Journal of Sound and Vibration 432 (2018) 213–234

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

# Sparsity-enhanced signal decomposition via generalized minimax-concave penalty for gearbox fault diagnosis

Gaigai Cai [a,b], Ivan W. Selesnick [b], Shibin Wang [b,c,*], Weiwei Dai [b], Zhongkui Zhu [d]

[a] Key Laboratory of Ministry of Education for Electronic Equipment Structure Design, Xidian University, Xi'an, 710071, PR China
[b] Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, NY 11201, USA
[c] State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China
[d] School of Rail Transportation, Soochow University, Suzhou, Jiangsu 215137, PR China

ARTICLE INFO

*Article history:*
Received 2 December 2017
Revised 6 May 2018
Accepted 15 June 2018
Available online XXX
Handling Editor: K. Shin

*Keywords:*
Gearbox fault diagnosis
Signal decomposition
Sparse representation
Generalized minimax-concave penalty
Convex optimization

ABSTRACT

Vibration signals arising from faulty gearboxes are often a mixture of the meshing component and the periodic transient component, and simultaneously contaminated by noise. Sparsity-assisted signal decomposition is an effective technique to decompose a signal into morphologically distinct components based on sparse representation and optimization. In this paper, we propose a sparsity-enhanced signal decomposition method which uses the generalized minimax-concave (GMC) penalty as a nonconvex regularizer to enhance sparsity in the sparse approximation compared to classical sparsity-assisted signal decomposition methods, and thus to improve the decomposition accuracy for gearbox fault diagnosis. Even though the GMC penalty itself is nonconvex, it maintains the convexity of the GMC regularized cost function to be minimized. Hence, similar to the classical L1-norm regularization methods, the global optimal solution can be guaranteed via convex optimization. Moreover, we present and validate a straight-forward way to choose transforms and set parameters for the proposed method. Through simulation studies, it is demonstrated that the proposed sparsity-enhanced signal decomposition method can effectively decompose the simulated faulty gearbox signal into the meshing component and the periodic transient component. Comparisons with the classical L1-norm regularized signal decomposition method and spectral kurtosis show that the proposed method can accurately preserve the amplitude of the periodic transient component and provide a more accurate estimation result. Experiment and engineering case studies further verify that the proposed method can accurately estimate the periodic transient component from vibration signals, which demonstrate that the proposed method is a promising tool for gearbox fault diagnosis.



## 1. Introduction

Gearboxes are used extensively for the transmission of speed and power in mechanical systems, such as wind turbines, aero-engines, helicopters, automobiles, and mining equipment. However, gearboxes are extremely prone to various faults when operated continuously in a harsh working environment. If a fault occurs in the key gear during its operation, it may cause an abnormal operation or system failure, which results in long downtimes, increased maintenance loss, or even casualties. For example, gearbox failures are regarded as one of the most serious causes of breakdown in wind turbines, and the reliability of

* Corresponding author. State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China.
*E-mail addresses:* wangshibin2008@gmail.com, wangshibin2008@mail.xjtu.edu.cn (S.Wang).

https://doi.org/10.1016/j.jsv.2018.06.037
0022-460X/

gearboxes is an important concern in the wind energy industry [1–3]. Therefore, condition monitoring and fault diagnosis of gearboxes have attracted considerable attention during the past decades [4].

Vibration signals generated by mechanical faults are important indicators of machinery operating condition, and contain information not only about the machine health condition but also the severity of the fault [5,6]. For a fault-free gear pair, the vibration signal of the gearbox tends to be dominated by the meshing frequency and its harmonics [7]. When there exists a localized fault on the gear, such as spall, pit or breakage, it will interact with another gear and thus generate a series of transients which are periodic at the inverse of the shaft rotational frequency [8]. Therefore, the vibration signal arising from a faulty gearbox system is usually a mixture of the meshing component and the periodic transient component, and these two components are polluted by heavy background noise from other machine components and the working environment [9]. How to accurately estimate the periodic transients corresponding to the fault from the complex noisy signal is a key task for gearbox fault diagnosis.

Over the past decades, various signal processing methods have been widely studied and applied to vibration signal analysis and fault diagnosis, such as short-time Fourier transform (STFT) [10], wavelet transform [11–13], empirical mode decomposition [14–16], spectral kurtosis (SK) [17–20], stochastic resonance [21–23] and time-frequency analysis [24–29]. These contributions have greatly enriched researches in machinery fault diagnosis. However, the effectiveness of most above-mentioned methods is diminished by two inevitable limitations. First, most of these methods reduce the amplitude of the component of interest while eliminating the noise and interference. If the signal of interest is preserved, the noise or interference is also preserved simultaneously. Second, most of the aforementioned methods analyze signals in the frequency domain or in the time-frequency domain. Vibration signals collected from faulty gearboxes are a combination of the meshing component, the periodic transient component, and noise. However, the meshing component and the periodic transient component not only overlap in the time domain but they may also overlap in the frequency domain. Hence, these two mixed components are difficult to decompose by these methods.

As a new branch of the signal processing method, sparse representation has received considerable attentions and has been widely used in the field of machinery fault diagnosis [30–35]. Sparsity-assisted signal decomposition using morphological component analysis (MCA) has proven to be a useful technique for machine fault signal decomposition. MCA is implemented via sparse representations and optimization and it is designed to separate additively-mixed components in a signal using morphological diversity rather than frequency or scale information [36]. It was first used for the separation of texture and piecewise smooth components in an image, and later widely used in a variety of fields. Abrial et al. extended MCA to the analysis of spherical data map [37]. Yong et al. generalized MCA to separate multi-channel EEG signal sources [38]. Farshchian et al. used MCA with two STFTs to separate the wing-beat signature from the body signature of radar bird data [39]. Considering that many complex signals arising from physiological and physical processes are often a mixture of components with different oscillation behaviors, Selesnick proposed a sparsity-assisted signal decomposition method using MCA and wavelet transform to decompose a signal into its components [40]. Cai et al. introduced the sparsity-assisted signal decomposition method using MCA to gearbox fault diagnosis in which the faulty gearbox vibration signal is separated into the meshing component and the periodic transient component [41]. The effectiveness of the sparsity-assisted signal decomposition method using MCA in faulty gearbox vibration signal decomposition is based on the fact that the meshing component and the periodic transient component possess different morphologies. Subsequently, other researchers began to study the usage of sparsity-assisted signal decomposition via MCA for gearbox fault diagnosis. Li et al. introduced the kernel spectral regression framework into MCA to diagnose the marine propulsion gearbox fault [42]. Yu et al. employed the principle of the minimum entropy of information to select the optimal dictionary for the transient component representation and proposed an improved MCA for signal decomposition to diagnose compound faults in gearboxes [43]. Zhang et al. proposed a resonance-based sparse signal decomposition with a comb filter method using MCA for gearbox multi-fault diagnosis [44]. These studies provide new insights on how to extract fault features for gearbox fault diagnosis. In these studies, the L1-norm is classically used to regularize the MCA problem because the L1-norm induces sparsity most effectively among convex penalties [45]. However, the L1-norm regularizer often underestimates the high-amplitude components of the sparse coefficients and tends to underestimate the signal of interest while eliminating noise and interference signal [46], which may cause missed alarm or underestimation of the fault severity.

In recent years, researches have addressed the design of penalties to strongly promote sparsity. Nonconvex sparsity-inducing penalties can provide a more accurate estimation of the high-amplitude component. However, most nonconvex penalties do not seek to maintain the convexity of the cost function to be minimized. Hence, the cost function is generally nonconvex and has extraneous suboptimal local minimizers. Blake, Zisserman and Nikolova designed convexity-preserving nonconvex penalties, and these penalties were further developed [47–50]. However, these penalties are separable, so when they are parameterized to maintain cost function convexity, they can only improve on the L1-norm penalty to a very limited extent. Recently, Selesnick proposed a novel class of nonconvex penalty functions, named the generalized minimax-concave (GMC) penalty, which can induce sparsity effectively while maintaining the cost function convexity [51]. Theoretical analysis proves that it is easy to prescribe the GMC penalty to maintain the convexity of the cost function to be minimized. Hence, the cost function has no suboptimal local minimizers and can be minimized via convex optimization comprising simple computations. The superiority of the GMC penalty is that it can effectively enhance the sparsity level in a sparse approximation problem and achieve a more accurate estimation of the signal of interest.

In this paper, we propose a sparsity-enhanced signal decomposition method for gearbox fault diagnosis by using the GMC penalty to improve the signal decomposition performance of the L1-norm regularized MCA. The paper proves that the convexity of the sparsity-enhanced MCA cost function can be maintained by prescribing the GMC penalty appropriately, and provides an

iterative algorithm to solve the convex sparsity-enhanced MCA problem for signal decomposition. This paper further proposes this method for decomposing gearbox vibration signals into the meshing component and the periodic transient component, and presents a straight-forward way to choose transforms and set parameters for the implementation of the proposed sparsity-enhanced signal decomposition method. A simulation study verifies the effectiveness of the proposed method and shows that the signal decomposition accuracy is improved. Experiment and engineering case studies further verify the effectiveness of the proposed method for signal decomposition as well as the transform choose and parameter setting strategy. Comparisons show that the proposed method can more accurately estimate the amplitude of the periodic transient component and thus can significantly improve the gearbox fault diagnosis accuracy compared with the L1-norm regularized MCA method or SK.

The main contributions of this paper are in two categories. First, we propose a sparsity-enhanced signal decomposition method via using the GMC penalty to regularize the MCA problem, give a condition to guarantee the convexity of the sparsity-enhanced MCA cost function, and derive a practical iterative algorithm based on the forward-backward splitting (FBS) algorithm to obtain the global minimum. Second, through simulation, an experiment case study and an engineering case study, we verify the effectiveness and superiority of the proposed sparsity-enhanced signal decomposition method for gearbox fault diagnosis. We also present and validate a straight-forward way to choose transforms and set parameters for the proposed method.

The remainder of this paper is organized as follows. Section 2 recalls the L1-norm regularized MCA for sparsity-assisted signal decomposition. In Section 3, we propose the sparsity-enhanced signal decomposition method using the GMC penalty, show how to maintain the convexity of the MCA cost function, and present a proximal algorithm to minimize the cost function. In Section 4, we provide a procedure to apply the sparsity-enhanced signal decomposition method to gearbox fault diagnosis. Section 5 validates the effectiveness and superiority of the proposed method for gearbox fault diagnosis by simulated signal analysis, and investigates the transform choice and parameter setting strategy. Section 6 applies the method to gearbox fault diagnosis (for an automobile transmission gearbox and the reduction gearbox of a hot-strip finishing milling machine) and compares its performance to the L1-norm regularized MCA and SK. Section 7 draws conclusions.

## 2. Review of sparsity-assisted signal decomposition using MCA

The vibration signal $\mathbf{y} \in \mathbb{R}^M$ measured from a faulty gearbox is modeled as

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 + \mathbf{n} \tag{1}$$

where $\mathbf{y}_1 \in \mathbb{R}^M$ comprises the meshing component and its harmonics, $\mathbf{y}_2 \in \mathbb{R}^M$ comprises fault-induced periodically occurring transients, and $\mathbf{n} \in \mathbb{R}^M$ is noise. The meshing component and the periodic transient component have different morphologies: the former exhibits a high oscillatory behavior and the latter exhibits a low oscillatory behavior [41].

For the signal in Eq. (1), the sparsity-assisted signal decomposition using MCA assumes that there are two specific transforms so that the component $\mathbf{y}_1$ can be sparsely represented by one transform but cannot be sparsely represented (or at least significant less sparsely) by the other transform, and vice versa. That is,

$$\mathbf{y}_1 = \mathbf{A}_1\mathbf{x}_1, \quad \mathbf{y}_2 = \mathbf{A}_2\mathbf{x}_2, \tag{2}$$

where the matrix $\mathbf{A}_1 \in \mathbb{R}^{M\times N_1}$ and $\mathbf{A}_2 \in \mathbb{R}^{M\times N_2}$ (with $N_i \geqslant M, i = 1, 2$) correspond to the two specific transforms, and $\mathbf{x}_1 \in \mathbb{R}^{N_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{N_2}$ are sparse coefficients vectors. The goal of MCA is to recover $\mathbf{y}_1$ and $\mathbf{y}_2$ from $\mathbf{y}$. Under the MCA framework with the classical L1-norm penalty, the MCA regularizer is

$$\phi(\mathbf{x}_1, \mathbf{x}_2; \lambda_1, \lambda_2) = \lambda_1\|\mathbf{x}_1\|_1 + \lambda_2\|\mathbf{x}_2\|_1 \tag{3}$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the regularization parameters. Then the components $\mathbf{y}_1$ and $\mathbf{y}_2$ can be estimated by solving the following unconstrained optimization problem,

$$(\mathbf{x}_1^{\text{opt}}, \mathbf{x}_2^{\text{opt}}) = \arg\min_{\mathbf{x}_1,\mathbf{x}_2} \left\{ F(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}_1\mathbf{x}_1 - \mathbf{A}_2\mathbf{x}_2\|_2^2 + \lambda_1\|\mathbf{x}_1\|_1 + \lambda_2\|\mathbf{x}_2\|_1 \right\} \tag{4}$$

where $F$ is the cost function to be minimized.

Convex optimization algorithms, such as split augmented Lagrangian shrinkage algorithm (SALSA) [40], alternating direction method of multipliers (ADMM) [52], majorization-minimization (MM) [53], and FBS [54], can be used to solve the problem in Eq. (4). Then the two additively-mixed components $\mathbf{y}_1$ and $\mathbf{y}_2$ in the signal can be decomposed: the estimated meshing component is $\widehat{\mathbf{y}}_1 = \mathbf{A}_1\mathbf{x}_1^{\text{opt}}$; the estimated periodic transient component is $\widehat{\mathbf{y}}_2 = \mathbf{A}_2\mathbf{x}_2^{\text{opt}}$.

Eq. (4) shows the most common case of MCA using L1-norm as the penalty. However, L1-norm tends to underestimate the true value of the signal [46]. Hence, the estimated components are also underestimated. To overcome this deficiency, we introduce the newly developed GMC penalty to regularize the MCA least squares problem to enhance the signal decomposition accuracy of MCA in the next section.

