

Research Article

Sparse Representation of Transients Based on Wavelet Basis and Majorization-Minimization Algorithm for Machinery Fault Diagnosis

Wei Fan,¹ Gaigai Cai,^{1,2} Weiguo Huang,¹ Li Shang,³ and Zhongkui Zhu^{1,2}

¹ School of Urban Rail Transportation, Soochow University, Suzhou 215137, China

² State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China

³ College of Electronic Information, Suzhou Vocational University, Suzhou 215104, China

Correspondence should be addressed to Zhongkui Zhu; zkzhu@ustc.edu

Received 5 April 2014; Accepted 21 May 2014; Published 19 June 2014

Academic Editor: Ruqiang Yan

Copyright © 2014 Wei Fan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vibration signals captured from faulty mechanical components are often associated with transients which are significant for machinery fault diagnosis. However, the existence of strong background noise makes the detection of transients a basis pursuit denoising (BPD) problem, which is hard to be solved in explicit form. With sparse representation theory, this paper proposes a novel method for machinery fault diagnosis by combining the wavelet basis and majorization-minimization (MM) algorithm. This method converts transients hidden in the noisy signal into sparse coefficients; thus the transients can be detected sparsely. Simulated study concerning cyclic transient signals with different signal-to-noise ratio (SNR) shows that the effectiveness of this method. The comparison in the simulated study shows that the proposed method outperforms the method based on split augmented Lagrangian shrinkage algorithm (SALSA) in convergence and detection effect. Application in defective gearbox fault diagnosis shows the fault feature of gearbox can be sparsely and effectively detected. A further comparison between this method and the method based on SALSA shows the superiority of the proposed method in machinery fault diagnosis.

1. Introduction

Detection of transients has always been an important task in image and signal processing, and it has also been increasingly significant in the field of machinery fault diagnosis in recent years [1–3]. Most of the machinery is operated by means of gears and other rotating parts. Fault in these parts may lead to the failure of the whole machine and even loss of lives. Thus, proper analysis for transient detection from the machinery vibration signals has attracted sustained attention during the past decades [4–8].

Vibration signals derived from the defective gearboxes are generally observed as periodic transient impulses due to the rotating status [9]. Researches have shown that the transients in the vibration signals generated from a machinery component usually correspond to the localized fault or mechanical defect, such as flaking, crack, breakage, and fracture [10]. Therefore, it is very important to extract the

useful information, such as the feature of transient impulse and the cycle of multiple transient impulses, by analyzing the transients in the vibration signal [11, 12]. However, the strong background noise in practical vibration signals will corrupt the fault-induced transient impulses. Hence, detection of transients hidden in the noisy vibration signals is of great importance for machinery fault diagnosis.

Different methods have been proposed to extract the fault features from the vibration signals, such as wavelet analysis [13], time-frequency analysis [14, 15] and empirical mode decomposition (EMD) [16], and so forth. Meanwhile, another feature extraction method based on sparse representation is widely used in image processing, which has a close analogy with the effect of those fault feature extraction methods for 1-D signal.

Generally, transients to be detected can be described as a series of sparse coefficients multiplied by a certain wavelet basis as long as the signal being processed has

a sparse representation with respect to the known wavelet basis [17]. Hence these signal processing problems can be viewed as a series of restoration and reconstruction problems, which are classical linear inverse problems. However, most inverse problems are ill-posed [18, 19]; thus, these inverse problems can only be solved satisfactorily by adopting some sort of regularization. Based on sparse representation theory, a standard formulation for coping with these problems called basis pursuit denoising (BPD), consisting of a data fidelity term and a penalty term, has been proposed in [20]. One approach to derive suitable algorithms is based on the majorization-minimization (MM) method from the optimization theory. To apply the MM algorithm, either the data fidelity term (a quadratic function) or the penalty term (usually nonquadratic) can be majorized to handle this class of BPD problems. Typical approach for solving the former term is the so-called split augmented Lagrangian shrinkage algorithm (SALSA), which has been proposed recently to handle the image recovery problems [21]; while, for the latter term, an algorithm associated with an alternative quadratic function employing majorization-minimization also provides the solution to the corresponding inverse problems [22], called MM algorithm based on nonquadratic majorization.

SALSA, a recent algorithm to solve the BPD problem by using the Hessian of the data fidelity term, has better convergence properties than the earlier algorithms, such as iterative shrinkage/thresholding algorithm (ISTA) and fast ISTA (FISTA) [23, 24]. Based on variable splitting, this algorithm obtains an equivalent constrained optimization formulation, which is then addressed with an augmented Lagrangian method, and then, the convergence is guaranteed. The transient feature detection method based on SALSA has been increasingly popular at present. Sparsity-enabled signal decomposition method based on SALSA is used to decompose the vibration signal of faulty gearbox into high-oscillatory component and low-oscillatory component, which is employed by Cai et al. to extract the fault feature of gearbox [25]. Application of the sparse algorithm SALSA is employed by Selesnick to extract the frequency component of a sinusoidal signal [26]. However, some problems still remain. For instance, it is hard to select the optimal basis for signal representation to avoid the loss of useful components and the computation cost is large because two quadratic functions associated with nonconvex regularization functions should be minimized during each iteration of this algorithm.

