two subproblems until convergence:

$$\begin{cases} L_t = \arg \min_{\text{rank}(\mathbf{L}) \leq r} \|\mathbf{M} - \mathbf{L} - \mathbf{S}_{t-1}\|_F^2 \\ \mathbf{S}_t = \arg \min_{\|\mathbf{S}\|_{L1} \leq k} \|\mathbf{M} - \mathbf{L}_t - \mathbf{S}\|_F^2 \end{cases} \quad (10)$$

Let the objective value $\|\mathbf{M} - \mathbf{L} - \mathbf{S}\|_F^2$ after solving the two subproblems in Equation (10) be $E_{t,1}$ and $E_{t,2}$ in the t^{th} iteration. Then, we have

$$E_{t,1} = \|\mathbf{M} - \mathbf{L}_t - \mathbf{S}_{t-1}\|_F^2, E_{t,2} = \|\mathbf{M} - \mathbf{L}_t - \mathbf{S}_t\|_F^2 \quad (11)$$

$$E_{t,2} = \|\mathbf{M} - \mathbf{L}_t - \mathbf{S}_t\|_F^2, E_{t+1,1} = \|\mathbf{M} - \mathbf{L}_{t+1} - \mathbf{S}_t\|_F^2 \quad (12)$$

The global optimality of $\mathbf{S}_t$ yields $E_{t,1} \geq E_{t,2}$ and the global optimality of $\mathbf{L}_{t+1}$ yields $E_{t,2} \geq E_{t+1,1}$. Therefore, the objective values $\|\mathbf{M} - \mathbf{L} - \mathbf{S}\|_F^2$ keep decreasing throughout Equation (10):

$$E_{1,1} \geq E_{1,2} \geq E_{2,1} \geq \cdots \geq E_{t,1} \geq E_{t,2} \geq E_{t+1,1} \geq \cdots \quad (13)$$

Since the objective of Equation (9) is monotonically decreasing, and the constraints are satisfied all the time, Equation (10) produces a sequence of objective values that converge to a local minimum.

Nevertheless, matrix $\mathbf{L}$ cannot be directly used to reconstruct the transient signal because the phase information is lost. The low-rank

complex spectrum matrix can be recovered by taking over the phase information of matrix $\mathbf{Y}$ denoted by $[\mathbf{X}] = \sqrt{[\mathbf{L}]}e^{j\angle[\mathbf{Y}]}$ and the sparse complex spectrum matrix is $[\mathbf{N}] = \sqrt{[\mathbf{S}]}e^{j\angle[\mathbf{Y}]}$. It is noteworthy to point out that the spectrum matrix $\mathbf{X}$ is often contaminated by noise. Therefore, a binary time-frequency mask method can be further used to refine the final results, which is written as follows:

$$\mathbf{F}(i,j) = \begin{cases} 1 & \text{if } |\mathbf{X}(i,j)| > G|\mathbf{N}(i,j)| \\ 0 & \text{if } |\mathbf{X}(i,j)| \leq G|\mathbf{N}(i,j)| \end{cases} \quad (14)$$

where G is a parameter to control the amplitude gain ratio between $|\mathbf{X}(i,j)|$ and $|\mathbf{N}(i,j)|$.

Finally, the estimated time-frequency spectrum of the signal can be transformed back to the time domain using the inverse transform. It is worth mentioning that the proposed method is very robust to STFT parameters and hyperparameter of SSGD according to the experimental verification. The transient signal extraction algorithm is shown in Algorithm 1. In addition, we give a guidance for parameter tuning in Appendix B.

Algorithm 1. The transient signal extraction algorithm.

1. Input measured signal $y(n)$ with N sampling length.
 2. Calculate the spectrogram matrix $\mathbf{M}$ using Equations (2), (3) and (4) with the Hanning window.
 3. Decompose spectrogram matrix $\mathbf{M}$ into the low-rank matrix $\mathbf{L}$ and the sparse matrix $\mathbf{S}$ using Equation (8).
 4. Recover low-rank complex matrix $[\mathbf{X}] = \sqrt{[\mathbf{L}]}e^{j\angle[\mathbf{Y}]}$.
 5. Filter out the noise in $\mathbf{X}$ using Equation (9) and $\mathbf{Q} = \mathbf{F} \odot \mathbf{X}$, where $\odot$ denotes Hadamard product.
 6. Extract the transient signal using inverse transform by Equation (5).
-

VALIDATION ON SYNTHETIC SIGNALS

In this section, we validate the proposed method with synthetic signals. We built the signal as a sequence of transients. It is monotonic and generated by $x = s \times h$. Here, we use “ $\times$ ” to denote convolution, and s is an impulse train, as:

$$s[n] = \sum_{m=0}^{N_m} \delta(n - mD_T) \text{ for } n = 1, \dots, N \quad (15)$$

where $\delta(n)$ denotes the Dirac impulse, and h is a filter with the transfer function

$$H(z) = \frac{1 - z^{-1}}{1 - 2 \cos(2\pi \cdot 0.2)rz^{-1} + r^2z^{-2}} \quad (16)$$

Firstly, an impulse signal $s[n]$ (time duration is 1 s and the sampling frequency is 10^4 Hz) with length $N = 10^4$ is produced, where $D_T = 600$ and $N_m = 16$. Secondly, the impulse signal $s[n]$ is passed through a filter $H(z)$, with $r = 0.95$ to synthesize the transient signal $x(n)$. Finally, Gaussian noise is added to the transient signal to obtain the simulated signal $y(n)$ while maintaining the SNR = -10 dB. A simple illustrative example is shown in

Figure 1, where the transient signal (in green color) is buried in strong additive noise (in blue color).

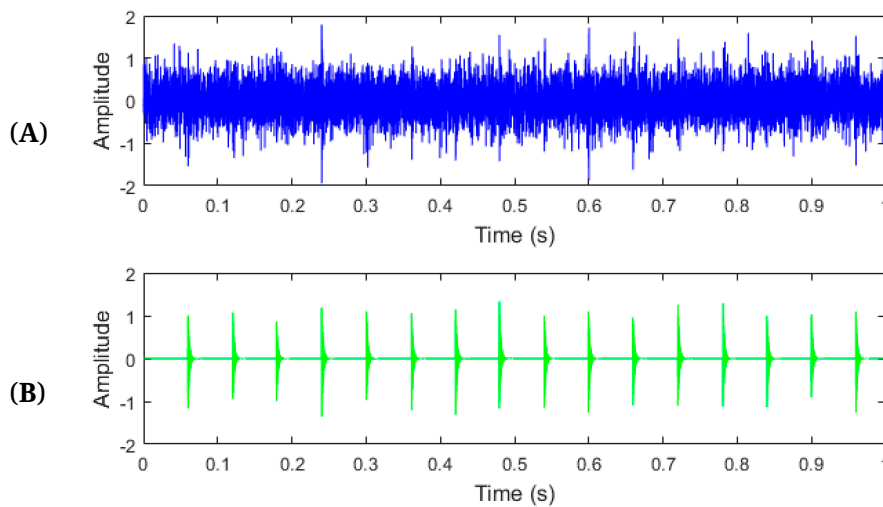


Figure 1. (A) A synthetic measured signal with white noise (SNR = -10 dB). (B) A synthetic periodic transient signal.