## 3. GMC penalty and sparsity-enhanced signal decomposition method

The GMC penalty is a nonconvex regularizer designed to maintain the convexity of sparse-regularized linear least squares problems. It is derived based on a multivariate generalization of the Huber function. In this section, we firstly recall the definition of the GMC penalty. Then we propose a sparsity-enhanced signal decomposition method by using the GMC penalty to penalize

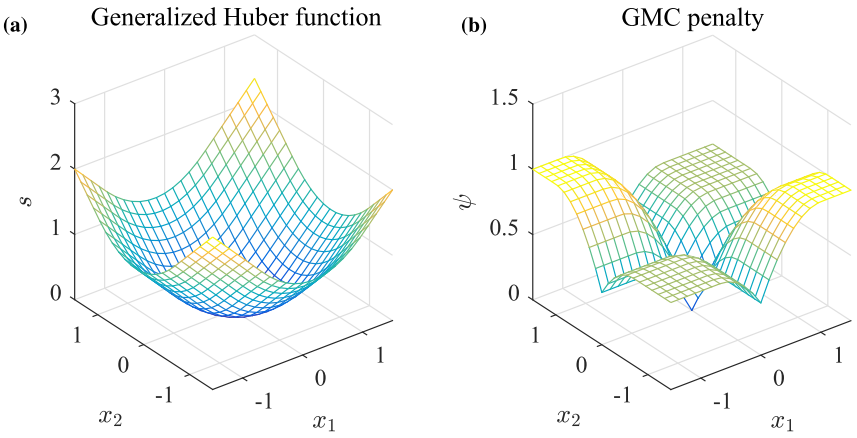


**Fig. 1.** (a) The generalized Huber function and (b) the GMC penalty.

the MCA problem. We also derive a condition to ensure that the GMC penalty maintains the total convexity of the sparsity-enhanced MCA cost function. Finally, we provide an iterative FBS algorithm to solve the sparsity-enhanced MCA problem for signal decomposition.

### 3.1. GMC penalty

The GMC penalty is prescribed via the generalized Huber function, which is actually a multivariate generalization of the scaled Huber function. Detailed information about the scaled Huber function can refer to reference [55]. Given a matrix $\mathbf{B} \in \mathbb{R}^{M\times N}$, the generalized Huber function $S : \mathbb{R}^N \to \mathbb{R}$ is defined as

$$S(\mathbf{x};\lambda,\mathbf{B}) = \min_{\mathbf{v}} \left\{ \lambda\|\mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{B}(\mathbf{x}-\mathbf{v})\|_2^2 \right\}, \tag{5}$$

which itself is an L1-norm regularized optimization problem. The GMC penalty function $\psi : \mathbb{R}^N \to \mathbb{R}$ is defined as

$$\psi(\mathbf{x}) = \lambda\|\mathbf{x}\|_1 - S(\mathbf{x};\lambda,\mathbf{B}) = \lambda\|\mathbf{x}\|_1 - \min_{\mathbf{v}} \left\{ \lambda\|\mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{B}(\mathbf{x}-\mathbf{v})\|_2^2 \right\}. \tag{6}$$

To illustrate the property of the generalized Huber function $S(\mathbf{x};\lambda,\mathbf{B})$ and the GMC penalty $\psi(\mathbf{x};\lambda,\mathbf{B})$, we set matrix $\mathbf{B}$ as

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 0 & 1 \end{bmatrix} \tag{7}$$

and $\lambda = 1$. Fig. 1(a) shows the generalized Huber function, and Fig. 1(b) shows the GMC penalty. It can be observed from Fig. 1(b) that the GMC penalty itself is nonconvex, and the GMC penalty approximates the L1-norm penalty around zero. Moreover, the GMC penalty also has the regularization property: large values are penalized more than (or the same as) small values.

### 3.2. Sparsity-enhanced signal decomposition using GMC penalty

Now we consider how to use the GMC penalty to regularize the MCA problem and thus to improve its performance for signal decomposition. To accurately estimate the components in the compound signal, a sparsity-enhanced signal decomposition method is proposed by using the GMC penalty to regularize the MCA problem.

First, the GMC penalty for signal decomposition is defined as

$$\psi(\mathbf{x}_1,\mathbf{x}_2;\lambda_1,\lambda_2,\mathbf{B}_1,\mathbf{B}_2) = \lambda_1\|\mathbf{x}_1\|_1 + \lambda_2\|\mathbf{x}_2\|_1 - \min_{\mathbf{v}_1,\mathbf{v}_2} \left\{ \lambda_1\|\mathbf{v}_1\|_1 + \lambda_2\|\mathbf{v}_2\|_1 + \frac{1}{2}\|\mathbf{B}_1(\mathbf{x}_1-\mathbf{v}_1) + \mathbf{B}_2(\mathbf{x}_2-\mathbf{v}_2)\|_2^2 \right\}, \tag{8}$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$. Then the GMC penalty is used to regularize the MCA problem. The cost function is

$$F(\mathbf{x}_1,\mathbf{x}_2) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}_1\mathbf{x}_1 - \mathbf{A}_2\mathbf{x}_2\|_2^2 + \psi(\mathbf{x}_1,\mathbf{x}_2;\lambda_1,\lambda_2,\mathbf{B}_1,\mathbf{B}_2). \tag{9}$$

The sparse MCA solution is obtained via minimizing of the cost function in Eq. (9), that is,

$$\begin{aligned}(\mathbf{x}_1^{\text{opt}}, \mathbf{x}_2^{\text{opt}}) &= \arg\min_{\mathbf{x}_1,\mathbf{x}_2} \{F(\mathbf{x}_1, \mathbf{x}_2)\} \\ &= \arg\min_{\mathbf{x}_1,\mathbf{x}_2} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}_1\mathbf{x}_1 - \mathbf{A}_2\mathbf{x}_2\|_2^2 + \lambda_1\|\mathbf{x}_1\|_1 + \lambda_2\|\mathbf{x}_2\|_1 \right. \\ &\quad \left. - \min_{\mathbf{v}_1,\mathbf{v}_2} \left\{ \lambda_1\|\mathbf{v}_1\|_1 + \lambda_2\|\mathbf{v}_2\|_1 + \frac{1}{2}\|\mathbf{B}_1(\mathbf{x}_1 - \mathbf{v}_1) + \mathbf{B}_2(\mathbf{x}_2 - \mathbf{v}_2)\|_2^2 \right\} \right\}. \end{aligned} \tag{10}$$

Next, we consider how to prescribe the GMC penalty to maintain the convexity of the cost function in Eq. (9). During the derivation of the convexity condition, the following notations are used to simplify the expression of the cost function,

$$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2], \mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2], \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}, \text{ and } \lambda = \begin{bmatrix} \lambda_1 \mathbf{1}_{N_1} \\ \lambda_2 \mathbf{1}_{N_2} \end{bmatrix}, \tag{11}$$

where $\mathbf{1}_N = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$ is the $N$-dimension vector of ones. Then Eq. (9) is rewritten as

$$F(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\lambda \odot \mathbf{x}\|_1 - \min_{\mathbf{v}} \left\{ \|\lambda \odot \mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{B}(\mathbf{x} - \mathbf{v})\|_2^2 \right\} \tag{12}$$

where $\odot$ denotes the element-wise multiplication of equal-size vectors. Next, Eq. (12) is written as

$$\begin{aligned} F(\mathbf{x}) &= \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\lambda \odot \mathbf{x}\|_1 - \min_{\mathbf{v}} \left\{ \|\lambda \odot \mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{B}(\mathbf{x} - \mathbf{v})\|_2^2 \right\} \\ &= \max_{\mathbf{v}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\lambda \odot \mathbf{x}\|_1 - \|\lambda \odot \mathbf{v}\|_1 - \frac{1}{2}\|\mathbf{B}(\mathbf{x} - \mathbf{v})\|_2^2 \right\} \\ &= \max_{\mathbf{v}} \left\{ \frac{1}{2}\mathbf{x}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{A} - \mathbf{B}^{\mathrm{T}}\mathbf{B})\mathbf{x} + \|\lambda \odot \mathbf{x}\|_1 + \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y} - \|\lambda \odot \mathbf{v}\|_1 + (\mathbf{v}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B} - \mathbf{y}^{\mathrm{T}}\mathbf{A})\mathbf{x} - \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{v} \right\} \\ &= \frac{1}{2}\mathbf{x}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{A} - \mathbf{B}^{\mathrm{T}}\mathbf{B})\mathbf{x} + \|\lambda \odot \mathbf{x}\|_1 + \max_{\mathbf{v}} \{g(\mathbf{x}, \mathbf{v})\} \end{aligned}$$

where $g(\mathbf{x}, \mathbf{v}) = \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y} - \|\lambda \odot \mathbf{v}\|_1 + (\mathbf{v}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B} - \mathbf{y}^{\mathrm{T}}\mathbf{A})\mathbf{x} - \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{v}$ is affine in $\mathbf{x}$. The function $\max_{\mathbf{v}} g(\mathbf{x}, \mathbf{v})$ is convex in $\mathbf{x}$ as it is the pointwise maximum of a set of convex function in $\mathbf{x}$ (where the set is indexed by $\mathbf{v}$). Hence, the condition for $F(\mathbf{x})$ to be convex is

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} - \mathbf{B}^{\mathrm{T}}\mathbf{B} \succeq 0 \tag{13}$$

i.e. the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A} - \mathbf{B}^{\mathrm{T}}\mathbf{B}$ is positive semi-definite. Hence, during the application, for a given $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2]$, we can simply set

$$\mathbf{B} = \sqrt{\gamma}\mathbf{A}, \text{ i.e., } \mathbf{B}_1 = \sqrt{\gamma}\mathbf{A}_1, \ \mathbf{B}_2 = \sqrt{\gamma}\mathbf{A}_2 \tag{14}$$

with $0 \leqslant \gamma \leqslant 1$ to satisfy the convexity condition in Eq. (13). The convexity parameter $\gamma$ controls the nonconvexity of the GMC penalty. If $\gamma = 0$, then $\mathbf{B}_1 = 0$, $\mathbf{B}_2 = 0$, and the penalty in Eq. (8) reduces to the L1-norm penalty. If $\gamma = 1$, then Eq. (13) is satisfied with equality and the penalty is "maximally" nonconvex such that the total convexity of the cost function can be guaranteed. In this case, $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \gamma\mathbf{A}^{\mathrm{T}}\mathbf{A}$, and the convexity condition is satisfied. Then $\mathbf{B} = \sqrt{\gamma}\mathbf{A}$ is substituted into Eq. (12), and the optimization problem for the sparsity-enhanced MCA is rewritten as

$$\mathbf{x}^{\text{opt}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\lambda \odot \mathbf{x}\|_1 - \min_{\mathbf{v}} \left\{ \|\lambda \odot \mathbf{v}\|_1 + \frac{\gamma}{2}\|\mathbf{A}(\mathbf{x} - \mathbf{v})\|_2^2 \right\} \right\}. \tag{15}$$

That is,

$$\begin{aligned}(\mathbf{x}_1^{\text{opt}}, \mathbf{x}_2^{\text{opt}}) &= \arg\min_{\mathbf{x}_1,\mathbf{x}_2} \{F(\mathbf{x}_1, \mathbf{x}_2)\} \\ &= \arg\min_{\mathbf{x}_1,\mathbf{x}_2} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}_1\mathbf{x}_1 - \mathbf{A}_2\mathbf{x}_2\|_2^2 + \lambda_1\|\mathbf{x}_1\|_1 + \lambda_2\|\mathbf{x}_2\|_1 \right. \\ &\quad \left. - \min_{\mathbf{v}_1,\mathbf{v}_2} \left\{ \lambda_1\|\mathbf{v}_1\|_1 + \lambda_2\|\mathbf{v}_2\|_1 + \frac{\gamma}{2}\|\mathbf{A}_1(\mathbf{x}_1 - \mathbf{v}_1) + \mathbf{A}_2(\mathbf{x}_2 - \mathbf{v}_2)\|_2^2 \right\} \right\}. \end{aligned} \tag{16}$$

Then $\mathbf{y}_1$ and $\mathbf{y}_2$ can be estimated by $\widehat{\mathbf{y}}_1 = \mathbf{A}_1\mathbf{x}_1^{\text{opt}}$ and $\widehat{\mathbf{y}}_2 = \mathbf{A}_2\mathbf{x}_2^{\text{opt}}$. For gearbox fault diagnosis, $\widehat{\mathbf{y}}_1$ and $\widehat{\mathbf{y}}_2$ are the estimates of the meshing component and the periodic transient component, respectively.

### 3.3. Convex optimization for solving sparsity-enhanced MCA problem

In this subsection, we consider how to solve the optimization problem for the sparsity-enhanced MCA.

Because the GMC penalty in Eq. (8) itself is an L1-norm regularized optimization problem, and does not have an explicit formula, and so cannot be directly evaluated, it may appear necessary to use an iterative algorithm comprising a double nested loop: an outer loop to iteratively solve for the optimal solution $\mathbf{x}^{\text{opt}}$ and an inner loop to solve for the optimal solution $\mathbf{v}^{\text{opt}}$. Fortunately, because $F$ can be expressed as a saddle-point function which is convex in $\mathbf{x}$ and concave in $\mathbf{v}$, a global minimum can be readily calculated using a proximal algorithm with no nested iterations. There is no need to explicitly evaluate the GMC penalty or its gradient [56].