Instead of utilization of variable splitting, the MM algorithm based on nonquadratic majorization utilizes a sequence of simpler convex optimization problems to replace the original ill-posed inverse problems yet is an effective and widely applicable method [27]. The algorithm has wide applications in image deconvolution or restoration under certain regularization due to its flexibility in the design of the sequence of simpler optimization problems so does in the field of speech denoising [22, 28–30]. However, there are still some remaining issues to be studied when applying the MM algorithm. One of the most important issues is how to select the optimal wavelet basis to satisfy the characteristics of the machinery vibration signals, which is very close to

the important issue of SALSA, and this is the vital point to improve the detection performance.

Considering that the wavelet basis of nonquadratic majorization-based MM algorithm in recent image processing is not suitable for feature extraction of gearbox fault vibration signal, and the selection of appropriate wavelet basis is the premise of the detection effect, a new method must be proposed to accommodate to the actual signal. Mass data show that the waveform of Morlet wavelet is in shape similar to signal transients caused by gearbox localized defects [3, 31]; thus, this paper proposes a new technique by synthesizing MM algorithm and Morlet wavelet basis. To implement the sparsity of the results, optimal wavelet basis is selected by applying correlation filtering, which has been widely applied in mechanical vibration signal processing [32]. The key issue to minimize the objective function of the ill-posed inverse problem is solved by nonquadratic majorization-based MM algorithm; thus a meaningful sparse coefficient would be got. Finally, the transients can be detected from the sparse coefficients, which are the key features for machinery fault diagnosis.

The remainder of the paper is organized as follows. In Section 2, the basic theoretical background concerning the proposed method is introduced. Section 3 gives a simulated study and analysis to verify the proposed method. Section 4 applies the presented method to the fault diagnosis of a gearbox in an automobile transmission gearbox, with a comparison to the extraction results of SALSA method. Finally, conclusions are drawn in Section 5.

2. Theory and Method

In transient detection problems, the goal is to detect the transients buried in noise from an observed signal. The transients can be represented as a series of sparse coefficients as long as the signal being analyzed has a sparse representation with respect to a known transformation. In this section, MM algorithm, solving ill-posed inverse problems, is introduced to model the transient detection problem. The following part provides a brief description on main steps of the MM-based transient detection method.

2.1. Mathematical Model of Transient Detection Problem.

Considering that the signal sampled from the transducer always has much noise, the observed signal $y(t)$ can be modeled as

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

where $y(t)$ is the sampled signal from the transducer, $x(t)$ is the true signal without noise, and $n(t)$ is the noise. The true signal $x(t)$ can be represented as a sparse linear combination of certain atoms. The representation of x can be expressed as $x = Ac$, where c is the vector of representation coefficients and also represents the transients; and the set of columns of A is a wavelet basis or dictionary; to be convenient, we also call A as wavelet basis. With this sparse representation, the estimation model (1) becomes

$$y(t) = Ac + n(t), \quad (2)$$

where A is an $N \times M$ matrix, $x = Ac$ is a length- N vector, and c is a length- M vector, with $M > N$. The more the similarity between the basis A and the signal x is, the sparser the vector c will be. Here assume that AA^* (where $(\cdot)^*$ denotes complex conjugate transpose) is invertible; therefore, the system of (2) has infinitely many solutions. As in many recent publications [21, 22, 26], we adopt the l_1 -norm regularizer to handle the ill-posed nature of the problem of inferring c and thus leads to the BPD problem:

$$\arg \min_c F(c) \quad \text{with } F(c) = \frac{1}{2} \|y - Ac\|_2^2 + \lambda \|c\|_1, \quad (3)$$

where $\|c\|_1$ is the l_1 -norm of M -length vector c , defined as $\|c\|_1 = \sum_{m=1}^M |c(m)|$, $\|c\|_2$ is the l_2 -norm of M -length vector c , defined as $\|c\|_2^2 = \sum_{m=1}^M |c(m)|^2$, and λ is the regularization parameter.