Then, a Hanning window with length $N_w = 50$ is selected for STFT-based SST, and the corresponding spectrogram of the raw signal and the transient signal are shown in Figure 2. It can be seen that the structure of the transient signal can be well captured after the transformation. On the contrary, the noise is spread over the entire time–frequency matrix. In other words, the low-rank of transient signal will not be actively changed after transform and can be considered as low-rank or semi-low-rank in the transformed domain. Therefore, the SSGD algorithm is used to decompose the spectrum of a raw signal into its low-rank and sparse components.

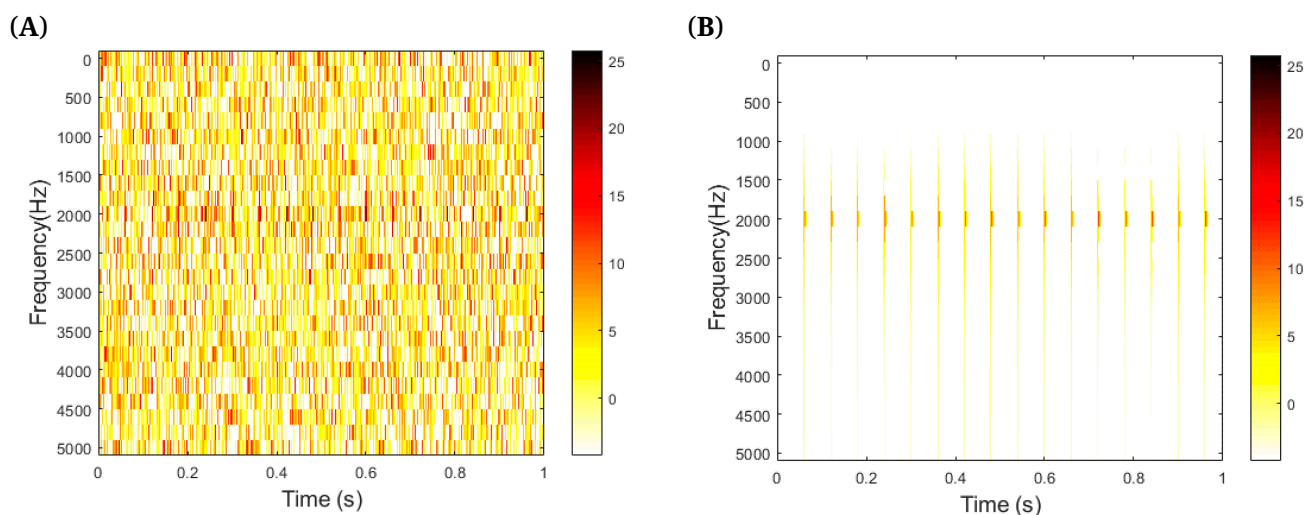


Figure 2. (A) Spectrogram of synthetic measured signal with a Hanning window with length $N_w = 50$ (SNR = -10 dB). (B) Spectrogram of the synthetic clean transient signal with Hanning window with length $N_w = 50$.

In this case, two important parameters in the SSGD is set to $r = 1$ and $\lambda = 0.02$. It is found that the decomposed low-rank and sparse spectrum can be attributed to the synthetic transient signal and the noise, respectively, as shown in Figure 3. However, there are differences between the extracted low-rank spectrum and spectrogram of the clean synthetic transient signal. These differences come from the fact that the transient signal in the time–frequency domain is not exactly low-rank, but it would be approximately low-rank. Therefore, the decomposed low-rank matrix can still be used to reconstruct transient signals.

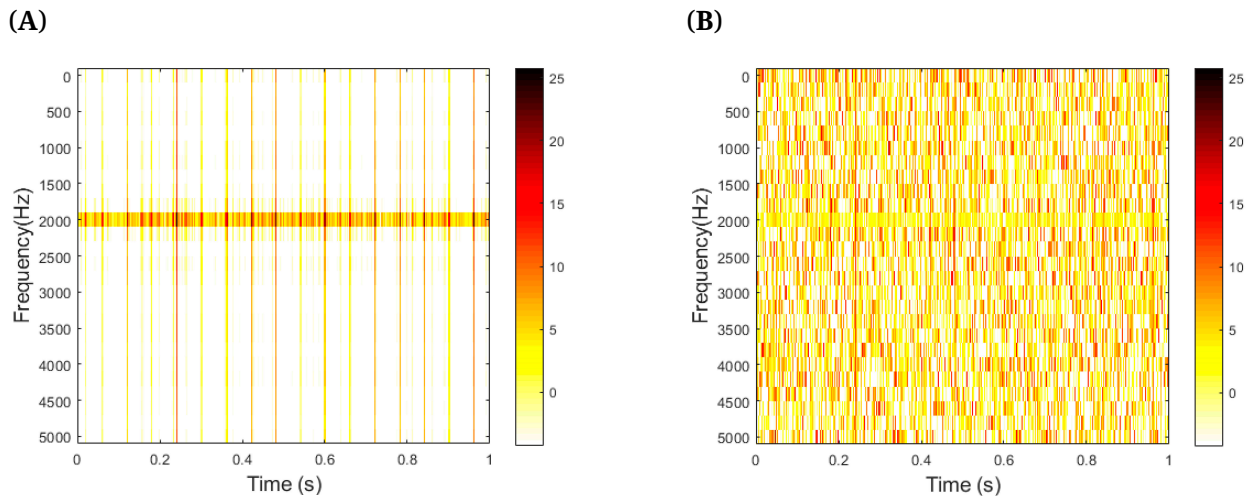


Figure 3. (A) Low-rank spectrum decomposed by SSGD algorithm with $r = 1$ and $\lambda = 0.02$. (B) Sparse spectrum decomposed by SSGD algorithm with $r = 1$ and $\lambda = 0.02$.