The minimization problem in Eq. (15) can be rewritten as the saddle-point problem

$$(\mathbf{x}^{\text{opt}}, \mathbf{v}^{\text{opt}}) = \arg\min_{\mathbf{x}}\min_{\mathbf{v}} \left\{ F_s(\mathbf{x}, \mathbf{v}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\lambda \odot \mathbf{x}\|_1 - \|\lambda \odot \mathbf{v}\|_1 - \frac{\gamma}{2}\|\mathbf{A}(\mathbf{x} - \mathbf{v})\|_2^2 \right\} \tag{17}$$

with $0 < \gamma < 1$. The solution $(\mathbf{x}^{\text{opt}}, \mathbf{v}^{\text{opt}})$ is a saddle-point of $F_s(\mathbf{x}, \mathbf{v})$. This saddle point problem can be solved via the FBS algorithm, as shown in Algorithm 1, where $\mu$ is a constant satisfying [36]

$$0 \leqslant \mu \leqslant \frac{2}{\max\{1, \gamma/(1-\gamma)\}\|\mathbf{A}^{\mathrm{T}}\mathbf{A}\|_2} \tag{18}$$

where $\|\mathbf{A}^{\mathrm{T}}\mathbf{A}\|_2$ is the maximum eigenvalue of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. The detailed derivation is listed in Appendix A. In this algorithm, we usually use $0.5 \leqslant \gamma \leqslant 0.8$. The soft thresholding operator with threshold $T \in \mathbb{R}_+$ is defined as

$$\text{soft}(y; T) = y \cdot \max\{0, 1 - T/|y|\}, y \in \mathbb{C}. \tag{19}$$

It can be seen from Algorithm 1 that the computation cost of the proposed sparsity-enhanced signal decomposition method depends on the computation costs of the transforms $\mathbf{A}_1$ and $\mathbf{A}_2$ for the meshing component and the periodic transient component, as well as the iterative algorithm. Hence, for the purpose of sparse representation of the meshing component and periodic transient component, and taking the computation cost into account, we choose DCT and STFT among kinds of transforms to minimize the computation cost of the proposed method. This is because DCT can sparsely represent the meshing component while it is not sparse for the transient component, and STFT with the Gaussian window can sparsely represent the transient component while it is not sparse for the meshing component. Thus, $\mathbf{A}_1$ is an oversampled inverse DCT and $\mathbf{A}_1^{\mathrm{T}}$ is an oversampled DCT. The columns of $\mathbf{A}_1$ form a normalized tight frame, i.e., $\mathbf{A}_1\mathbf{A}_1^{\mathrm{T}} = \mathbf{I}$. Similarly, $\mathbf{A}_2$ is an oversampled inverse STFT and $\mathbf{A}_2^{\mathrm{T}}$ is an oversampled STFT. The columns of $\mathbf{A}_2$ also form a normalized tight frame, that is, $\mathbf{A}_2\mathbf{A}_2^{\mathrm{T}} = \mathbf{I}$. Moreover, the operation of DCT and STFT is matrix-free and efficiently implemented using FFT. Thus Algorithm 1 is also matrix-free and efficient for minimizing the cost function in Eq. (16). Then, the computational cost of the proposed sparsity-enhanced signal decomposition method depends on the computational costs of DCT and STFT. The computation cost of DCT is $O(N_1 \log_2 N_1)$ where $N_1$ is the transform length for DCT, and the computation cost of STFT is $O(M/L \cdot N_{ft} \log_2 N_{ft})$ where $M$ is the signal length, $L$ is the window hop-length for STFT, and $N_{ft}$ is the Fourier transform length for STFT. Hence, the total computation cost of the proposed method is $O(N_{iter}(N_1 \log_2 N_1 + M/L \cdot N_{ft} \log_2 N_{ft})$, where $N_{iter}$ is the iteration in the FBS algorithm.

**Algorithm 1** FBS algorithm for sparsity-enhanced MCA problem.

**Initialize:** $\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}, \mathbf{v}_1^{(0)}, \mathbf{v}_2^{(0)}, 0 \leqslant \mu \leqslant \frac{2}{\max\{1,\gamma/(1-\gamma)\}\|\mathbf{A}^{\mathrm{T}}\mathbf{A}\|_2}$.
Number of iteration: $N_{iter}$
Stopping threshold: $\delta$
**for** $k = 0$ to $N_{iter}$ **do**
  $\mathbf{w}_1^{(k)} = \mathbf{x}_1^{(k)} - \mu\mathbf{A}_1^{\mathrm{T}}(\mathbf{A}_1(\mathbf{x}_1^{(k)} + \gamma(\mathbf{v}_1^{(k)} - \mathbf{x}_1^{(k)})) + \mathbf{A}_2(\mathbf{x}_2^{(k)} + \gamma(\mathbf{v}_2^{(k)} - \mathbf{x}_2^{(k)})) - \mathbf{y})$,
  $\mathbf{w}_2^{(k)} = \mathbf{x}_2^{(k)} - \mu\mathbf{A}_2^{\mathrm{T}}(\mathbf{A}_1(\mathbf{x}_1^{(k)} + \gamma(\mathbf{v}_1^{(k)} - \mathbf{x}_1^{(k)})) + \mathbf{A}_2(\mathbf{x}_2^{(k)} + \gamma(\mathbf{v}_2^{(k)} - \mathbf{x}_2^{(k)})) - \mathbf{y})$,
  $\mathbf{u}_1^{(k)} = \mathbf{v}_1^{(k)} - \mu\gamma\mathbf{A}_1^{\mathrm{T}}(\mathbf{A}_1(\mathbf{v}_1^{(k)} - \mathbf{x}_1^{(k)}) + A_2(v_2^{(k)} - \mathbf{x}_2^{(k)}))$,
  $\mathbf{u}_2^{(k)} = \mathbf{v}_2^{(k)} - \mu\gamma\mathbf{A}_2^{\mathrm{T}}(\mathbf{A}_1(\mathbf{v}_1^{(k)} - \mathbf{x}_1^{(k)}) + A_2(v_2^{(k)} - \mathbf{x}_2^{(k)}))$,
  $\mathbf{x}_1^{(k+1)} = \text{soft}(\mathbf{w}_1^{(k)}; \mu\lambda_1)$,
  $\mathbf{x}_2^{(k+1)} = \text{soft}(\mathbf{w}_2^{(k)}; \mu\lambda_2)$,
  $\mathbf{v}_1^{(k+1)} = \text{soft}(\mathbf{u}_1^{(k)}; \mu\lambda_1)$,
  $\mathbf{v}_2^{(k+1)} = \text{soft}(\mathbf{u}_2^{(k)}; \mu\lambda_2)$,
  **if** $\max\{\|\mathbf{x}_1^{(k+1)} - \mathbf{x}_1^{(k)}\|_2, \|\mathbf{x}_2^{(k+1)} - \mathbf{x}_2^{(k)}\|_2\} \leqslant \delta$, **then** break; **end if**
**end for**
**return** $\mathbf{x}_1^{\text{opt}} = \mathbf{x}_1^{(k+1)}, \mathbf{x}_2^{\text{opt}} = \mathbf{x}_2^{(k+1)}$

For gearbox fault diagnosis, based on the above iteration, the meshing component and the periodic transient component can be decomposed by $\widehat{\mathbf{y}}_1 = \mathbf{A}_1\mathbf{x}_1^{\text{opt}}$ and $\widehat{\mathbf{y}}_2 = \mathbf{A}_2\mathbf{x}_2^{\text{opt}}$, respectively. In the next section, we consider how to use the proposed

sparsity-enhanced signal decomposition method to improve the signal decomposition accuracy and diagnose gearbox fault.

## 4. Gearbox fault diagnosis via sparsity-enhanced signal decomposition method

To improve the signal decomposition accuracy and obtain a more reliable fault diagnosis result, we introduce the proposed sparsity-enhanced signal decomposition method for the analysis of gearbox vibration signal. Fig. 2 illustrates the flowchart describing the use of sparsity-enhanced signal decomposition method for gearbox fault diagnosis, which consists of three steps:

*Step 1*: We collect vibration signals from the gearbox.
*Step 2*: The vibration signal is decomposed using the proposed sparsity-enhanced signal decomposition method. In the implementation of sparsity-enhanced signal decomposition via MCA, we first choose two distinct transforms which can sparsely represent the meshing component and the periodic transient component, respectively. MCA relies on the use of appropriate transforms for effective sparse decomposition. In this paper, we prescribe DCT and STFT as the transforms for the meshing component and the periodic transient component, respectively. Second, transform-specific parameters are set. Third, the regularization parameters $\lambda_1$, $\lambda_2$ and the convexity parameter $\gamma$ are set. Then, the sparsity-enhanced MCA cost function is defined using the GMC penalty as a regularizer. Finally, the iterative algorithm in Algorithm 1 is used to solve the sparse coefficients of each signal component.
*Step 3*: Based on the decomposition result of Step 2, the meshing component and the periodic transient component are decomposed. The fault feature for the gearbox is obtained from the periodic transient component. Finally, a gearbox fault diagnosis decision is made.

## 5. Simulation study and parameter selection

To verify the effectiveness of the proposed sparsity-enhanced signal decomposition method for gearbox fault diagnosis, a simulation experiment is studied in this section. The result is compared with that of the L1-norm regularized MCA and SK to illustrate the benefit of the sparsity-enhanced MCA for signal decomposition. Moreover, a strategy for choosing transforms and setting parameters is discussed in this section.

### 5.1. Simulation study

Considering the characteristic of the faulty gearbox vibration signal, a noisy simulated signal is constructed:

$$y(t) = y_1(t) + y_2(t) + n(t) = (0.8\cos(2\pi f_1 t) + 0.6\cos(2\pi f_2 t)) + \sum_k h(t - T_0 - kT) + n(t)$$

where $y_1(t) = 0.8\cos(2\pi f_1 t) + 0.6\cos(2\pi f_2 t)$ represents the meshing component with $f_1 = 45$ Hz and $f_2 = 500$ Hz, $y_2(t) = \sum_k h(t - T_0 - kT)$ represents the periodic transient component. Here, $h(t)$ is a single transient and $kT + T_0$ is the time center of the $k$-th transient with a time-offset of $T_0 = 0.08$ s and a period of $T = 0.1$ s, defined as

$$h(t) = \mathrm{e}^{-\xi/\sqrt{1-\xi^2}\cdot(2\pi f_0 t)^2}\cos(2\pi f_0 t)$$

where the frequency is $f_0 = 180$ Hz and the damping ratio is $\xi = 0.015$. The noise $n \sim N(0, \sigma^2)$ is the additive white Gaussian noise with standard deviation $\sigma = 0.6$. The sampling frequency is 4096 Hz and the signal length is $M = 4096$. The waveforms of the simulated signals are illustrated in Fig. 3(a)–(d).

We use the proposed method to analyze the simulated signal. According to the procedure illustrated in Fig. 2, we first choose two transforms to represent the meshing component $y_1$ and the periodic transient component $y_2$, respectively. By taking the computation cost into consideration, DCT and STFT which have high computation efficiency and are suitable for the sparse representation of the meshing component and the periodic transient component are introduced in this study. Next, the transform parameters of DCT $\mathbf{A}_1^{\mathrm{T}}$ and STFT $\mathbf{A}_2^{\mathrm{T}}$ are set to perform the proposed method. The transform length for DCT is $N_1 = 2.5$ M. For STFT $\mathbf{A}_2^{\mathrm{T}}$, a Gaussian window with a length of $R = 128$ is employed. The window hop-length is set to be $L = 16$ samples. The Fourier transform length for STFT is $N_{ft} = 4R$. The detailed discussion about parameter setting will be provided in the next subsection.

Then, we must set the regularization parameters $\lambda_1$ and $\lambda_2$. During the implementation of the proposed method, we found the regularization parameters $\lambda_1$ and $\lambda_2$ are critical for the performance of the proposed method and they must be set appropriately. In this example, $\lambda_1$ is set to be 3.558 and $\lambda_2$ is set to be 0.332 (the detailed strategy for parameter setting is discussed in the next subsection).

Finally, we must also set $\gamma$ to guarantee the convexity condition $\mathbf{A}^{\mathrm{T}}\mathbf{A} - \mathbf{B}^{\mathrm{T}}\mathbf{B} \succeq 0$. It has been mentioned that the suggested range for $\gamma$ is $0.5 \leqslant \gamma \leqslant 0.8$. Here we simply set $\gamma = 0.8$ to maintain the convexity of the sparsity-enhanced MCA cost function. Besides the value of 0.8, some other values between 0.5 and 0.8 also can be assigned to $\gamma$ to obtain a reasonable analysis result.

Fig. 4 shows the decomposition result using the proposed sparsity-enhanced signal decomposition method with the above parameters, where the dashed line represents the noise-free signal and the solid line represents the estimated signal obtained