The function $F(c)$ cannot be easily minimized unless it is quadratic. To minimize $F(c)$, an iterative algorithm must be used. According to the MM algorithm from the optimization theory, instead of minimizing $F(c)$, the MM algorithm solves a sequence of simpler minimization problems:

$$c_{k+1} = \arg \min_c G_k(c), \quad (4)$$

where k is the iteration counter, $k = 1, 2, 3, \dots$. The MM algorithm asks that each function $G_k(c)$ should be a majorizer (upper bound) of $F(c)$ and that it coincides with $F(c)$ at $c = c_k$. That is

$$\begin{aligned} \forall c, G_k(c) &\geq F(c), \\ G_k(c_k) &= F(c_k). \end{aligned} \quad (5)$$

2.2. Majorization Iterative Algorithm Based on MM Approach. Traditional MM approaches to this problem consider three possible majorization strategies and thus lead to three different classes of algorithms as shown in [16]. As mentioned above, we can solve this problem easier if the function is quadratic.

Note that the data fidelity term $\|y - Ac\|_2^2$, presented in the function $F(c)$, is already a quadratic function. Then we just need to focus our attention on the penalty term $\|c\|_1$.

We mark penalty term $\|c\|_1$ as $\Psi(c)$, term $|c|$ as $\phi(c)$; then $\Psi(c) = \|c\|_1 = \sum_{m=1}^M |c(m)| = \sum_{m=1}^M \phi(c(m))$. $\phi(c)$ is an absolute value function and thus is nondifferentiable and nonstrictly convex, which makes (3) difficult to solve. According to the MM algorithm as shown in (5), we should find a quadratic function $g(c)$ to majorize $\phi(c)$ of the general form:

$$g(c) = mc^2 + nc + b, \quad (6)$$

where the parameters m , n and b are constants, the majorizer $g(c)$ should be the upper bound for $\phi(c)$ that coincides with $\phi(c)$ at a specified point c_k as shown in Figure 1. For this quadratic majorizer, conditions in (5) are equivalent to

$$\begin{aligned} g(c_k) &= \phi(c_k), \\ g'(c_k) &= \phi'(c_k). \end{aligned} \quad (7)$$

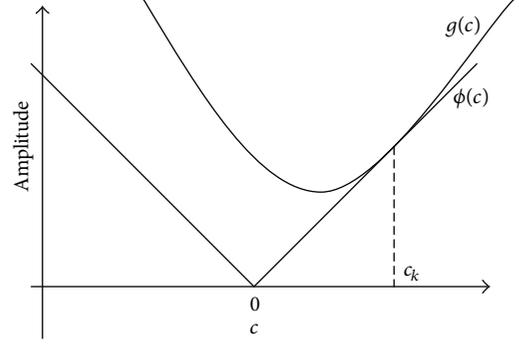


FIGURE 1: The penalty function and its quadratic majorizer.

Solving for m and b gives $m = (\phi'(c_k)/2c_k) - (n/2c_k)$, $b = \phi(c_k) - (c_k/2)\phi'(c_k) - (n/2)c_k$, thus leading to the majorizer $g(c)$ in (6) given by

$$\begin{aligned} g(c) &= \left(\frac{\phi'(c_k)}{2c_k} - \frac{n}{2c_k} \right) c^2 + nc \\ &+ \left(\phi(c_k) - \frac{c_k}{2} \phi'(c_k) - \frac{n}{2} c_k \right). \end{aligned} \quad (8)$$

Considering a special form of function $g(c)$, we set the unknown parameter $n = 0$; then the parameters m and b become $m = (\phi'(c_k)/2c_k)$, $b = \phi(c_k) - (c_k/2)\phi'(c_k)$, thus leading to the majorizer $g(c)$ in (8) given by

$$g(c) = \frac{\phi'(c_k)}{2c_k} c^2 + \phi(c_k) - \frac{c_k}{2} \phi'(c_k). \quad (9)$$

Considering the concrete form of $\phi(c)$, the function $G_k(c)$ can be written in a matrix format as

$$G_k(c) = \frac{1}{2} \|y - Ac\|_2^2 + \lambda \left(\frac{1}{2} c^* \Lambda_k^{-1} c + \frac{1}{2} \|c_k\|_1 \right), \quad (10)$$

where Λ_k denotes the diagonal matrix with vector $|c_k|$ along its diagonal,

$$\Lambda_k = \text{diag}(|c_k|) = \begin{bmatrix} |c_k(0)| & & \\ & \cdots & \\ & & |c_k(M-1)| \end{bmatrix}. \quad (11)$$

Then, the MM update (4) for c_k is

$$c_{k+1} = \arg \min_c \left[\frac{1}{2} \|y - Ac\|_2^2 + \lambda \left(\frac{1}{2} c^* \Lambda_k^{-1} c + \frac{1}{2} \|c_k\|_1 \right) \right]. \quad (12)$$

The last term of (12) can be omitted because it does not depend on c ; thus a new update equation also called cost function is got as

$$c_{k+1} = \arg \min_c \frac{1}{2} \|y - Ac\|_2^2 + \frac{\lambda}{2} c^* \Lambda_k^{-1} c. \quad (13)$$

Equation (13) is quadratic in c so the solution to this problem can be written explicitly using linear algebra as

$$c_{k+1} = (A^* A + \lambda \Lambda_k^{-1})^{-1} A^* y. \quad (14)$$

Although (14) is mathematically valid, there are still some remaining issues with this update. Due to the sparsity of c , components of c will go to zero as the iteration progresses. Thus the first issue is that as the elements of c go to zero, elements of Λ_k^{-1} would go to infinity, which will make the update numerically inaccurate. The second important issue is the selection of the optimal wavelet basis to ensure the sparsity of c when applying MM algorithm to transient detection problem.