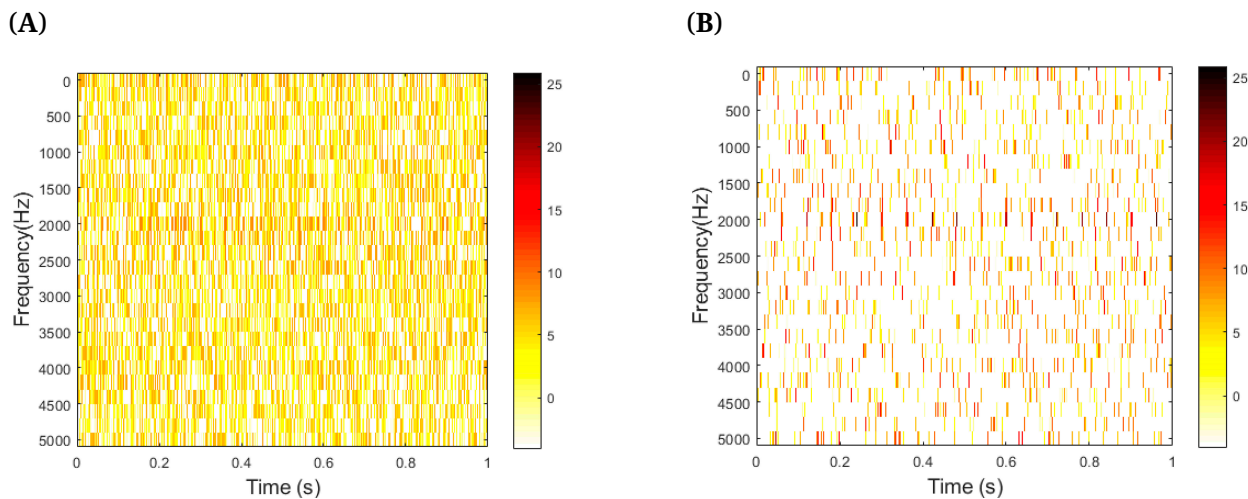


Figure 4. (A) Low-rank spectrum decomposed by RPCA algorithm with $\lambda = 0.02$. (B) Sparse spectrum decomposed by RPCA algorithm with $\lambda = 0.02$.

Similar to the previous work using RPCA for low-rank decomposition [30], here, RPCA is used to decompose the STFT-based SST spectrum and the results are presented in Figure 4. It is not easy to determine whether the transient signal is in the low-rank or the sparse component, compared with Figure 3. Further, the singular values of the decomposed matrices are

calculated to test the validity of low-rank and sparse decomposition algorithm. Distributions of singular values of the decomposed low-rank and sparse matrix by SSGD and RPCA are displayed in Figure 5. It can be found that the rank of the low-rank matrix obtained by RPCA is even larger than the sparse matrix. The reason for this may be the convex relaxation of RPCA. Due to its convexity relaxation, the decomposed results are insufficiently promoted for low-rank or sparsity. On the contrary, the low-rank matrix decomposed by SSGD satisfies its definition well just because the algorithm imposes a hard constraint on the rank. Besides, SSGD is significantly less time-consuming than RPCA.

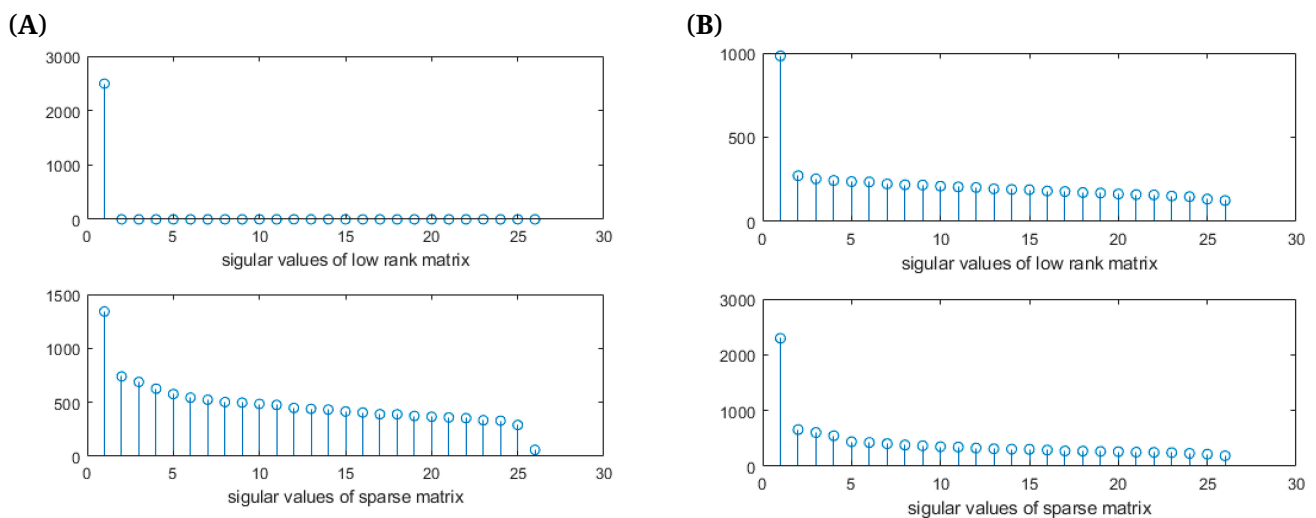


Figure 5. (A) Distribution of singular values of the decomposed low-rank and sparse matrix by SSGD. (B) Distribution of singular values the decomposed low-rank and sparse matrix by RPCA.

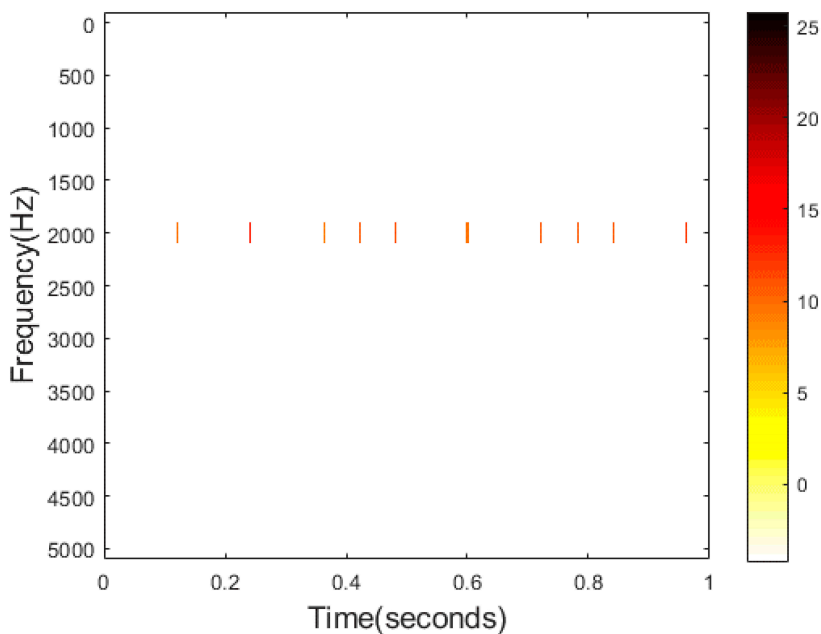


Figure 6. Filtered low-rank spectrum matrix $Q = F \odot X$ with $G = 6$.