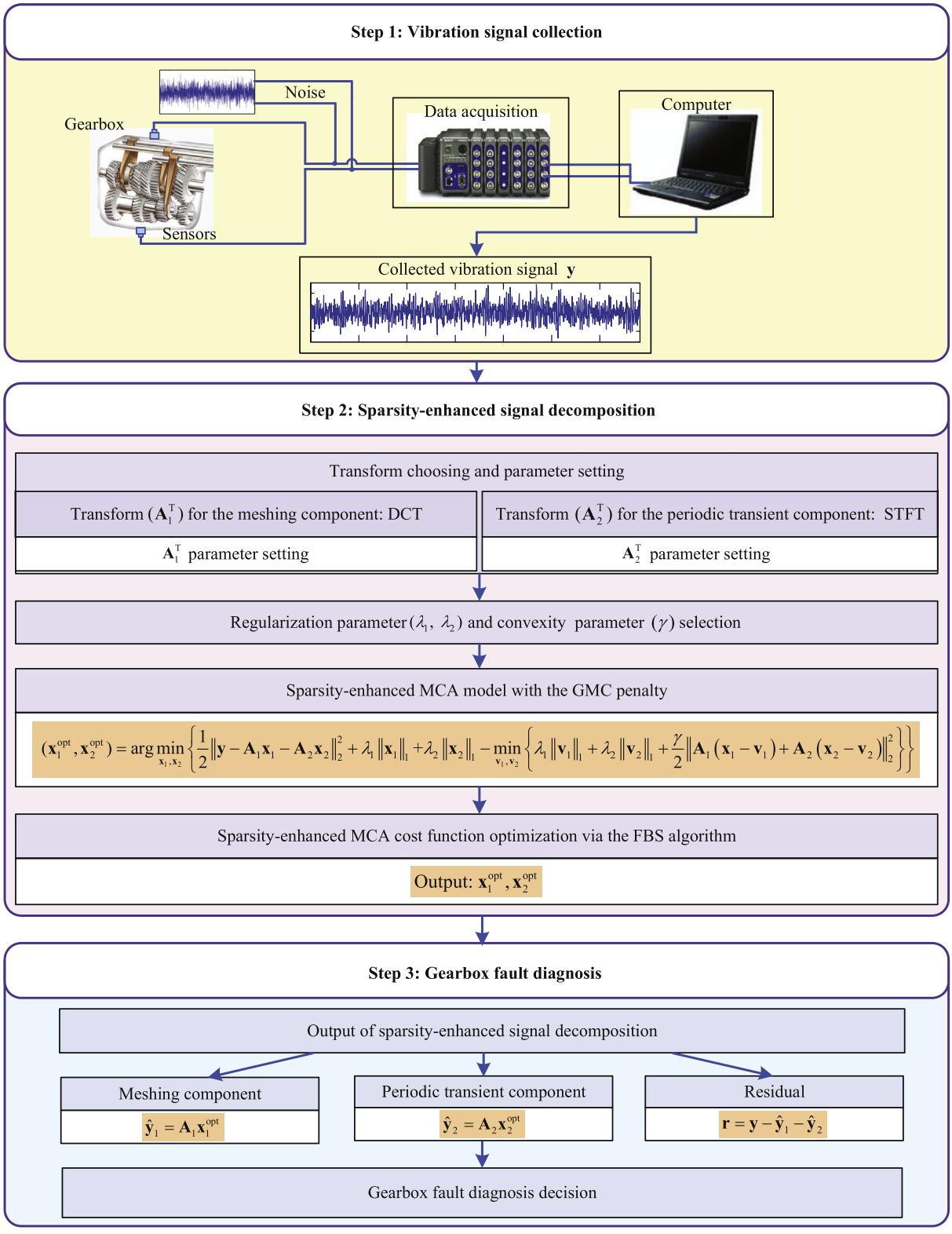


**Fig. 2.** Flowchart of the sparsity-enhanced signal decomposition method for gearbox fault diagnosis.

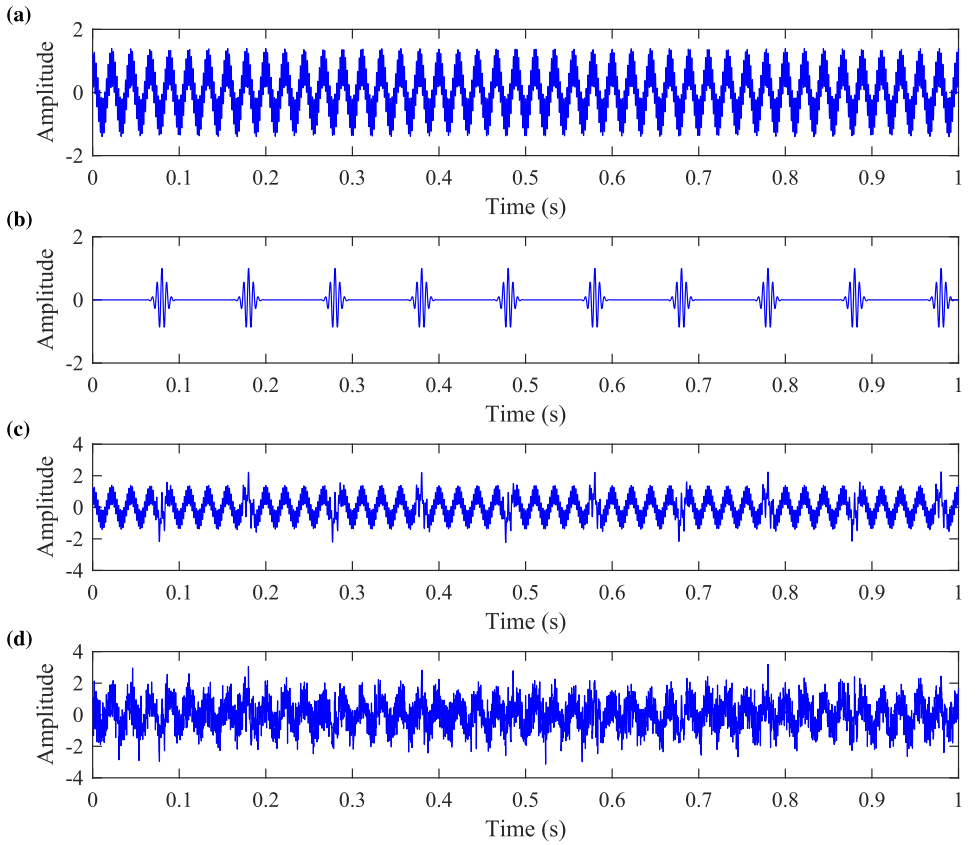


**Fig. 3.** The simulated signal: (a) the meshing component $y_1(t)$; (b) the periodic transient component $y_2(t)$; (c) the noise free simulated signal $y_1(t) + y_2(t)$; (d) the noisy simulated signal $y(t)$.

using the proposed method. Fig. 4(a) is the estimated meshing component and Fig. 4(b) is the estimated periodic transient component. The comparison between the estimated transients and the noise-free signal shows that the proposed method can provide a reasonably accurate result. Moreover, we use the root-mean-square-error (RMSE) to evaluate the performance of the sparsity-enhanced MCA. The RMSE is defined as

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(x(m) - \widehat{x}(m))^2} \tag{20}$$

where $x$ and $\widehat{x}$ are the original component and the estimated component. The smaller the RMSE is, the better the decomposition performance is. It can be seen from Fig. 4(a) and (b) that the RMSE values of the estimated meshing component and the periodic transient component are 0.088 and 0.053, respectively. Fig. 4(c) and (d) show the details of the first and fifth estimated transients in Fig. 4(b). It can be observed from Fig. 4(c) that the estimated transient is almost the same as the noise-free signal. Fig. 4(d) shows the fifth transient in detail. Even though the estimation result is not as accurate as that of the first transient in Fig. 4(c), we observe that it quite accurately preserves the maximum amplitude value of the noise-free signal.

For comparison, the simulated signal is also analyzed using the L1-norm regularized MCA and SK. We use the same transforms and transform parameters with the proposed method for the L1-norm regularized MCA. We properly adjust the regularization parameters for the L1-norm solution to minimize the RMSE for $y_2$ (the detailed strategy for parameter setting is discussed in the next subsection). Fig. 5 shows the decomposition result using the L1-norm regularized MCA with the minimum RMSE for the periodic transient component, where the dashed line represents the ground-truth and the solid line represents the L1-norm regularized MCA solution. The estimated meshing component and the periodic transient component are illustrated in Fig. 5(a) and (b), respectively. The RMSE values for these two components are 0.091 and 0.103, respectively. Compared with the result of the proposed method, it can be seen that the proposed method reduces the RMSE value of the periodic transient component by about 50% relative to the L1-norm regularized MCA. Fig. 5(c) and (d) show the details of the first and fifth estimated transients in Fig. 5(b). It can be seen from Fig. 5(b)–(d) that the amplitude of the estimated transient is diminished. Moreover, there also exist interferences in the estimated transient component. In practical signal analysis, the noise is usually more difficult to be removed and thus the weak fault feature may still be disturbed by the noise. If we increase the regularization parameters to remove more noise, the amplitude of the transient will be diminished even heavier. From this comparison, it can be seen that the

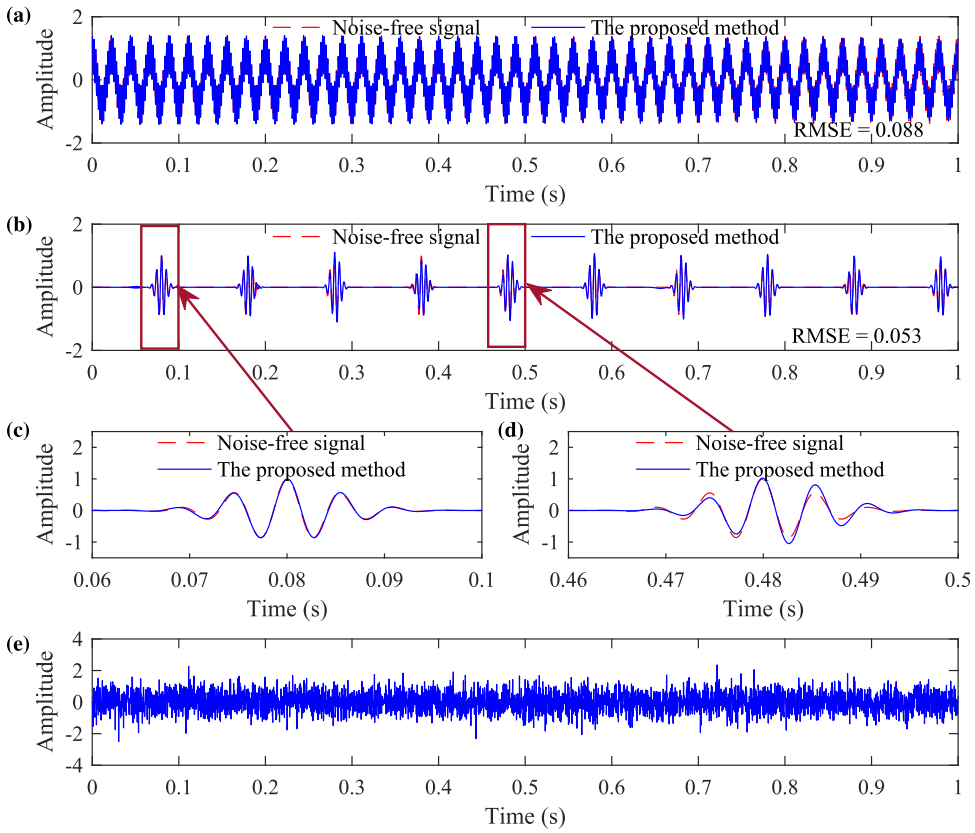


**Fig. 4.** Decomposition result using the proposed sparsity-enhanced signal decomposition method: (a) the estimated meshing component; (b) the estimated periodic transient component; (c) the zoomed-in plot for the first transient; (d) the zoomed-in plot for the fifth transient; (e) the residual component.

proposed method more accurately preserves the signal of interest while reducing noise, thus improves the signal decomposition performance.

SK is a powerful analysis technique for transients detection in vibration signals and it has been widely studied and used in rotating machine fault diagnosis [18,19]. In this study, SK is also used to analyze the simulated signal. The kurtogram of SK and the resulting filtered signal are illustrated in Fig. 6. It can be observed that SK effectively detects the location (Fig. 6(b)) and the period (Fig. 6(c)) of the transients. However, the amplitude of the periodic transients in Fig. 6(b) is diminished, and there are many artifacts in the filtered result. The proposed method also outperforms SK in this example.

### *5.2. Parameter setting strategy*

In this subsection, we discuss a guideline for choosing the transforms and setting the regularization parameters.

#### *5.2.1. Transform choice*

The success of MCA relies on the sparse representations of morphologically-distinct components. That is, each component should be sparsely represented by a prescribed transform which should not be able to sparsely represent the other component. This gives us a basic principle for choosing transforms in the use of the proposed method. In this paper, the goal is to separate the meshing component and the periodic transient component in the signal. Hence, we choose two distinct transforms to represent the meshing component and the periodic transient component. For the meshing component, we use DCT with a transform length $N_1 = 2.5\,\mathrm{M}$. For the sparse representation of the periodic transient component, we use STFT with a Gaussian window. As for the window length, we suggest choosing a relatively short window because a short window gives good time resolution, which is useful for the identification of the location of transients and their periodicity. If the window is too long, then it will enhance the ability of STFT to sparsely represent the meshing component, which is unfavorable for the decomposition of the raw signal into the meshing component and the periodic transient component. In this study, we always use a Gaussian window with a window length $R = 128$. Moreover, we suggest choosing the length of the DFT for each window 2 or 4 times the window length. This is because if DFT for each window is too long, then it will affect the computation efficiency. In this paper, the length of DFT is always 4 times the window length, that is $N_{ft} = 4R$. And after each step, we always shift the analysis window along by

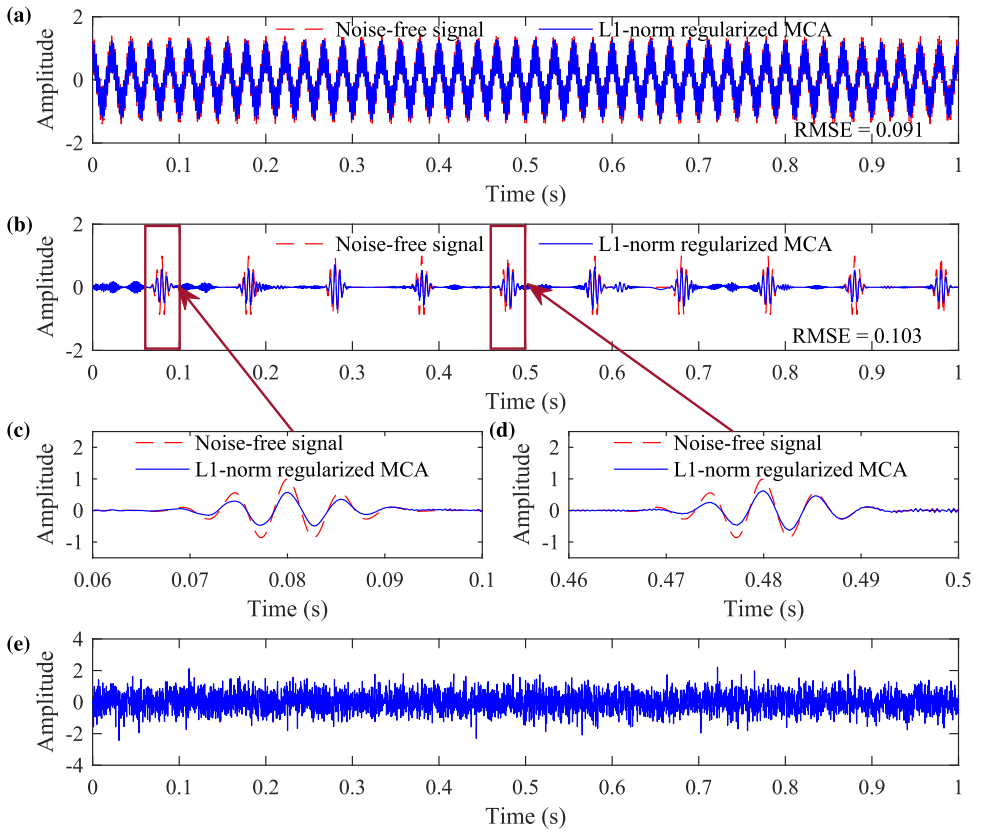


**Fig. 5.** Decomposition result using the L1-norm regularized MCA: (a) the estimated meshing component; (b) the estimated periodic transient component; (c) the zoomed-in plot; (d) the zoomed-in plot; (e) the residual component.