The first issue can be avoided, as described in [22], by using the matrix inverse lemma, so the update equation can be written as

$$c_{k+1} = \frac{1}{\lambda} \Lambda_k \left[A^* y - A^* (A \Lambda_k A^* + \lambda I)^{-1} A \Lambda_k A^* y \right], \quad (15)$$

where λ can be chosen according to the regularization in [30].

The second important issue, the selection of the optimal wavelet basis, can be avoided by employing correlation filtering, which will be described in the next section.

2.3. Selection of Optimal Basis by Correlation Filtering. Morlet wavelet is one of the most popular nonorthogonal wavelets, defined in the time domain as a harmonic wave multiplied by a Gaussian time domain window:

$$\psi(t) = \exp\left(-\frac{\beta^2 t^2}{2}\right) \cos(\pi t). \quad (16)$$

It is a cosine signal that decays exponentially on both the left and the right sides. This feature makes it very similar to an impulse. It has been used for impulse isolation and mechanical fault diagnosis [3, 32–34].

Considering that the waveform of Morlet wavelet is in shape similar to the signal transients caused by gearbox localized defects at a constant speed, we incorporate Morlet wavelet into the above algorithm. The Morlet wavelet is firstly chosen as the basis.

The parametric formulation of Morlet wavelet is

$$\begin{aligned} \psi(f, \zeta, \tau, t) &= \psi_\gamma(t) \\ &= e^{(-\zeta/\sqrt{1-\zeta^2})[2\pi f(t-\tau)]^2} \cos(2\pi f(t-\tau)), \end{aligned} \quad (17)$$

where parameter vector $\gamma = (f, \zeta, \tau)$ determines the wavelet properties, these parameters (f, ζ, τ) are denoted frequency $f \in R^+$, damping ratio $\zeta \in [0, 1) \subset R^+$, and time index $\tau \in R$, respectively.

The discrete parameters f , ζ , and τ belong to the subsets of F , Z , and T_C , respectively:

$$\begin{aligned} F &= \{f_1, f_2, \dots, f_i\} \subset R^+, \\ Z &= \{\zeta_1, \zeta_2, \dots, \zeta_j\} \subset R^+ \cap [0, 1), \\ T_C &= \{\tau_1, \tau_2, \dots, \tau_k\} \subset R. \end{aligned} \quad (18)$$

With different parameters, dictionary of Morlet wavelet can be constructed as follows:

$$\begin{aligned} \Psi &= \{\psi_\gamma(t) : \gamma \in F \times Z \times T_C\} \\ &= \{\psi(f, \zeta, \tau, t) : f \in F, \zeta \in Z, \tau \in T_C\}, \end{aligned} \quad (19)$$

and each item in the dictionary is called an atom.

With the constructed dictionary, correlation filtering is exploited to identify the optimal set of parameters $(\bar{f}, \bar{\zeta}, \bar{\tau})$, which is the most similar to the original signal [32, 34]. Correlation, which shows the similarity between the basis and the original signal, can be measured by an inner product operation. Then, a correlation function c_γ is defined to quantify the correlation degree between the basis $\psi_\gamma(t)$ and the original signal $x(t)$;

$$c_\gamma = \cos \theta = \frac{|\langle \psi_\gamma(t), x(t) \rangle|}{\|\psi_\gamma(t)\|_2 \|x(t)\|_2}, \quad (20)$$

where θ is the angle between $\psi_\gamma(t)$ and $x(t)$, and the smaller the angle, the nearer the $\psi_\gamma(t)$ and $x(t)$ in space and the more resemble the $\psi_\gamma(t)$ and $x(t)$. Thus, the single similar wavelet function $\psi(t, \bar{f}, \bar{\zeta}, \bar{\tau})$ with optimal parameters $(\bar{f}, \bar{\zeta}, \bar{\tau})$ can be found by maximizing the correlation function c_γ at each time value from the constructed Morlet wavelet dictionary. Peaks of c_γ for a given time value τ can be represented as

$$k_\gamma(\tau) = \max_{f \in F, \zeta \in Z} c_\gamma = c(\bar{f}, \bar{\zeta}, \bar{\tau}), \quad (21)$$

which relates the wavelets with the strongest correlation to the signal, and the time index parameter $\bar{\tau}$ can be found by maximizing the coefficient $k_\gamma(\tau)$; thus the most similar wavelet function is $\psi(t, \bar{f}, \bar{\zeta}, \bar{\tau})$.