Then, the phase information of the raw signal spectrum is used to recover the low-rank complex spectrum. A binary time-frequency mask is used to remove noise to obtain a more accurate low-rank spectrum. The filtered low-rank complex spectrum $Q = F \odot X$ is presented in Figure 6. Compared with Figure 2B, it is found that the result is good with the clean transient signal spectrum.

Finally, the estimated filtered low-rank spectrum matrix can be transformed back to the time domain using the inverse STFT-based SST, to obtain the transient signal, which is presented in Figure 7C. It can be seen that the periodic transient signals are almost entirely recovered, compared with Figure 7A,B, which proves the effectiveness of the proposed method.

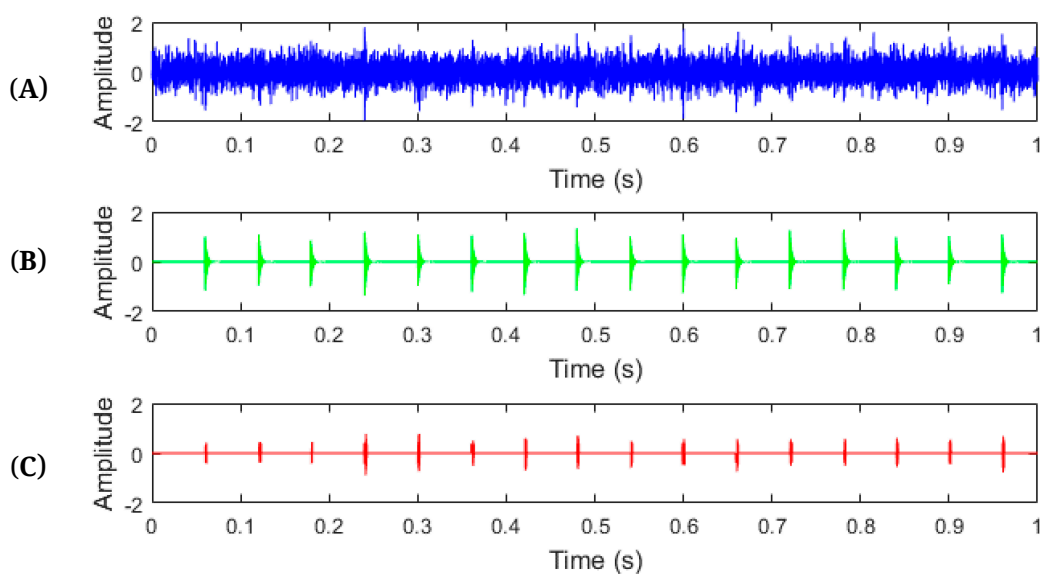


Figure 7. (A) A synthetic measured signal with white noise (SNR = -10 dB). (B) A synthetic periodic transient signal. (C) The extracted transient signal by our proposed method (Figure 6 shows that only 10 transients are recognized because the other values are too small to display).

In addition, there are many methods to recover sparse transients from noise. To emphasize the advantage of the low-dimensional modeling of fault signal, a Matching Pursuit (MP) based algorithm is used for comparison [31]. The crucial step of MP-based algorithms is the establishment of a sparsity-promoted dictionary, and the majority of known ways for establishing a dictionary can be divided into two general categories: analytic- and learning-based approaches [31]. An important issue is that such an analytic dictionary always needs to be established in advance, which inevitably causes the effect of extraction to be greatly affected by the dictionary. Often, in practical applications, it is very difficult to choose and build a dictionary properly.

Figure 8 shows the extraction of transient signals by the MP algorithm. It can be found that both methods can capture the time of transient signal well, but the transient signal extracted by the MP-based algorithm is still

mixed with some noise, and the proposed method performs well in noise reduction. In comparison, our method requires only a small number of parameters to achieve good results, and the method is robust to parameters.

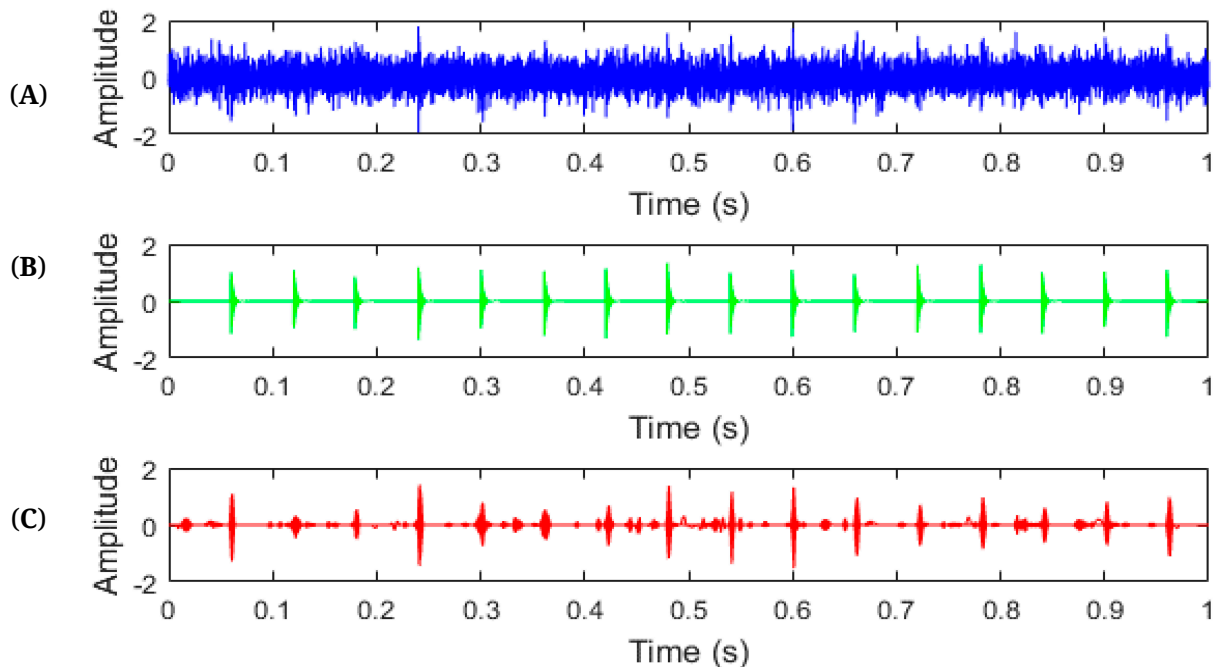


Figure 8. (A) A synthetic measured signal with white noise (SNR = -10 dB). (B) A synthetic periodic transient signal. (C) The extracted transient signal by MP-based method.