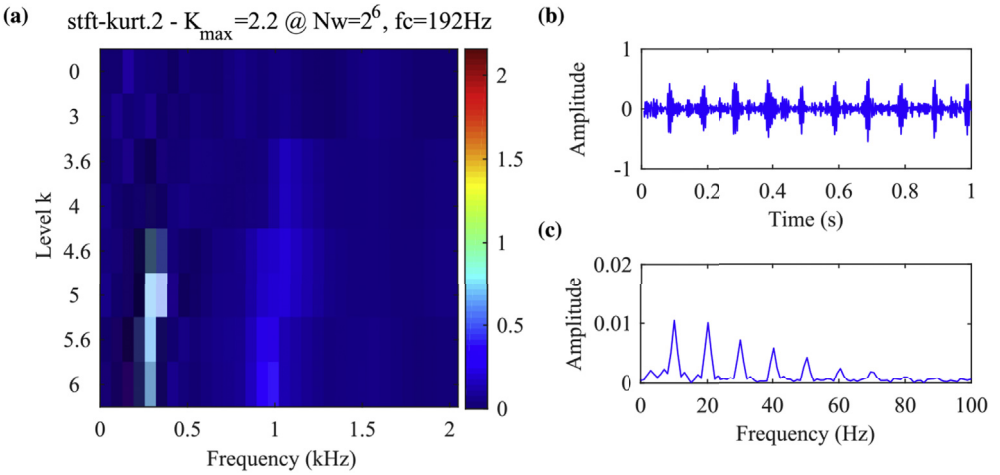


**Fig. 6.** The analysis result of the simulation signal via SK: (a) the kurtogram; (b) the resulting filtered signal; and (c) the envelope spectrum.

$L = 16$ samples each time (hop-length). Other hop-length for $L$ which can satisfy the perfect reconstruction condition also can be used in this method. The larger the hop-length $L$ is, the more computationally efficient the algorithm is.

Fig. 7 shows the ability of DCT and STFT to represent the simulated signal. It can be observed from Fig. 7(b) and (c) that DCT can sparsely represent the meshing component, while its representation of the periodic transient component is inefficient (non-sparse). From Fig. 7(f) and (g), it can be seen that STFT can sparsely represent the periodic transient component, while it cannot achieve a sparse representation of the meshing component. This figure demonstrates that the transforms used in this example meet the basic principle and are appropriate for the separation of the meshing component and the periodic transient component

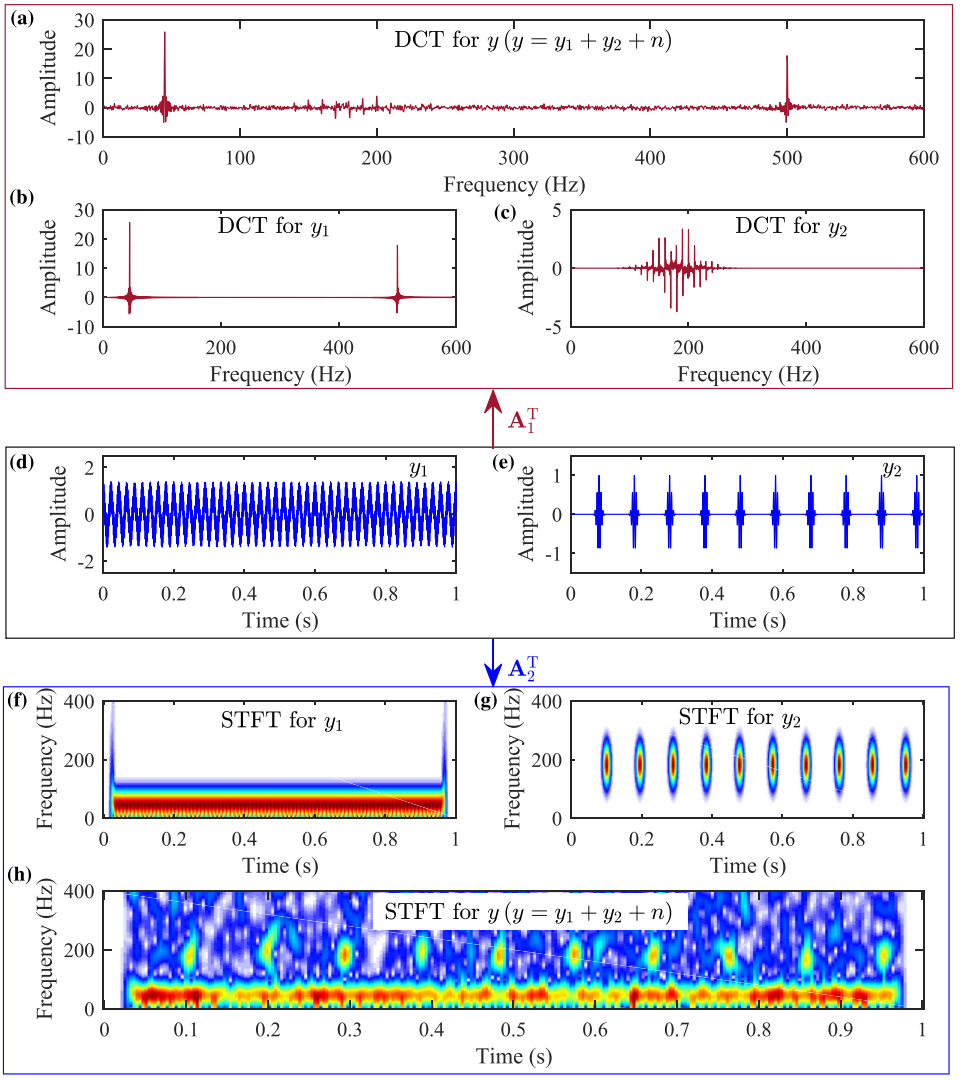


**Fig. 7.** Representation ability of DCT and STFT for different components in the simulated signal: (a) DCT for $y$; (b) DCT for $y_1$; (c) DCT for $y_2$; (d) the waveform of $y_1$; (e) the waveform of $y_2$; (f) STFT for $y_1$; (g) STFT for $y_2$; (h) STFT for $y$.

via MCA. It must be mentioned that other transforms which have similar representation properties can also be employed in the study. Other transforms which can give a sparse representation for the meshing component can also be employed as $\mathbf{A}_1^{\mathrm{T}}$, such as the discrete Fourier transform (DFT) and tunable-Q wavelet transform (TQWT) with a high Q-factor. Additionally, other transforms which can represent transients sparsely can be employed as $\mathbf{A}_2^{\mathrm{T}}$, such as TQWT with a low Q-factor and adaptive impulse dictionary.

*5.2.2. Regularization parameters setting*

The regularization parameters $\lambda_1$ and $\lambda_2$ are also critical for signal decomposition accuracy. The parameters $\lambda = [\lambda_1, \lambda_2]$ control the tradeoff between the regularization term and the data fidelity term in the objective function. Moreover, the parameters $\lambda_1$ and $\lambda_2$ also control the relative sparsity of the transform coefficients of the two morphological different components [57]. In this paper, it means that the parameters $\lambda_1$ and $\lambda_2$ control the relative sparsity between the DCT coefficients and the STFT coefficients. Furthermore, it can be seen from the FBS algorithm in Algorithm 1 that the regularization parameters determine the threshold of the soft threshold operator in the algorithm. The transforms (DCT and STFT) used in this paper have the property that each column of the inverse transforms $\mathbf{A}_1, \mathbf{A}_2$ have the same L2 norm $\eta_1, \eta_2$, respectively. Hence, we suggest setting $\lambda_1$ and $\lambda_2$ as follows,

$$\lambda_1 = \beta_1\eta_1\sigma,\ \lambda_2 = \beta_2\eta_2\sigma \tag{21}$$

Both parameters $\lambda_1$ and $\lambda_2$ consist of three factors. They are: (1) the standard deviation $\sigma$ of the noise in the signal, (2) the L2 norm of each column of the inverse transform $A_1$ and $A_2$ ($\eta_1$, $\eta_2$), and (3) the ratio parameters $\beta_1$ and $\beta_2$. Next, we consider how to set the parameters $\lambda_1$ and $\lambda_2$.

1) Parameter $\sigma$: The parameter $\sigma$ is the standard deviation of the noise in the signal. In this example, the standard deviation of the noise in the simulated signal is known. However, in practical application, the standard deviation of the noise is usually unknown. Fortunately, we can roughly estimate the standard deviation of the noise in a signal by [58]

$$\widehat{\sigma} = \mathrm{MAD}(y)/0.6745. \tag{22}$$

This is a conventional noise level estimator, and it depends only on the noisy observation. The MAD function is the 'median absolute deviation' operator, defined as

$$\mathrm{MAD}(y) = \mathrm{median}(|y - \mathrm{median}(y)|). \tag{23}$$

2) Parameters $\eta_1$ and $\eta_2$: The parameters $\eta_1$ and $\eta_2$ are the L2 norm of each column of the inverse transform of DCT and STFT, respectively. Transforms, DCT and STFT, used in this paper have the property that each column of the inverse transform has the same L2 norm. If the transform is determined, then its L2 norm of each column is determined accordingly. In this example, $\eta_1$ and $\eta_2$ can be calculated as follows [59],

$$\eta_1 = \sqrt{M/N_1},\ \eta_2 = \sqrt{L/N_{ft}}. \tag{24}$$

It has been mentioned above, $M = 4096$, $N_1 = 2.5\,\mathrm{M}$, $L = 16$ and $N_{ft} = 4R = 512$ in this example. Therefore, we have $\eta_1 = 0.6325, \eta_2 = 0.1768$.

3) Parameters $\beta_1$ and $\beta_2$: The remaining task for setting $\lambda_1$ and $\lambda_2$ is to set $\beta_1$ and $\beta_1$. It is obvious that $\lambda_1$ and $\lambda_2$ determine the threshold of the soft thresholding in FBS, and the threshold is proportional to the energy of the noise to be removed in the signal. In this paper, we set $\beta_1$ and $\beta_2$ as

$$\beta_1 = \beta\theta,\quad \beta_2 = \beta(1-\theta), \tag{25}$$

and then we have

$$\lambda_1 = \beta\theta\eta_1\sigma,\ \lambda_2 = \beta(1-\theta)\eta_2\sigma, \tag{26}$$

where the parameter $\beta$ is an unified ratio, and the parameter $\theta$ with $0 < \theta < 1$ controls the relative sparsity of the sparse coefficients of the two morphological-distinct components in the transform domain (DCT and STFT).

Next, we analyze how to set $\beta$ and $\theta$. We fix other parameters in the algorithm and tune $\beta$ and $\theta$ to obtain different decomposition results for the simulated signal, and use RMSE to evaluate the performance of the proposed method. We vary $\beta$ from 1 to 25 with an increment of 0.1, and vary $\theta$ from 0.2 to 0.9 with an increment of 0.025. Fig. 8(a) shows the RMSE between the periodic transient component and its estimation of the proposed method. During the analysis of the simulated signal in Section 5.1, we choose $\beta$ and $\theta$ from the area where the RMSE value of the estimated periodic transients is extremely small. We choose $\beta = 12.5$ and $\theta = 0.75$ which is marked in Fig. 8(a) to analyze the signal. Then the corresponding $\lambda_1$ and $\lambda_2$ are as follows,

$$\lambda_1 = \beta\theta\eta_1\sigma = 12.5 \times 0.75 \times 0.6325 \times 0.6 = 3.558,$$

$$\lambda_2 = \beta(1-\theta)\eta_2\sigma = 12.5 \times 0.25 \times 0.1768 \times 0.6 = 0.332.$$

This explains the reason why we use $\lambda_1 = 3.558$ and $\lambda_2 = 0.332$ to analyze the simulated signal in Section 5.1. Moreover, it also can be observed that, besides $\beta = 12.5$ and $\theta = 0.75$, there is a relatively large range where the RMSE is small, which means that we can also set $\beta$ and $\theta$ to these values to obtain reasonably accurate results using the proposed method.

Up to now, we have analyzed these three types of parameters contributing to $\lambda_1$ and $\lambda_2$. In the practical application to gearbox fault diagnosis, we can use Eq. (22) to estimate the standard deviation of the noise in the signal and calculate $\eta_1$ and $\eta_2$ according to the transforms employed in the proposed method. Then we choose $\beta$ and $\theta$ according to the cross validation result in Fig. 8(a) to determine $\beta_1$ and $\beta_2$. A general way to set the $\beta$ and $\theta$ parameters is to first use the parameter $\beta = 12.5$ and $\theta = 0.75$ to analyze the vibration signal. If the result is not satisfactory, then we can change $\beta$ and $\theta$ around (12.5, 0.75) according to Fig. 8(a) to obtain a reasonable result. Case-studies (below) in automobile transmission and hot strip milling reduction gearbox fault diagnosis will verify the effectiveness of this parameter setting strategy.

Moreover, for comparison, we use the same strategy to find the appropriate parameters for the L1-norm regularized MCA. Fig. 8(b) shows the RMSE value between the periodic transient component and its estimation of the L1-norm regularized MCA. There is also a relatively large area where the L1-norm regularized MCA can achieve a relatively good result. The parameters used for the analysis in Fig. 7 are also marked in Fig. 8(b). However, it can be seen from Fig. 8 that the RMSE value of the L1-norm

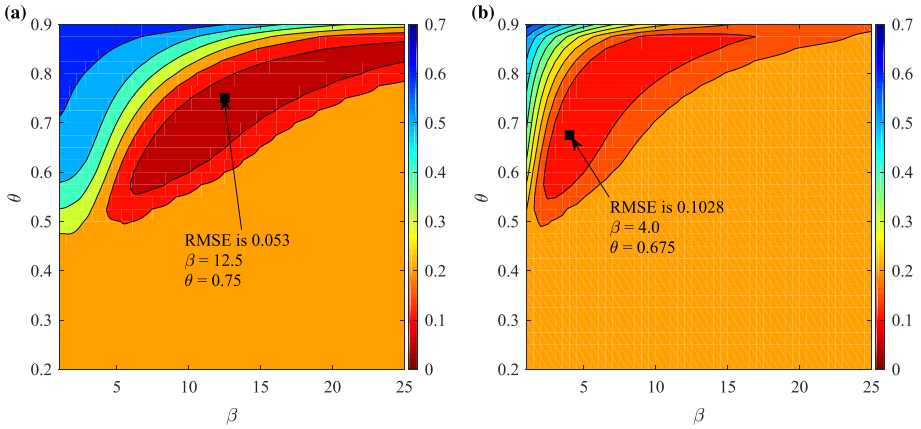


**Fig. 8.** RMSE of the periodic transient component and its estimate: (a) the proposed method; (b) the L1-norm regularized MCA.