The optimal wavelet basis $\text{Mor}(t, \tau)$ with optimal parameters $\bar{f}, \bar{\zeta}$ based on $\psi(t, \bar{f}, \bar{\zeta}, \bar{\tau})$ replacing the above basis A in (15) can be constructed; thus the second important issue of the MM algorithm has been solved.

2.4. Sparse Representation of Transients Based on Wavelet Basis and MM Algorithm. Motivated by the merits of MM algorithm, a new sparse representation of transients method based on MM algorithm incorporated wavelet basis is proposed in this study. As the signal being analyzed can be represented sparsely on certain wavelet basis, the transients detection problem can be viewed as a basis pursuit denoising problem; thus the MM algorithm can be employed to handle it.

Considering that the faulty gearbox vibration signal always performs as periodic double-side attenuation functions due to the rotating nature, which is in shape similar to the Morlet wavelet, this paper inserts Morlet wavelet basis $\text{Mor}(t, \tau)$ selected by correlation filtering into the MM algorithm and thus yields the iterative algorithm as follows:

$$\begin{aligned} c_{k+1} &= \frac{1}{\lambda} \Lambda_k \left[\text{Mor}^* y - \text{Mor}^* (\text{Mor} \Lambda_k \text{Mor}^* + \lambda I)^{-1} \right. \\ &\quad \left. \times \text{Mor} \Lambda_k \text{Mor}^* y \right]. \end{aligned} \quad (22)$$

The updating of (22) can be implemented as follows:

- (1) set $k = 0$, choose λ , and initialize c ;
- (2) set iterative number Nit ;

TABLE 1: Procedures of these two methods: the proposed algorithm and wavelet-assisted SALSA algorithm.

The proposed algorithm	Wavelet-assisted SALSA Algorithm
(1) Set k, c_0, λ	(1) Set $k = 0, v_0, d_0, c_0, \lambda, \mu$
(2) Repeat	(2) Repeat
$\Lambda_k = \text{diag}(c_k)$	$c_{k+1} = \arg \min_c \ \text{Mor} \cdot c_k - y\ _2^2 + \mu \ c_k - v_k - d_k\ _2^2$
$c_{k+1} = \arg \min_c \left(\frac{1}{2} \ y - \text{Mor} \cdot c_k\ _2^2 + \left(\frac{\lambda}{2} \right) c_k^* \Lambda_k^{-1} c_k \right)$	$v_{k+1} = \arg \min_v \lambda \ v_k\ _1 + \left(\frac{\mu}{2} \right) \ c_{k+1} - v_k - d_k\ _2^2$
$k \leftarrow k + 1$	$d_{k+1} = d_k - (c_{k+1} - v_{k+1})$
until stopping criterion is satisfied.	$k \leftarrow k + 1$
	until stopping criterion is satisfied.

(3) repeat

$$g = \frac{1}{\lambda} \text{Mor}^* y$$

$$\Lambda_k = \text{diag}(c_k)$$

$$F = \lambda I + \text{Mor} \Lambda_k \text{Mor}^* \quad (23)$$

$$c_{k+1} = \Lambda_k (g - \text{Mor}^* F^{-1} \text{Mor} \Lambda_k g)$$

$$k \leftarrow k + 1$$

until stopping criterion is satisfied ($k > \text{Nit}$).

The matrix Mor has two parameters $\bar{f}, \bar{\zeta}$ as defined above. Then the reconstructed signal \hat{x} can be expressed as $\hat{x} = \text{Mor} \cdot \hat{c}$, where \hat{c} is the final vector with sparse coefficients.

To illustrate the superiority of the computation cost of the proposed method, the procedures of this method and another sparse algorithm SALSA also based on Morlet wavelet basis are shown in Table 1. As shown in Table 1, wavelet-assisted SALSA algorithm has two convex functions to be minimized, and one of them includes a nonstrictly convex function, which is very hard to minimize. However, there is only one function to be minimized in the proposed method, and the function is quadratic, which makes the proposed method less computation cost.

In summary, the procedure of the proposed transient sparse representation method is illustrated in Figure 2. After the accomplishment of vibration signal measurement, the procedures of transient detection can be described as the following major parts.

- (1) Given a signal $y(t)$ with N data points, find the most similar function $\psi(t, \bar{f}, \bar{\zeta}, \bar{\tau})$ by using correlation filtering based on correlation value maximization rule according to (20)-(21);
- (2) construct optimal wavelet basis $\text{Mor}(t, \tau)$ of size $N \times M$ by using the function $\psi(t, \bar{f}, \bar{\zeta}, \bar{\tau})$ above as the basis atom;
- (3) construct new transient sparse representation method based on MM algorithm through replacing A with the optimal basis $\text{Mor}(t, \tau)$ in (15); update the new method in (23), and a new data vector \hat{c} of size $M \times 1$ representing the transients in signal is generated.