VALIDATION ON VIBRATION SIGNALS

This section illustrates the application of the proposed method on actual vibration signals, as often encountered in industrial applications. The rolling bearing experimental data are from the MFPT Challenge data [32], which comprises data from a bearing test rig (nominal bearing data, an outer race fault at various loads, and inner race fault and various loads). A fault is mainly identified by estimating the fault frequency and associating it to a given component in the machine. The prevailing method in the modern literature is surely the squared envelope spectrum, which reveals the repetition frequency of the transient signal series [33]. The following shows the application of the proposed method and envelope spectrum analysis in real vibration data.

As for inner race fault, the expected components in the envelope spectrum are ball pass frequency inner race (BPFI) and its harmonics, and sidebands spaced at shaft speed and their harmonics [5]. The envelope spectrum of the raw inner race fault shows that most of the energy is focused on BPFI and its harmonics, as shown in Figure 9. It indicates an inner race fault of the bearing, which matches the fault type of the signal. Therefore, the type of bearing fault can be detected by envelope spectrum analysis. Figure 10A indicates that the inner race fault signal has

significantly larger impulsiveness in the time domain, making envelope spectrum analysis capture the fault signature at BPFI effectively.

Then, the extracted transient signal by the proposed method is displayed in Figure 10B, which approximately captures the time position of the transient signal. Besides, the envelope spectrum of the extracted transient signal by the proposed method is presented in Figure 11. It is also found that most of the energy is focused on BPFI and its harmonics, which indicates the fault type of the signal. Moreover, the proposed method captures the sidebands in the low-frequency region better than the envelope spectrum directly on the raw signal. It also proves that the proposed algorithm is very effective in diagnosing fault types.

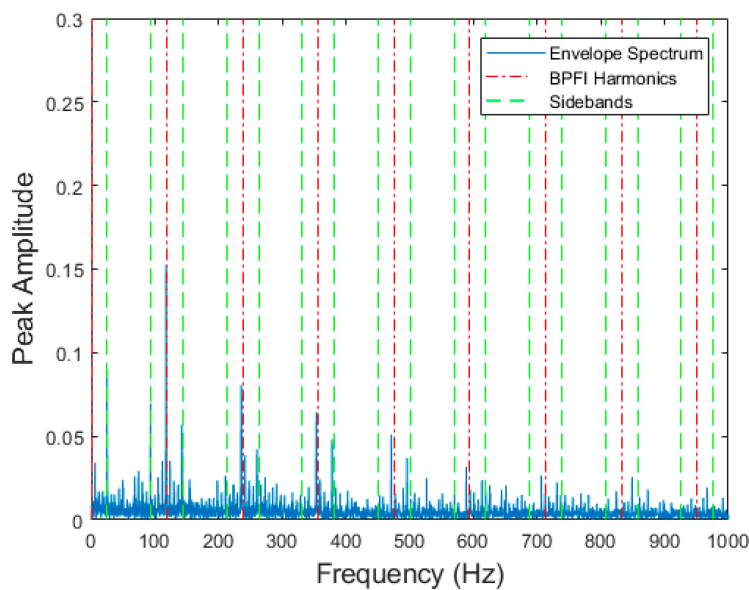


Figure 9. Envelope Spectrum of the raw inner race fault signal, the expected BPFI and its harmonics (red dashed line) and the expected sidebands (green dashed line).

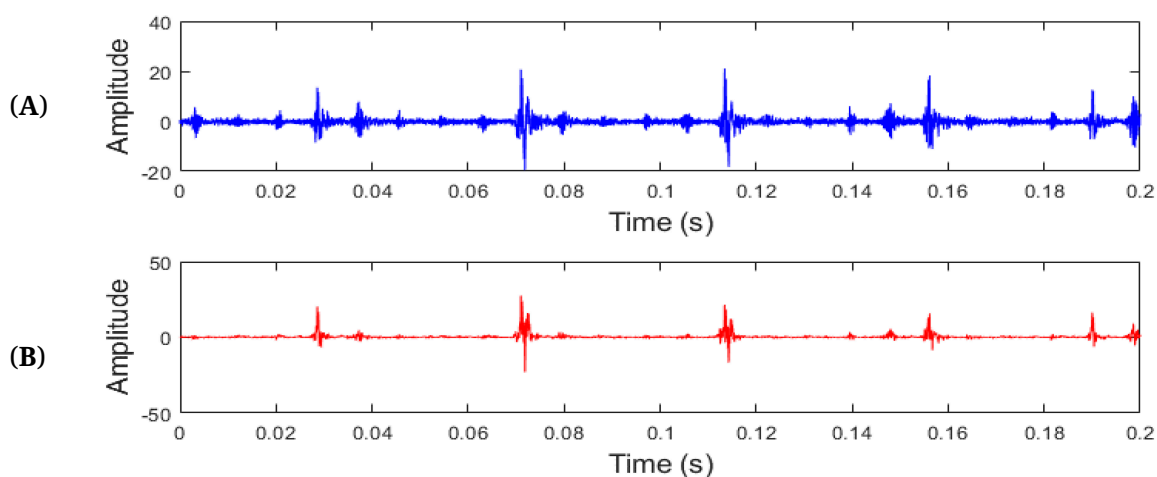


Figure 10. (A) The raw inner race fault signal data. (B) The extracted transient signal by the proposed method.