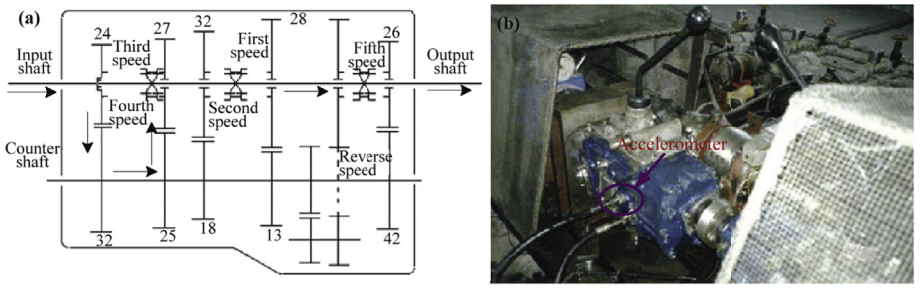


**Fig. 9.** The automobile transmission gearbox: (a) The structure of the gearbox and (b) the setup of the gearbox.

regularized MCA is much larger than that of the proposed method, which further verifies that the proposed sparsity-enhanced signal decomposition method can achieve a more accurate result than the L1-norm regularized MCA.

## 6. Practical applications

In this section, we apply the proposed method to analyze signals collected from a run-to-failure experiment test and an engineering case to further validate the performance of the proposed method, as well as the usability of the strategy for choosing the transforms and setting parameters.

### *6.1. Case 1: automobile transmission gearbox vibration signal analysis*

In this experiment case study, the vibration signals are collected from an automobile transmission. The tested gearbox has five forward speeds and one backward speed as shown in Fig. 9(a). Vibration signals were acquired using an accelerometer mounted on the outer case of the gearbox which is loaded on the third speed, as shown in Fig. 9(b). This experiment concerns a fatigue test of 7 cycles. At the beginning of the seventh cycle, a tooth of the third-speed driving gear was found to be broken. During the experiment, the input rotating speed is 1600 r/min ($f_I = 1600/60 = 26.7$ Hz) and the sampling frequency is 3000 Hz. Table 1 shows the working parameters of the third-speed gears. The vibration signal is taken for the following analysis.

**Table 1**
Working parameters of the third speed gears.

| | The third speed gears | | Constant meshing gears | |
|---|---|---|---|---|
| | Driving gear | Driven gear | Driving gear | Driven gear |
| Number of teeth | 25 | 27 | 24 | 32 |
| Rotating frequency (Hz) | 20 | 18.5 | 25 | 33.3 |
| Rotating period (s) | 0.05 | 0.054 | 0.04 | 0.03 |
| Meshing frequency | 500 | | 640 | |

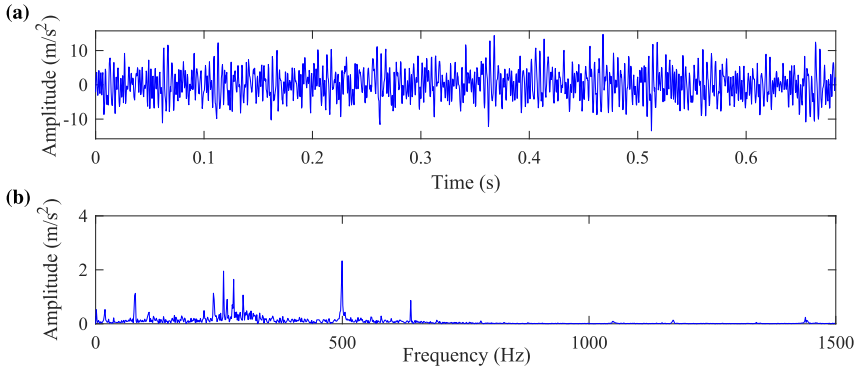


**Fig. 10.** (a) The measured vibration signal of gearbox with severe fault and (b) its Fourier spectrum.

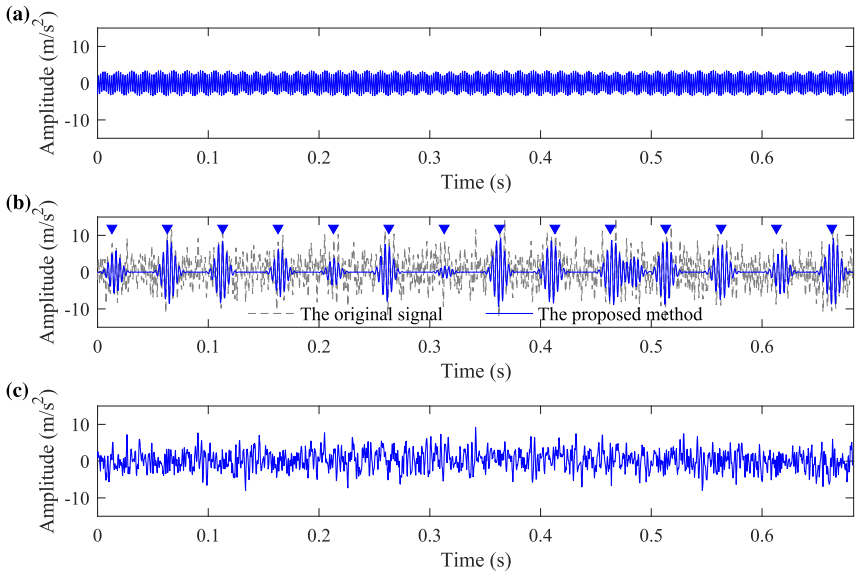


**Fig. 11.** Vibration signal decomposition via the proposed method: (a) the estimated meshing component; (b) the estimated periodic transient component; (c) the residual component.

We expect to find evidence of periodic transients with a frequency of $f_r = 20$ Hz corresponding to the fault frequency of the third-speed driving gear.

*6.1.1. Severe fault signal analysis*

Firstly, a vibration signal collected during the sixth cycle is taken for analysis. Fig. 10(a) shows the time-domain waveform of the collected signal with a length of 2048 samples, and Fig. 10(b) shows the Fourier spectrum of the vibration signal in Fig. 10(a). The fault signature cannot be identified in the time domain in Fig. 10(a) due to background noise contained in the signal. Its Fourier spectrum shows that the component with 500 Hz is predominant, which is the meshing frequency of the third-speed gears. There is no sufficient evidence in the time-domain waveform nor the Fourier spectrum to detect the fault.

We use the proposed method to decompose the vibration signal into the meshing component and the periodic transient component. To run the proposed method, DCT is used to sparsely represent the meshing component in the signal; STFT with a Gaussian window is used to represent the periodic transient component. In this application, both DCT and STFT use the same parameters as in the simulation study. The transform length of DCT is also 2.5 times the signal length. Additionally, the STFT window length is 128 and the window hop-length is 16 samples. The regularization parameters $\lambda_1$ and $\lambda_2$ are set via Eq. (26). The standard deviation of the noise in the signal is estimated via Eq. (22). The parameters $\beta$ and $\theta$ are chosen according to the analysis in subsection 5.2. Here we also choose $\beta = 12.5$ and $\theta = 0.75$ which are the same values as in the simulation study. The parameter $\gamma$ is simply set to be 0.8 in accordance with the simulation study. Fig. 11 shows the decomposition result: Fig. 11(a) shows the estimated meshing component and Fig. 11(b) shows the estimated periodic transient component. In Fig. 11(b) we observe a series of periodic transients with a period of 0.05 s, which corresponds to the rotational frequency of the third-speed

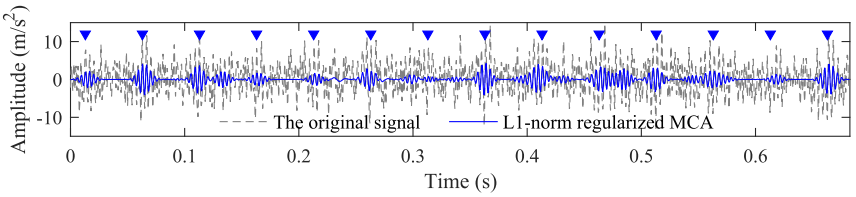


**Fig. 12.** The estimated periodic transient component via the L1-norm regularized MCA.

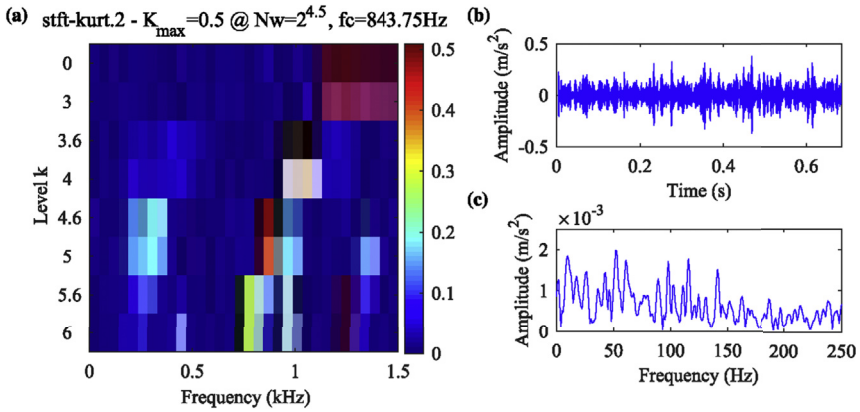


**Fig. 13.** The analysis result of the gearbox vibration signal via SK: (a) the kurtogram; (b) the resulting filter signal; (c) the envelope spectrum.

gears. The proposed method successfully decomposes the signal into the meshing component and the periodic transient component, and extracts fault signatures for gearbox fault diagnosis.

For comparison, the signal is also analyzed via the L1-norm regularized MCA as shown in Fig. 12, and the parameters are the same with that used in the simulation study. It can be seen that the estimated periodic transients component is not as accurate as that estimated by the proposed sparsity-enhanced signal decomposition method. A series of transients are also estimated by the L1-norm regularized MCA. However, the amplitudes of the transients are highly underestimated. Moreover, there also exist interferences in Fig. 12 which affects on the diagnosis of fault.

For further comparison, the signal in Fig. 10(a) is analyzed using SK. Fig. 13 shows the kurtogram and the analysis result using SK. It is difficult to find the periodic transient response in Fig. 13(b). Fig. 13(c) shows the Hilbert envelope of the filtered signal in Fig. 13(b). The fault frequency for the car transmission is still not obvious in Fig. 13(b). The proposed method also outperforms SK in this application.

### *6.1.2. Incipient fault vibration signal analysis*

To further illustrate the superiority of the proposed method, we try to analyze the vibration signal of incipient fault which is collected at the end of the fourth cycle during the experiment. Fig. 14(a) shows the time domain waveform of the signal and Fig. 14(b) shows its Fourier spectrum.

Then we use the proposed method to analyze the signal, and the parameters used in this case are the same with that used in the above. Fig. 15(a) shows the estimated periodic transient component by using the proposed sparsity-enhanced signal decomposition method. It can be seen that the transients in the decomposition result are apparent and the period of the transients is obvious. Fig. 16(a) shows the squared envelope spectrum (SES) of the estimated transient component in Fig. 15(a). It can be seen that the dominant frequencies are $f_r$ and its multiple ($2f_r$), which correspond to the rotating frequency of the third speed driving gear. Based on this analysis result, we can also conclude that there is a localized fault on the third speed driving gear.

Fig. 15(b) shows the analysis result by using the L1-norm regularized MCA. From Fig. 15(b), we can see that the periodic characteristic of the estimated transient component by using the L1-norm regularized MCA is inapparent, and the transient component are submerged in the noise or other interference. Fig. 16(b) shows the SES of the estimated transient component via the L1-norm regularized MCA. However, due to the underestimation problem of L1-norm and the influence of noise, there are a series of frequencies with the similar amplitude in the spectrum and no dominant frequency can be found. There is no evident signature for the fault of the third speed driving gear. From this comparison, it can be seen that the underestimation of L1-norm tends to lead to missed detection of fault. The proposed sparsity-enhanced signal decomposition method can more accurately preserve the signal of interest, thus can diagnose fault earlier in this case.

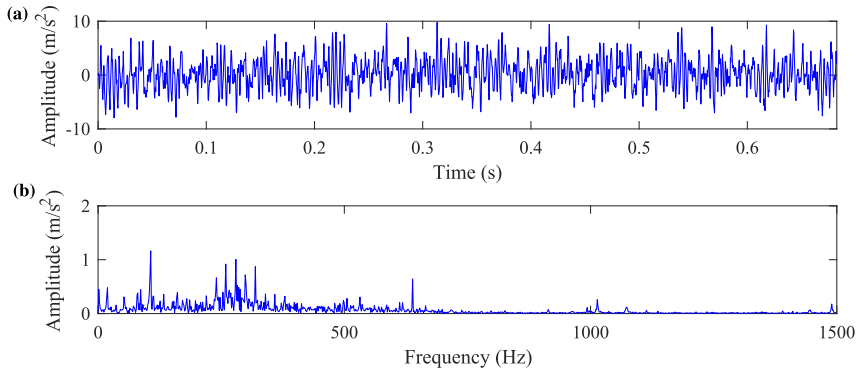


**Fig. 14.** (a) The vibration signal of gearbox with incipient fault and (b) its Fourier spectrum.

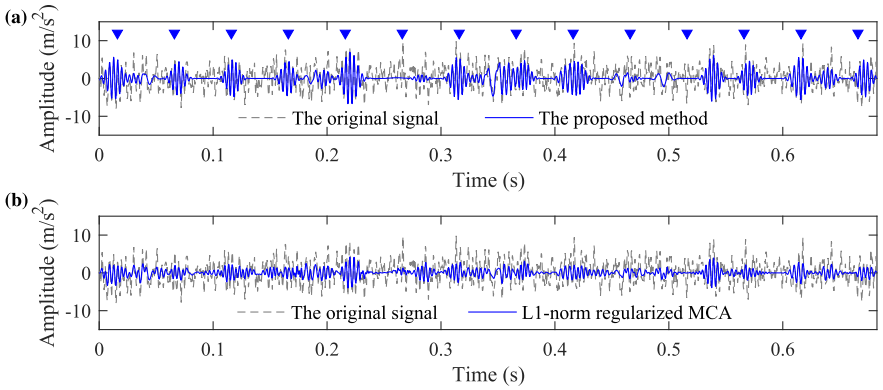


**Fig. 15.** The estimated periodic transient component of the signal with incipient fault using different method: (a) the proposed sparsity-enhanced method; (b) the L1-norm regularized MCA.