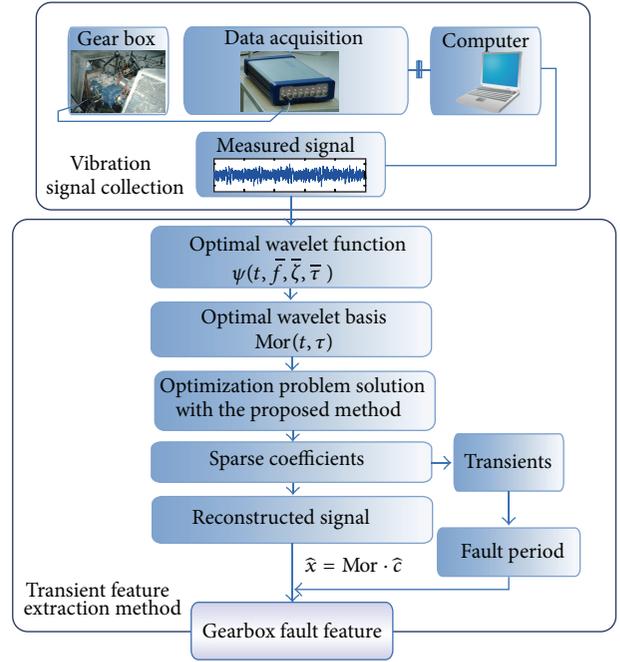


FIGURE 2: Procedure of the proposed transient sparse representation method.

Finally, the transients hidden in signal can be represented as a sparse vector \hat{c} , and the denoised signal \hat{x} is finally reconstructed by $\hat{x} = \text{Mor} \cdot \hat{c}$.

3. Simulation and Evaluation of the Proposed Method

To verify the effectiveness of the proposed method, a simulated vibration signal $y(t)$ is performed for transient feature extraction. Considering the characteristics of vibration signals of faulty gearbox, the simulated signal is constructed as

$$\begin{aligned}
 y(t) &= x(t) + n(t) \\
 &= \sum_k e^{(-\zeta_0/\sqrt{1-\zeta_0^2})[2\pi f_0(t-kT_0-\tau_0)]^2} \\
 &\quad \times \cos(2\pi f_0(t-kT_0-\tau_0)) + A_n n(t), \quad (24)
 \end{aligned}$$

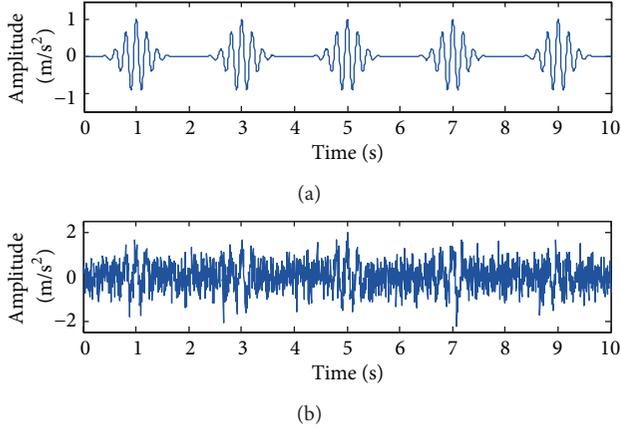


FIGURE 3: The simulated signal: (a) the original signal and (b) the noisy signal.

where frequency is $f_0 = 5$ Hz, $\zeta_0 = 0.01$ is the damping ratio, time index is $\tau_0 = 1$ s, and cyclic period is $T_0 = 2$ s. Obviously, $x(t)$ is a real periodic cyclic impulse response signal simulating the gear fault signal, and the signal $n(t)$ is white noise. The sampling frequency is 200 Hz in time ranges $[0, 10]$ s. Waveform of the simulated signal is shown in Figure 3, whose noise is weighted by $A_n = 0.5$. Figures 3(a) and 3(b) give the waveform of the simulated signal including noise or not, respectively.

The proposed transient sparse representation method is applied to analyze the simulated signal and extract the transients from the signal. According to the procedure in Figure 2, the first step is to calculate the optimal wavelet function by correlation filtering based on correlation value maximization rule. With this principle, parameters $\bar{f} = 5$ Hz, $\bar{\zeta} = 0.01$ are obtained for the optimal wavelet function. The parameters are exactly equal to the simulated values.