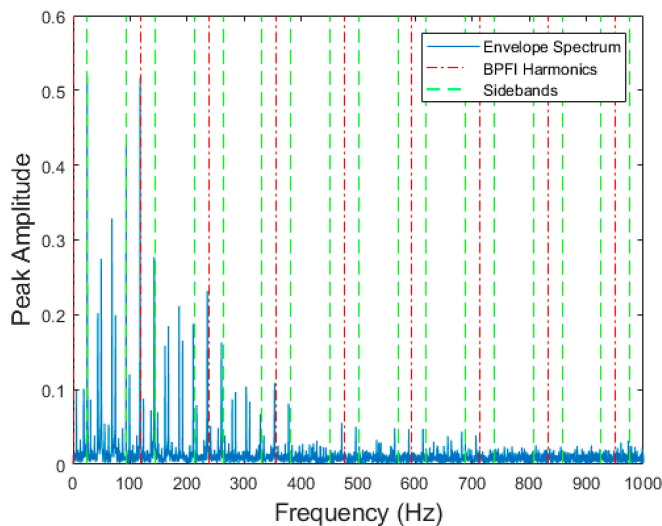


Figure 11. Envelope Spectrum of the extracted transient signal by the proposed method.

The expected components in the envelope spectrum are ball pass frequency outer race (BPFO) and harmonics of an outer race fault signal. Therefore, most of the energy in the envelope spectrum of the outer race signal should be concentrated at BPFO. However, it is found that there are no clear peaks at BPFO harmonics when the raw signal envelope spectrum is analyzed, as shown in Figure 12. It illustrates that envelope spectrum analysis on the raw signal does not provide useful diagnostic information. The outer race fault signal does not have a significantly larger impulsiveness, and it is masked by the loud noise, as shown in Figure 13A, making envelope spectrum analysis unable to capture the fault signature at BPFO effectively. Therefore, the raw signal needs to be preprocessed to extract the transient signal under strong signal before envelope spectrum analysis. Here, the proposed method is used for preprocessing, and the extracted transient signal is shown in Figure 13B.

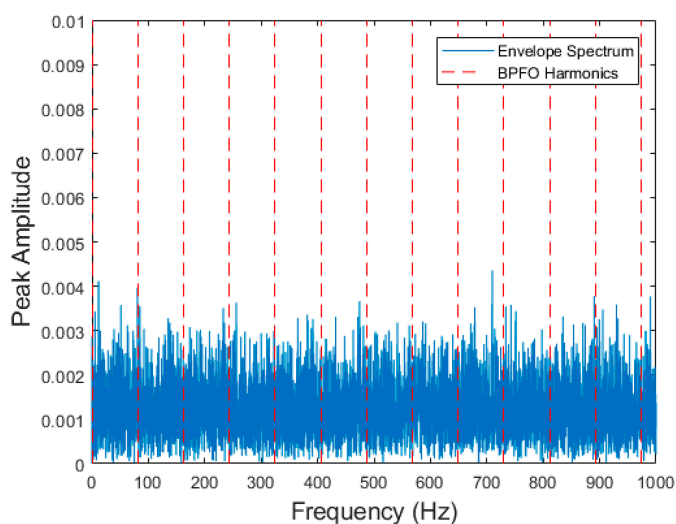


Figure 12. Envelope Spectrum of the raw outer race fault and the expected BPFI and its harmonics (red dashed line).

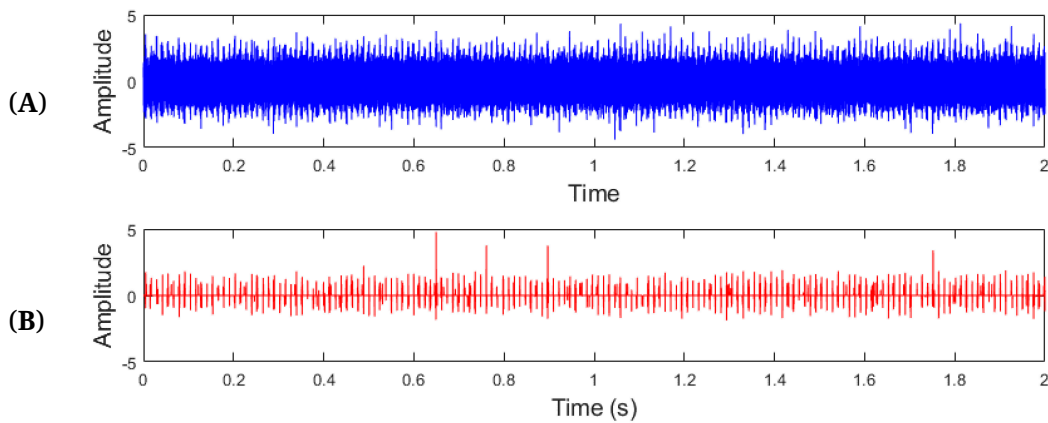


Figure 13. (A) Raw outer race fault signal. (B) The extracted transient signal by the proposed method.

Further, envelope spectrum analysis is carried out for the extracted transient signal, as shown in Figure 14A. It can be seen that there is a significant peak at BPFO, compared with Figure 12, which proves that the proposed preprocessing method is very useful for the extraction of fault information. Moreover, the envelope spectrum of the transient signal extracted by the proposed algorithm is compared with SK algorithm, as presented in Figure 14. It is found that the proposed method captures fault information well in high-frequency harmonics compared to the SK method. It further demonstrates the effectiveness of the proposed method in the diagnosis of vibration fault signals.

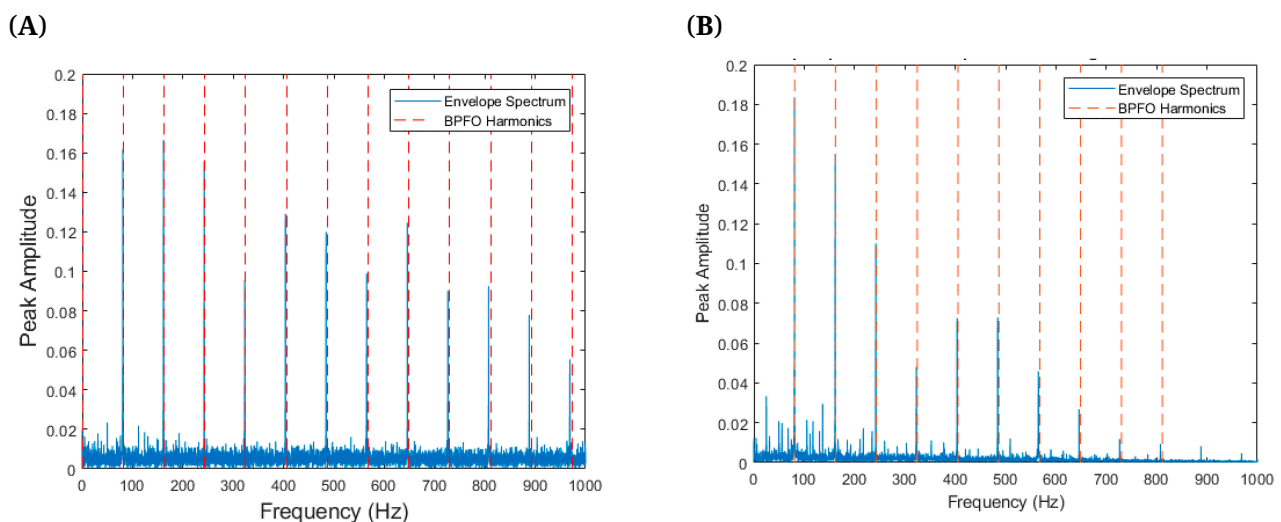


Figure 14. (A) Envelope Spectrum of the extracted transient signal by the proposed method. (B) Envelope Spectrum of the signal by SK algorithm.