### *6.2. Case 2: hot-strip finishing milling machine signal analysis*

In order to further validate the effectiveness of the proposed method and the parameter setting strategy, vibration signals collected from the reduction gearbox in a hot strip finishing milling production line are analyzed by the proposed sparsity-enhanced signal decomposition method. In this engineering case, the proposed method is applied to identify fault features of the reduction gearbox. Fig. 17 shows the schematic sketch of the reduction gearbox of the finishing milling. Its main drive system is a single-stage gearbox with Z22/Z65. Accelerometers were mounted on the accessible bearing housings, including the reduction gearbox and the distribution gearbox. These accelerometers were secured by means of magnetic bases. The collected vibration signals were sampled at a frequency of 5120 Hz and each record of the signal is of length 4096. A certain day, it was found that the vibration of the hot strip finishing milling housing exhibited abnormalities and the root mean square (RMS) value of the signal measured by accelerometer 4 were significantly greater than those of other accelerometers. Then attentions were focused on vibration signals collected from this accelerometer.

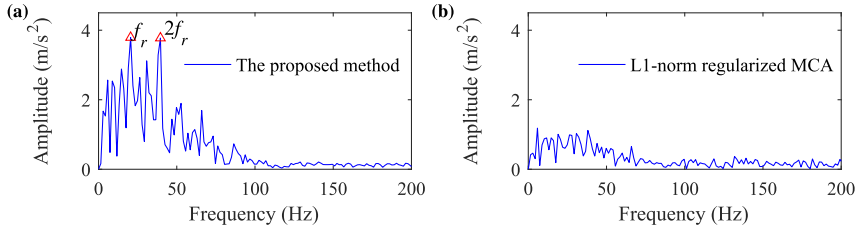


**Fig. 16.** The SES of the estimated transient component via different methods: (a) the proposed method; (b) the L1-norm regularized MCA.

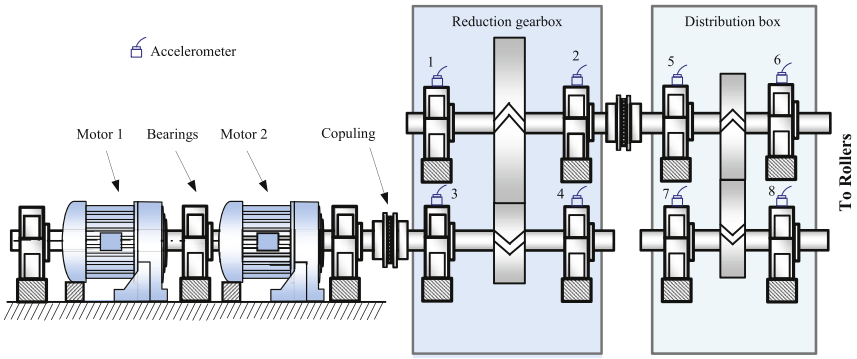


**Fig. 17.** Schematic sketch of the F3 milling stand.

**Table 2**
Characteristic parameters of the reduction gearbox in the milling machine.

| | |
|---|---|
| Transmission ratio (Z2/Z1) | 65/22 |
| Module of the pinion (mm) | 30 |
| Central distance (mm) | 1350 |
| Face-width (mm) | 560 |
| Rotating frequency of input shaft (Hz) | 4.546 |
| Rotating frequency of output shaft (Hz) | 1.539 |
| Meshing frequency (Hz) | 100.03 |

In an inspection, the rotational frequency of the input shaft was measured to be 4.546 Hz via a tachometer, which is equal to the rotational frequency of the pinion. Table 2 lists the characteristic parameters of the reduction gearbox under this condition. Fig. 18(a) shows the time-domain waveform of a record of the collected vibration signal measured by accelerometer 4 and Fig. 18(b) shows its Fourier spectrum. The time-domain waveform appears to have feature similar to local impulses, however, the Fourier spectrum mainly consists of the meshing frequency 100.0 Hz and its multiple harmonics can be clearly observed. No significant sidebands round meshing frequency and its multiple harmonics can be observed. However, there are some frequencies between 100.0 Hz and 200.0 Hz, which indicates a modulation phenomenon. Fig. 18(c) is the zoomed-in plot of the spectrum, and no significant sidebands round the meshing frequency 100.0 Hz can be seen.

Then, the proposed method is applied to analyze the vibration signal for further investigation. In the implementation of the proposed method, the transform parameters of DCT and STFT are also the same as in the simulation analysis. The standard deviation of the noise in the vibration signal is also estimated via Eq. (22). The parameters, $\beta$ and $\theta$, are chosen as $\beta = 12.5$ and $\theta = 0.75$, which are also the same with the former analysis. The parameter $\gamma$ is still set to be 0.8. The analysis results are shown

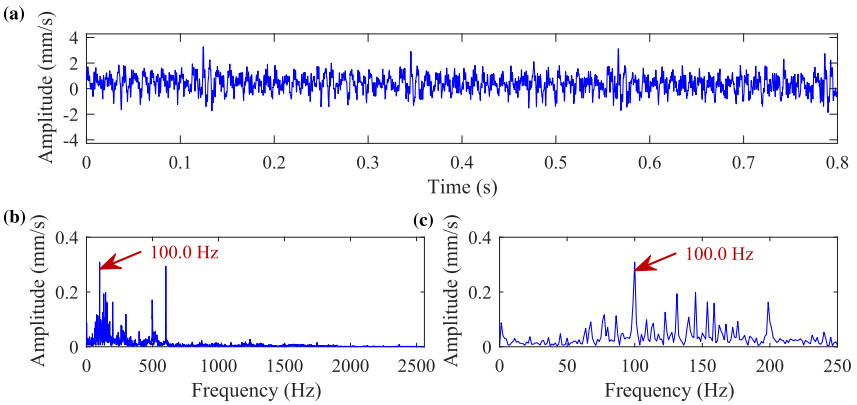


**Fig. 18.** Vibration signal measured by accelerometer 4 and its Fourier spectrum; (a) the time-domain waveform; (b) its Fourier spectrum; (c) its zoomed-in plot for the low-frequency band.

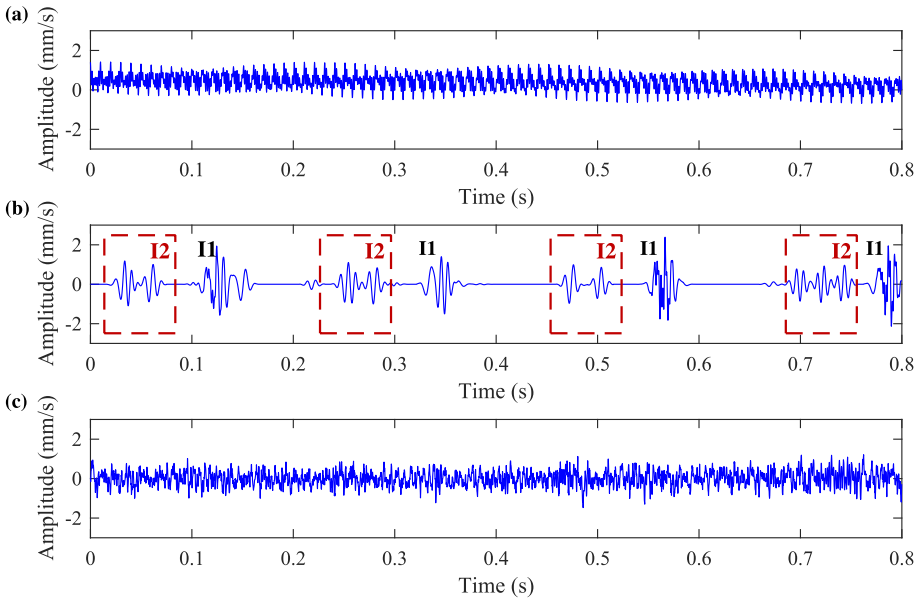


**Fig. 19.** Vibration signal decomposition result using the proposed method: (a) the estimated meshing component; (b) the estimated periodic transient component; (c) the residual component.

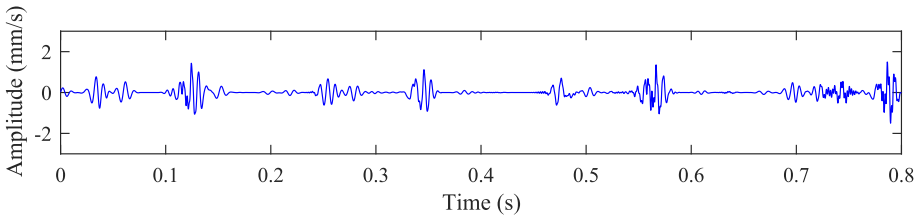


**Fig. 20.** The estimated periodic transient component using the L1-norm regularized MCA.

in Fig. 19. Obviously, a series of strong transients (marked with I1) can be seen in Fig. 19(b). Besides that, there is another series of successive relatively weak transients (marked with I2). Both the strong transients and weak transients occur with a period of 0.22 s, which is consistent with the rotational frequency of the input shaft. Therefore, it can be surmised that there are two sites of localized faults in the reduction gearbox on the pinion, and the fault corresponding to the periodic transients (I1) is worse than the fault corresponding to the periodic transients (I2). Moreover, the time interval between each transient in series I1 and the corresponding transient in series I2 is about 1/3 the rational period of the pinion, implying that these two faults are located about 1/3 the circle of the pinion apart.

For comparison purposes, the L1-norm regularized MCA is employed to analyze the signal, and Fig. 20 shows the decomposed periodic transient component. It can be seen from Fig. 20 that, even though both the strong transient I1 and the weak transient I2 are extracted from the original signal, the amplitudes of the transients are much smaller than that estimated by the proposed enhanced MCA method. The weak transient is almost invisible. Furthermore, SK is employed to analyze this signal for comparison, and the SK result is shown in Fig. 21, where Fig. 21(a) shows the kurtogram of SK and Fig. 21(b) shows the filtered signal. Fig. 21(c) shows the envelope spectrum of Fig. 21(b). We can see from Fig. 21(b) that the series of impulse I1 has been clearly extracted. However, the amplitude of I1 has been greatly diminished. And what's more, the impulse I2 cannot be observed from Fig. 21(b). From the envelope spectrum of the filtered signal in Fig. 21(c), fault feature information still cannot be observed.

Later, the production line was stopped and the reduction gearbox was disassembled for detailed examination. It was found that two-scuffing and wear faults occurred on the pinion. Moreover, the distance between these two faults is about 1/3 the circle of the pinion. Fig. 22(a) shows the damage corresponding to the I1 transient and Fig. 22(b) shows the damaged corresponding to the I2 transient. It can be seen from Fig. 22(b) that several adjacent teeth are damaged on the pinion. If we revisit the analysis result in Fig. 19(b), we can see that each I2 is not a single transient, there are a series of continuous transients. From this observation, we can also draw the conclusion that more than one tooth is broken at the localized fault I2 and the neighboring teeth

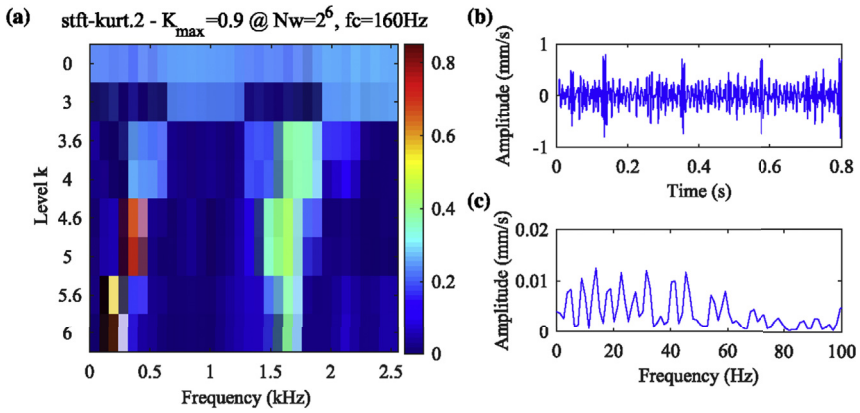


**Fig. 21.** The analysis result of the hot-strip finishing milling machine vibration signal via SK: (a) the kurtogram; (b) the resulting filter signal; (c) the envelope spectrum.

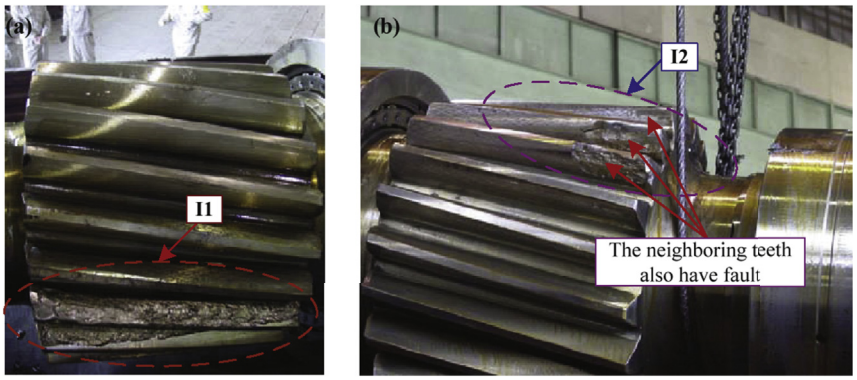


**Fig. 22.** Fault locations on the pinion: (a) the location of the fault I1; and (b) the location of the fault I2.

are supposed to also have fault, which is consistent with the examination. The proposed method analyzes signals reasonably accurately for gearbox fault diagnosis. And the performance of the proposed sparsity-enhanced signal decomposition method outperforms the L1-norm regularized MCA or SK.

## 7. Conclusions

This paper proposes a sparsity-enhanced signal decomposition method using the GMC penalty for gearbox fault diagnosis. The enhancement is implemented by using the GMC penalty to regularize the least squares problem of MCA. The GMC penalty is an effective sparsity-inducing regularizer which can preserve the convexity of the MCA cost function to be minimized, even though it is a nonconvex penalty. Thus the global minimum of the sparsity-enhanced MCA cost function can always be determined by convex optimization. In this paper, we minimize the MCA cost function using the FBS algorithm. Moreover, we provide a guideline for choosing the transforms and setting the regularization parameters. The effectiveness and advantages of the proposed method are illustrated by simulational, experimental and engineering case studies. The results demonstrate that the proposed method can more accurately estimate the components of vibration signals and tends to avoid fault miss detection, which can facilitate gearbox fault diagnosis.