The second step is to cope with the objective function as described in (3) (with $\lambda = 8.457$). The optimal wavelet basis $\text{Mor}(t, \tau)$ with parameters \bar{f} , $\bar{\zeta}$ in the first step should be founded so as to realize the feature detection sparsely. With the optimal basis $\text{Mor}(t, \tau)$, the transients in the noisy signal are converted into sparse vector \hat{c}_{MM} (represents the sparse vector by using MM method) as illustrated in Figure 4(a) through the iterative algorithm shown in (23). The reconstructed signal is shown in Figure 4(c). For better comparison of the sparse vector and the transients in the reconstructed signal, the first N elements of vector \hat{c} are taken as shown in Figure 4. The following analysis will refer to this one, and not explain in the next section.

In Figure 4(a), cyclic period $T = 2$ s is identified, which is equal to the simulated value ($T_0 = 2$ s). The impulse times are 1 s, 3 s, 5 s, 7 s, and 9 s, respectively, which can also be identified from Figure 4(a). Compared to the original signal shown in Figure 3(a), it can be found that the reconstructed signal shows an excellent effect on transient detection as the transients are captured clearly, while the noise is well discarded.

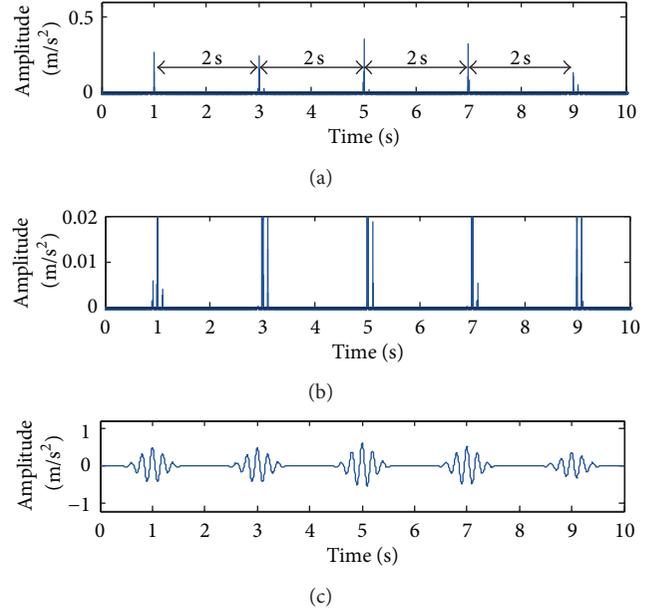


FIGURE 4: Detection results of the simulated signal by using the proposed method: (a) the sparse coefficients; (b) the amplified sparse coefficients, and (c) the reconstructed signal.

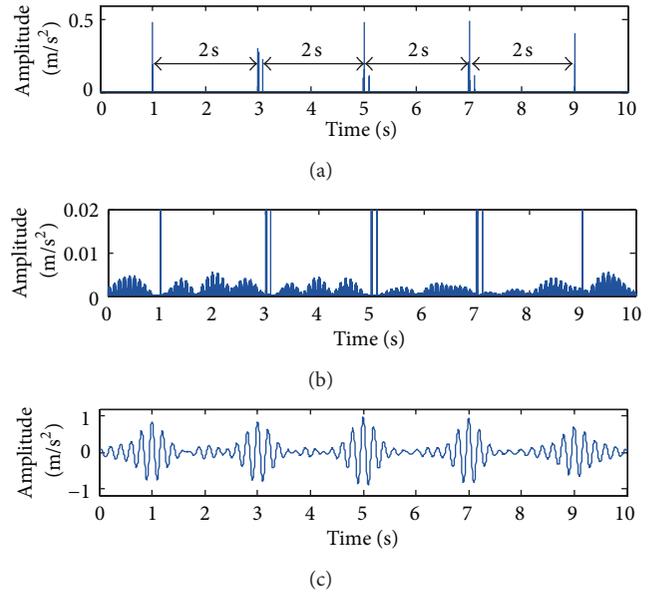


FIGURE 5: Detection results of the simulated signal by using SALSA method: (a) the sparse coefficients; (b) the amplified sparse coefficients, and (c) the reconstructed signal.

To confirm the superiority of this result, the same simulated signal is analyzed by SALSA method for comparison. Optimal Morlet wavelet basis is firstly constructed in the same way described in Section 2.3, and then SALSA algorithm is applied to deal with this detection problem. The detection results of SALSA method are illustrated in Figure 5. Figures 5(a) and 5(c) give the sparse vector \hat{c}_{SALSA} (represents the sparse vector by using SALSA) and the reconstructed signal, respectively.

TABLE 2: The analysis results of the proposed method at different noise amplitudes.