CONCLUSIONS

This paper has introduced a low-dimensional model for the vibration signal of rolling element bearing faults, which has the typical form of a series of transients overlaid with stationary background noise. It is based on the sparse representation of the signal and the characteristics of

transient signals expressed in a sparse representation. In this work, STFT-based SST is utilized to transform the noisy measured signal to the high-resolution time–frequency domain. The low-rank of the transient signal will be preserved in the transformed domain. The filtered transient signal can be recovered through the inverse transformation of the low-rank component of the SST magnitude spectrum, which is obtained by solving the optimization problem using the Semi-Soft GoDec algorithm. This paper aims to exploit the low-dimensional model of periodic transient signals and establish a low rank and sparse model for fault detection of bearings. The performance of the proposed method has been tested on several synthetic signals and real vibration signals. In terms of reconstruction of the fault signal, results demonstrate that the method is effective and robust. The proposed model also deals with signals acquired from the actual vibration machine. The advantage of the proposed method is that it does not require prior training or other particular features and only requires the setting of a few parameters.

AUTHOR CONTRIBUTIONS

LY conceived the method and carried out the derivation procedure. WD wrote the manuscript and finished the simulation and experimental validation. SH validated the formulations and participated in drafting the manuscript. WJ coordinated the research. All authors gave final approval of the version to be submitted.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 11704248, 51835008, 11574212) and the Open Fund of the Key Laboratory of Aerodynamic Noise Control (ANCL20180302).

ACKNOWLEDGMENTS

The authors would like to thank the continuous support from Prof. Jerome Antoni and Prof. Quentin Leclerc (Laboratoire Vibrations Acoustique, INSA de Lyon).

APPENDICES

Appendix A

Fast low-rank approximation [24]:

Given r bilateral random projections (BRP) of an $m \times n$ dense matrix X , i.e., $Y_1 = XA_1$ and $Y_2 = X^T A_2$, wherein $A_1 \in \mathbb{R}^{m \times r}$ and $A_2 \in \mathbb{R}^{n \times r}$ are random matrices, $L = Y_1(A_2^T Y_1)^{-1} Y_2^T$ is a fast rank- r approximation of matrix X . When singular values of X decay slowly, L may perform poorly. A modification for this situation based on the power scheme is used. In the

power scheme modification, we instead calculate BRP of a matrix $\tilde{X} = (XX^T)^p X$, whose singular values decay faster than X . To obtain the approximation of X with r , we calculate the QR decomposition of Y_1 and Y_2 , i.e., $Y_1 = Q_1 R_1$, $Y_2 = Q_2 R_2$. Therefore, the low-rank approximation of X is $L = Q_1 [R_1 (A_2^T Y_1)^{-1} R_2^T]^{\frac{1}{2p+1}} Q_2^T$. In addition, the power scheme modification reduces the error of low-rank approximation by increasing p . In this paper, the values of p are all set to 6.

A soft thresholding operator for any matrix Q is defined as:

$$\text{Soft_threshold}(Q, \lambda) = \max(|q| - \lambda, 0) \text{sgn}(q)$$

where q denotes each entry of the matrix Q and $\text{sgn}()$ is the sign function.

Algorithm. Semi-Soft Go Decomposition.

Input: Spectrogram matrix X and parameter $r, \lambda, \varepsilon, p$

Output: low-rank matrix L and sparse matrix S

1. Initialize: $L_0 = X, S_0 = \mathbf{0}, t = 0$
 2. **while** $\frac{\|X - L_t - S_t\|_F}{\|X\|_F} > \varepsilon$ **do**
 3. $t = t + 1$
 4. $\tilde{L} = [(X - S_{t-1})(X - S_{t-1})^T]^p (X - S_{t-1})$
 5. $Y_1 = \tilde{L} A_1, A_2 = Y_1$
 6. $Y_2 = \tilde{L}^T Y_1 = Q_2 R_2, Y_1 = \tilde{L}^T Y_2 = Q_1 R_1$
 7. if $\text{rank}(A_2^T Y_1) < r$ then $r = \text{rank}(A_2^T Y_1)$, go to the step 3; end;
 8. $L_t = Q_1 [R_1 (A_2^T Y_1)^{-1} R_2^T]^{\frac{1}{2p+1}} Q_2^T$;
 9. $S_t = \text{Soft_threshold}(X - L_t, \lambda)$;
 10. **end while**
-

Appendix B

Parameter (r, λ, G) selection strategy:

1. Fix the rank $r = 1$.
2. Set a larger parameter value for λ and G and visualize the decomposed L and S matrices.
3. Alternately reduce the value of G until the low-rank of the L matrix and the sparse of the S matrix can be clearly seen.

REFERENCES

1. Randall RB, Antoni J. Rolling element bearing diagnostics—a tutorial. *Mech Syst Signal Process.* 2011;25(2):485-520. doi: 10.1016/j.ymssp.2010.07.017
2. Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. *Mech Syst Signal Process.* 2015;64-65:100-31. doi: 10.1016/j.ymssp.2015.04.021
3. Ge X, Hamzaoui N, Antoni J. Semi-automated diagnosis of bearing faults based on a hidden Markov model of the vibration signals. *Measurement.* 2018;127:141-66. doi: 10.1016/j.measurement.2018.05.040