When we use MCA to decompose vibration signals into two morphologically-distinct components, in this paper, taking the computational efficiency into consideration, DCT is employed for the sparse representation of the meshing component, and STFT with the Gaussian window is employed for the periodic transient component. Other transforms can also be employed for providing the sparse representation of the meshing component and the transient component. For example, DFT for the meshing component and wavelet transform for the transient component. This topic is also interesting for machinery fault diagnosis. Moreover, the application of the proposed method is not limited to gearbox fault diagnosis. It could also be used for the fault diagnosis of other rotating machinery whose vibration signals exhibit similar characteristics. It has been verified that the proposed method can obtain reasonably accurate estimates of fault features. Hence, it is a promising method for the condition monitoring and degradation assessment of gearbox and other rotating machinery.

## Acknowledgements

This work was supported by the National Nature Science Foundation of China (No. 51405321, 51605366, and 51375322), the China Postdoctoral Science Foundation (No. 2016M590937, 2017T100740), and the Fundamental Research Funds for the Central Universities. Thanks also go to the anonymous reviewers for their constructive suggestions.

## Appendix A

In this appendix, we illustrate the derivation of the FBS algorithm for solving the saddle-point problem in Eq. (17).

The solution to the problem in Eq. (17) is a saddle-point of $F_s(\mathbf{x}, \mathbf{v})$. The point $(\mathbf{x}^{\text{opt}}, \mathbf{v}^{\text{opt}})$ is a saddle-point of $F_s(\mathbf{x}, \mathbf{v})$ only if $\mathbf{0} \in \partial F_s(\mathbf{x}^{\text{opt}}, \mathbf{v}^{\text{opt}})$, where $\partial F_s$ is the subdifferential of $F_s$. For the function $F_s$ in Eq. (17), we have

$$\partial_{\mathbf{x}} F_s(\mathbf{x}, \mathbf{v}) = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y}) - \gamma \mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{x} - \mathbf{v}) + \lambda \operatorname{sign}(\mathbf{x}) \tag{27}$$

$$\partial_{\mathbf{v}} F_s(\mathbf{x}, \mathbf{v}) = \gamma \mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{x} - \mathbf{v}) - \lambda \operatorname{sign}(\mathbf{v}). \tag{28}$$

Then $\mathbf{0} \in \partial F_s(\mathbf{x}^{\text{opt}}, \mathbf{v}^{\text{opt}})$ is expressed as

$$\mathbf{0} \in \begin{bmatrix} \partial_{\mathbf{x}} F_s(\mathbf{x}, \mathbf{v}) \\ \partial_{\mathbf{v}} F_s(\mathbf{x}, \mathbf{v}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y}) - \gamma \mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{x} - \mathbf{v}) + \lambda \operatorname{sign}(\mathbf{x}) \\ \gamma \mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{x} - \mathbf{v}) - \lambda \operatorname{sign}(\mathbf{v}) \end{bmatrix}. \tag{29}$$

According to the FBS algorithm, we denote $P(\mathbf{x}, \mathbf{v})$ and $Q(\mathbf{x}, \mathbf{v})$ as

$$P(\mathbf{x}, \mathbf{v}) = \begin{bmatrix} (1-\gamma)\mathbf{A}^{\mathrm{T}}\mathbf{A} & \gamma \mathbf{A}^{\mathrm{T}}\mathbf{A} \\ -\gamma \mathbf{A}^{\mathrm{T}}\mathbf{A} & \gamma \mathbf{A}^{\mathrm{T}}\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix} - \begin{bmatrix} \mathbf{A}^{\mathrm{T}}\mathbf{y} \\ \mathbf{0} \end{bmatrix}, \tag{30}$$

$$Q(\mathbf{x}, \mathbf{v}) = \begin{bmatrix} \lambda \operatorname{sign}(\mathbf{x}) \\ \lambda \operatorname{sign}(\mathbf{v}) \end{bmatrix}. \tag{31}$$

Then the iteration process for FBS is as follows:

$$\begin{bmatrix} \mathbf{w}^{(k)} \\ \mathbf{u}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^{(k)} \\ \mathbf{v}^{(k)} \end{bmatrix} - \mu P(\mathbf{x}^{(k)}, \mathbf{v}^{(k)}) \tag{32}$$

$$\begin{bmatrix} \mathbf{x}^{(k+1)} \\ \mathbf{v}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \operatorname{soft}(\mathbf{w}^{(k)}; \mu\lambda) \\ \operatorname{soft}(\mathbf{u}^{(k)}; \mu\lambda) \end{bmatrix} \tag{33}$$

From this derivation, we can obtain the FBS algorithm shown in Algorithm 1.

## References

[1] J. Puigcorbe, A. de-Beaumont, Wind Turbine Gearbox Reliability, 2010, http://www.renewableenergyworld.com/articles/print/volume-13/issue-3/wind-power/wind-turbine-gearbox-reliability.html.
[2] W. Qiao, D. Lu, A survey on wind turbine condition monitoring and fault diagnosis-Part I: components and subsystems, IEEE Trans. Ind. Electron. 62 (10) (2015) 6536–6545.
[3] W. Qiao, D. Lu, A survey on wind turbine condition monitoring and fault diagnosis-Part II: signals and signal processing methods, IEEE Trans. Ind. Electron. 62 (10) (2015) 6546–6557.
[4] X. Chen, S. Wang, B. Qiao, Q. Chen, Basic research on machinery fault diagnostics: past, present, and future trends, Front. Mech. Eng. 13 (2) (2018) 264–291.
[5] R.B. Randall, Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications, John Wiley & Sons, 2011.
[6] H. Hong, M. Liang, Fault severity assessment for rolling element bearings using the Lempel-Ziv complexity and continuous wavelet transform, J. Sound Vib. 320 (1–2) (2009) 452–468.
[7] 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 Process. 1 (2) (1987) 173–183.
[8] P.D. Samuel, D.J. Pines, A review of vibration-based techniques for helicopter transmission diagnostics, J. Sound Vib. 282 (1–2) (2005) 475–508.
[9] R.B. Randall, A new method of modeling gear faults, J. Mech. Des. 104 (2) (1982) 259–267.
[10] C. Kar, A. Mohanty, Vibration and current transient monitoring for gearbox fault detection using multiresolution Fourier transform, J. Sound Vib. 311 (1–2) (2008) 109–132.
[11] J. Lin, M.J. Zuo, Extraction of periodic components for gearbox diagnosis combining wavelet filtering and cyclostationary analysis, J. Vib. Acoust. 126 (3) (2004) 449–451.
[12] R. Yan, R.X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: a review with applications, Signal Process. 96 (2014) 1–15.
[13] S. Wang, W. Huang, Z. Zhu, Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis, Mech. Syst. Signal Process. 25 (4) (2011) 1299–1320.
[14] Y. Lei, J. Lin, Z. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 35 (1–2) (2013) 108–126.
[15] J. Wang, R.X. Gao, R. Yan, Integration of EEMD and ICA for wind turbine gearbox diagnosis, Wind Energy 17 (5) (2014) 757–773.
[16] A. Parey, M. El Badaoui, F. Guillet, N. Tandon, Dynamic modelling of spur gear pair and application of empirical mode decomposition-based statistical analysis for early detection of localized tooth defect, J. Sound Vib. 294 (3) (2006) 547–561.
[17] J. Antoni, The spectral kurtosis: a useful tool for characterising non-stationary signals, Mech. Syst. Signal Process. 20 (2) (2006) 282–307.

[18] J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process. 20 (2) (2006) 308–331.
[19] Y. Wang, J. Xiang, R. Markert, M. Liang, Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: a review with applications, Mech. Syst. Signal Process. 66 (2016) 679–698.
[20] T. Barszcz, R.B. Randall, Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine, Mech. Syst. Signal Process. 23 (4) (2009) 1352–1365.
[21] J. Li, X. Chen, Z. Du, Z. Fang, Z. He, A new noise-controlled second-order enhanced stochastic resonance method with its application in wind turbine drivetrain fault diagnosis, Renew. Energy 60 (2013) 7–19.
[22] S. Lu, Q. He, H. Zhang, F. Kong, Enhanced rotating machine fault diagnosis based on time-delayed feedback stochastic resonance, J. Vib. Acoust. 137 (5) (2015) 051008.
[23] J. Li, X. Chen, Z. He, Multi-stable stochastic resonance and its application research on mechanical fault diagnosis, J. Sound Vib. 332 (22) (2013) 5999–6015.
[24] S. Wang, X. Chen, I.W. Selesnick, Y. Guo, C. Tong, X. Zhang, Matching synchrosqueezing transform: a useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis, Mech. Syst. Signal Process. 100 (2018) 242–288.
[25] S. Wang, X. Chen, G. Cai, B. Chen, X. Li, Z. He, Matching demodulation transform and synchrosqueezing in time-frequency analysis, IEEE Trans. Signal Process. 62 (1) (2014) 69–84.
[26] S. Wang, X. Chen, C. Tong, Z Zhao, Matching synchrosqueezing wavelet transform and application to aeroengine vibration monitoring, IEEE Trans. Instrument. Measure. 66 (2) (2017) 360–372.
[27] S. Wang, X. Chen, Y. Wang, G. Cai, B. Ding, X Zhang, Nonlinear squeezing time-frequency transform for weak signal detection, Signal Process. 113 (2015) 195–210.
[28] S. Wang, L. Yang, X. Chen, C. Tong, B. Ding, J. Xiang, Nonlinear squeezing time-frequency transform and application in rotor rub-impact fault diagnosis, J. Manuf. Sci. Eng. 139 (10) (2017) 101005.
[29] W. Yang, Z. Peng, K. Wei, P. Shi, W. Tian, Superiorities of variational mode decomposition over empirical mode decomposition particularly in time-frequency feature extraction and wind turbine condition monitoring, IET Renew. Power Gener. 11 (4) (2016) 443–452.
[30] L. Cui, N. Wu, C. Ma, H. Wang, Quantitative fault analysis of roller bearings based on a novel matching pursuit method with a new step-impulse dictionary, Mech. Syst. Signal Process. 68 (2016) 34–43.
[31] L. Cui, J. Wang, S. Lee, Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis, J. Sound Vib. 333 (10) (2014) 2840–2862.
[32] Y. Ding, W. He, B. Chen, Y. Zi, I.W. Selesnick, Detection of faults in rotating machinery using periodic time-frequency sparsity, J. Sound Vib. 382 (2016) 357–378.
[33] Z. Du, X. Chen, H. Zhang, R. Yan, Sparse feature identification based on union of redundant dictionary for wind turbine gearbox fault diagnosis, IEEE Trans. Ind. Electron. 62 (10) (2015) 6594–6605.
[34] G. He, K. Ding, H. Lin, Fault feature extraction of rolling element bearings using sparse representation, J. Sound Vib. 366 (2016) 514–527.
[35] L. Wang, G. Cai, W. You, W. Huang, Z.K. Zhu, Transients extraction based on averaged random orthogonal matching pursuit algorithm for machinery fault diagnosis, IEEE Trans. Instrument. Measure. 12 (66) (2017) 3237–3248.
[36] M. Elad, J.-L. Starck, P. Querre, D.L. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA), Appl. Comput. Harmon. Anal. 19 (3) (2015) 340–358.
[37] P. Abrial, Y. Moudden, J.-L. Starck, B. Afeyan, J. Bobin, J. Fadili, M. Nguyen, Morphological component analysis and inpainting on the sphere: application in physics and astrophysics, J. Fourier Anal. Appl. 13 (6) (2007) 729–748.
[38] X. Yong, R.K. Ward, G.E. Birch, Generalized morphological component analysis for EEG source separation and artifact removal, Neural Engineering, 2009, in: NER'09. 4th International IEEE/EMBS Conference on, IEEE, 2009, pp. 343–346.
[39] M. Farshchian, I. Selesnick, A. Parekh, Bird body and wing-beat radar Doppler signature separation using sparse optimization, compressed sensing theory and its applications to radar, sonar and remote sensing (CoSeRa), in: 2016 4th International Workshop on, IEEE, 2016, pp. 71–74.
[40] I.W. Selesnick, Resonance-based signal decomposition: a new sparsity-enabled signal analysis method, Signal Process. 91 (12) (2011) 2793–2809.
[41] G. Cai, X. Chen, Z. He, Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox, Mech. Syst. Signal Process. 41 (1–2) (2013) 34–53.
[42] Z. Li, Z. Peng, A new nonlinear blind source separation method with chaos indicators for decoupling diagnosis of hybrid failures: a marine propulsion gearbox case with a large speed variation, Chaos, Solit. Fractals 89 (2016) 27–39.
[43] D. Yu, M. Wang, X. Cheng, A method for the compound fault diagnosis of gearboxes based on morphological component analysis, Measurement 91 (2016) 519–531.
[44] D. Zhang, D. Yu, Multi-fault diagnosis of gearbox based on resonance-based signal sparse decomposition and comb filter, Measurement 103 (2017) 361–369.
[45] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (1) (2009) 34–81.
[46] I.W. Selesnick, I. Bayram, Enhanced sparsity by non-separable regularization, IEEE Trans. Signal Process. 64 (9) (2016) 2298–2313.
[47] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, 1987.
[48] M. Nikolova, Markovian reconstruction using a GNC approach, IEEE Trans. Image Process. 8 (9) (1999) 1204–1220.
[49] M. Nikolova, M.K. Ng, C.-P. Tam, Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction, IEEE Trans. Image Process. 19 (12) (2010) 3073–3088.
[50] W. He, Y. Ding, Y. Zi, I.W. Selesnick, Sparsity-based algorithm for detecting faults in rotating machines, Mech. Syst. Signal Process. 72 (2016) 46–64.
[51] I. Selesnick, Sparse regularization via convex analysis, IEEE Trans. Signal Process. 65 (17) (2017) 4481–4494.
[52] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122.
[53] M.A. Figueiredo, J.M. Bioucas-Dias, R.D. Nowak, Majorization-minimization algorithms for wavelet-based image restoration, IEEE Trans. Image Process. 16 (12) (2007) 2980–2991.
[54] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Model. Simul. 4 (4) (2005) 1168–1200.
[55] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer Science and Business Media, 2011.
[56] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, X. Chen, Nonconvex sparse regularization and convex optimization for bearing fault diagnosis, IEEE Trans. Ind. Electron. 65 (6) (2018) 7332–7342.
[57] H. Huang, N. Baddour, M. Liang, Auto-OBSD: automatic parameter selection for reliable oscillatory behavior-based signal decomposition with an application to bearing fault signature extraction, Mech. Syst. Signal Process. 86 (2017) 237–259.
[58] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[59] I. Selesnick, L1-norm Penalized Least Squares with SALSA, Connexions, 2014.