4. Qi XX, Ji JW, Han XW. Fault Diagnosis Methods of Rolling Bearing: A General Review. *Key Engineer Mater.* 2011;480:986-92. doi: 10.4028/www.scientific.net/KEM.480-481.986
5. McFadden PD, Smith JD. Model for the vibration produced by a single point defect in a rolling element bearing. *J Sound Vib.* 1984;96(1):69-82. doi: 10.1016/0022-460x(84)90595-9
6. McDonald GL. Vibration Signal-Based Fault Detection for Rotating Machines [Master Thesis]. Edmonton (Canada): University of Alberta; 2011. doi: 10.7939/R3SK98
7. Cheng JS, Yu DJ, Yu Y. A fault diagnosis approach for roller bearings based on EMD method and AR model. *Mech Syst Signal Process.* 2006;20(2):350-62. doi: 10.1016/j.ymssp.2004.11.002
8. Wang WY, Wong AK. Autoregressive model-based gear fault diagnosis. *J Vib Acoust.* 2002;124:172-9. doi: 10.1115/1.1456905
9. Endo H, Randall RB. Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. *Mech Syst Signal Process.* 2007;21(2):906-19. doi: 10.1016/j.ymssp.2006.02.005
10. Feng ZP, Liang M, Chu FL. Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech Syst Signal Process.* 2013;38(1):165-205. doi: 10.1016/j.ymssp.2013.01.017
11. Rubini R, Meneghetti U. Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings. *Mech Syst Signal Process.* 2001;15(2):287-302. doi: 10.1006/mssp.2000.1330
12. Cocconcelli M, Zimroz R, Rubini R, Bartelmus W. STFT based approach for ball bearing fault detection in a varying speed motor. In: Fakhfakh T, Bartelmus W, Chaari F, Zimroz R, Haddar M, editors. *Condition Monitoring of Machinery in Non-Stationary Operations. Proceedings of the Second International Conference "Condition Monitoring of Machinery in Non-Stationary Operations" CMMNO'2012.* Berlin (Germany): Springer; 2012. p. 41-50. doi: 10.1007/978-3-642-28768-8_5
13. Daubechies I, Maes SH. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In: Aldroubi A, Unser MA, editors. *Wavelets in Medicine and Biology.* Boca Raton (US): CRC Press; 1996. p. 527-46.
14. Daubechies I, Lu J, Wu H-T. Synchrosqueezed Wavelet Transforms: an empirical mode decomposition-like tool. *Appl Comput Harmon Anal.* 2011;30(2):243-61. doi: 10.1016/j.acha.2010.08.002
15. Auger F, Flandrin P, Lin YT, McLaughlin S, Meignen S, Oberlin T, et al. Time-Frequency Reassignment and Synchrosqueezing: An Overview. *IEEE Signal Process Mag.* 2013;30(6):32-41. doi: 10.1109/MSP.2013.2265316
16. Antoni J, Randall RB. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mech Syst Signal Process.* 2004;20(2):308-31. doi: 10.1016/j.ymssp.2004.09.002
17. Wiggins RA. Minimum entropy deconvolution. *Geoeexploration.* 1978;16(1): 21-35. doi: 10.1016/0016-7142(78)90005-4

18. McDonald GL, Zhao Q, Zuo MJ. Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection. *Mech Syst Signal Process.* 2012;33:237-55. doi: 10.1016/j.ymssp.2012.06.010
19. Du ZH, Chen XF, Zhang H, Yang B, Zhai Z, Yan R. Weighted low-rank sparse model via nuclear norm minimization for bearing fault detection. *J Sound Vib.* 2017;400:270-87. doi: 10.1016/j.jsv.2017.03.044
20. Beckouche S, Ma JW. Simultaneous dictionary learning and denoising for seismic data. *Geophysics.* 2014;79(3):A27-31. doi: 10.1190/geo2013-0382.1
21. Gao HZ, Liang L, Chen XG, Xu GH. Feature Extraction and Recognition for Rolling Element Bearing Fault Utilizing Short-Time Fourier Transform and Non-negative Matrix Factorization. *Chin J Mech Eng.* 2015;28(1):96-105. doi: 10.3901/CJME.2014.1103.166
22. Candes EJ, Li XD, Ma Y, Wright J. Robust principal component analysis. *J ACM.* 2011;58(3):11. doi: 10.1145/1970392.1970395
23. Tao M, Yuan X. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J Optim.* 2011;21:57-81. doi: 10.1137/10078189
24. Zhou T, Tao D. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In: *Proceedings of the 28th International Conference on Machine Learning*. Bellevue (US): Omnipress; 2011. p. 33-40.
25. Ma JW. Three-dimensional irregular seismic data reconstruction via low-rank matrix completion. *Geophysics.* 2013;78(5):181-192. doi: 10.1190/geo2012-0465.1
26. Thakur G, Wu HT. Synchrosqueezing-based Recovery of Instantaneous Frequency from Nonuniform Samples. *SIAM J Math Anal.* 2011;43:2078-95. doi: 10.1137/100798818
27. Oberlin T, Meignen S, Perrier V. The Fourier-based Synchrosqueezing Transform. In: *Proceedings of the 2014 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*; 2014 May 4-9; Florence, Italy. Piscataway (US): IEEE; 2014. p. 315-9. doi: 10.1109/ICASSP.2014.6853609
28. Zhou TY, Tao DC. Bilateral random projection. In: *2012 IEEE International Symposium on Information Theory Proceedings*; 2012 Jul 1-6; Cambridge, MA, USA. Piscataway (US): IEEE; 2012. doi: 10.1109/ISIT.2012.6283064
29. Fazel M, Candes E, Recht B, Parrilo P. Compressed sensing and robust recovery of low-rank matrices. In: *2008 42nd Asilomar Conference on Signals, Systems and Computers*; 2008 Oct 26-29; Pacific Grove, CA, USA. Piscataway (US): IEEE; 2009. p. 1043-47. doi: 10.1109/ACSSC.2008.5074571
30. Yu L, Antoni J, Leclerc Q. Combined Regularization Optimization for Separating Transient Signal from Strong Noise in Rolling Element Bearing Diagnostics. In: *Surveillance 7*; 2013 Oct; Chartres, France. HAL ID: hal-00958622
31. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process.* 2006;54(11):4311-22. doi: 10.1109/TSP.2006.881199

32. Condition Based Maintenance Fault Database for Testing Diagnostics and Prognostic Algorithms. Available from: <https://mfpt.org/fault-data-sets/>. Accessed 2018 Nov 10.
33. Mcfadden PD, Smith JD. Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review. *Tribology Int.* 1984;17(1):3-10. doi: 10.1016/0301-679X(84)90076-8

How to cite this article:

Yu L, Dai W, Huang S, Jiang W. Sparse Time–Frequency Representation for the Transient Signal Based on Low-Rank and Sparse Decomposition. *J Acoust.* 2019;1:e190003. <https://doi.org/10.20900/joa20190003>