A_n	SNR (dB)	k_r	Impulse time (s)					Period parameter (s)
0	∞	1.000	1.000	3.000	5.000	7.000	9.000	2.000
0.1	10.054	1.000	1.000	3.000	5.000	7.000	9.000	2.000
0.2	4.033	0.998	1.000	3.000	5.000	7.000	9.000	2.000
0.3	0.512	0.997	1.000	3.000	5.000	7.000	9.000	2.000
0.4	-1.987	0.990	1.000	3.000	5.000	7.000	9.000	2.000
0.5	-3.925	0.989	1.000	3.000	5.000	7.000	9.000	2.000
0.6	-5.509	0.984	1.000	3.000	5.000	7.000	9.000	2.000
0.7	-6.848	0.969	1.000	3.000	5.000	7.000	8.995	1.999
0.8	-8.008	0.973	1.000	3.005	5.000	7.000	9.095	2.024
0.9	-9.031	0.965	1.000	3.005	5.000	7.000	9.095	2.024
1.0	-9.946	0.957	1.000	3.005	5.000	7.000	9.095	2.024
1.1	-10.774	0.946	0.995	3.095	5.000	7.000	9.095	2.000
1.2	-11.530	0.933	0.995	3.100	5.000	7.000	9.090	2.024

The period of the impulses can also be identified as marked in Figure 5(a) by SALSA method. It seems that SALSA method has a good performance for transient detection. However, some details about the analysis results are ignored. We amplify both the sparse coefficients \hat{c}_{MM} and \hat{c}_{SALSA} with their y -labels range $[0, 0.02]$ as shown in Figures 4(b) and 5(b), respectively. The sparsity of the vector can be measured by the l_0 -norm ($\|\cdot\|_0$) of it. We set the threshold value as $1e^{-10}$, then we get $\|\hat{c}_{MM}\|_0 = 71$, while we get $\|\hat{c}_{SALSA}\|_0 = 1031$. Thus, conclusions can be drawn that vector \hat{c}_{MM} is much sparser than vector \hat{c}_{SALSA} , which implies the proposed method has a better performance on transient sparse representation. Furthermore, their histories of the cost functions are illustrated in Figure 6, simultaneously. History of the cost function shown in (13) of the proposed method is illustrated in green dash-dot line, while the cost function history of the SALSA method [21] for this simulation is illustrated in blue dash line. Figure 6 shows the proposed method has a faster convergence than SALSA method.

Another simulated signal with larger noise ($A_n = 0.7$) as shown in Figure 7(a) is analyzed by these two methods, and the detection results of the proposed method (with $\lambda = 11.840$) and SALSA method are shown in Figures 7(b) and 7(c), respectively. Conclusions can be drawn that, with the increasing of noise level, detection ability of SALSA method fades away and even leads to faulty detection as shown in red circles in Figure 7(c), while the proposed method still has a good performance on extraction effect.

In order to test the noise tolerance of the proposed method, the simulation tests with different noise amplitudes A_n from (24) are investigated as shown in Table 2, in which noise amplitudes A_n , SNR, the correlation coefficients k_r between the reconstructed signal and the original signal, the impulse time, and the period parameters are listed.

Correlation coefficient k_r , used to measure the effect of the construction, also can be written as:

$$k_r = \frac{|\langle x, \hat{x} \rangle|}{\|x\|_2 \|\hat{x}\|_2}, \quad (25)$$

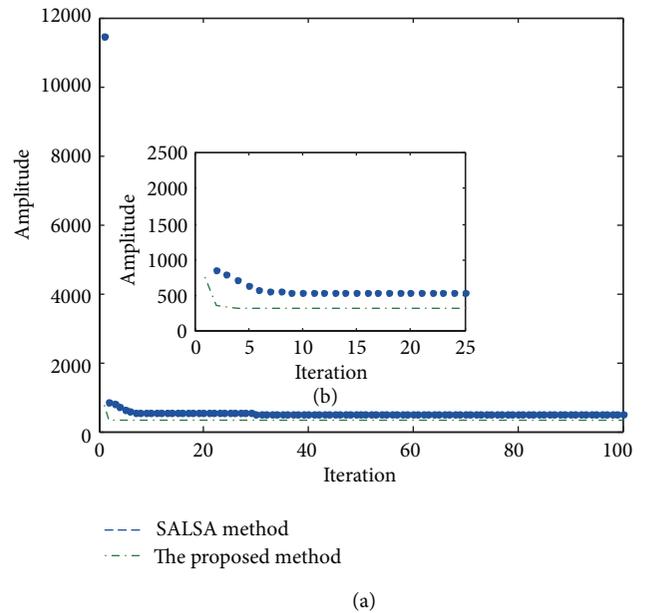


FIGURE 6: The cost function histories of the proposed method and SALSA method.

where x and \hat{x} are the original component and the corresponding estimate result, respectively.

SNR, the signal-to-noise ratio, used to weigh the noise level is defined as

$$\text{SNR} = 10 \times \lg \left(\frac{P_s}{P_n} \right), \quad (26)$$

where P_s is the energy of the useful information $x(t)$ and P_n is the energy of the noise $n(t)$. As shown in Table 2, it can be seen that with the increasing of the noise amplitude, the proposed method still has an excellent detection effect.

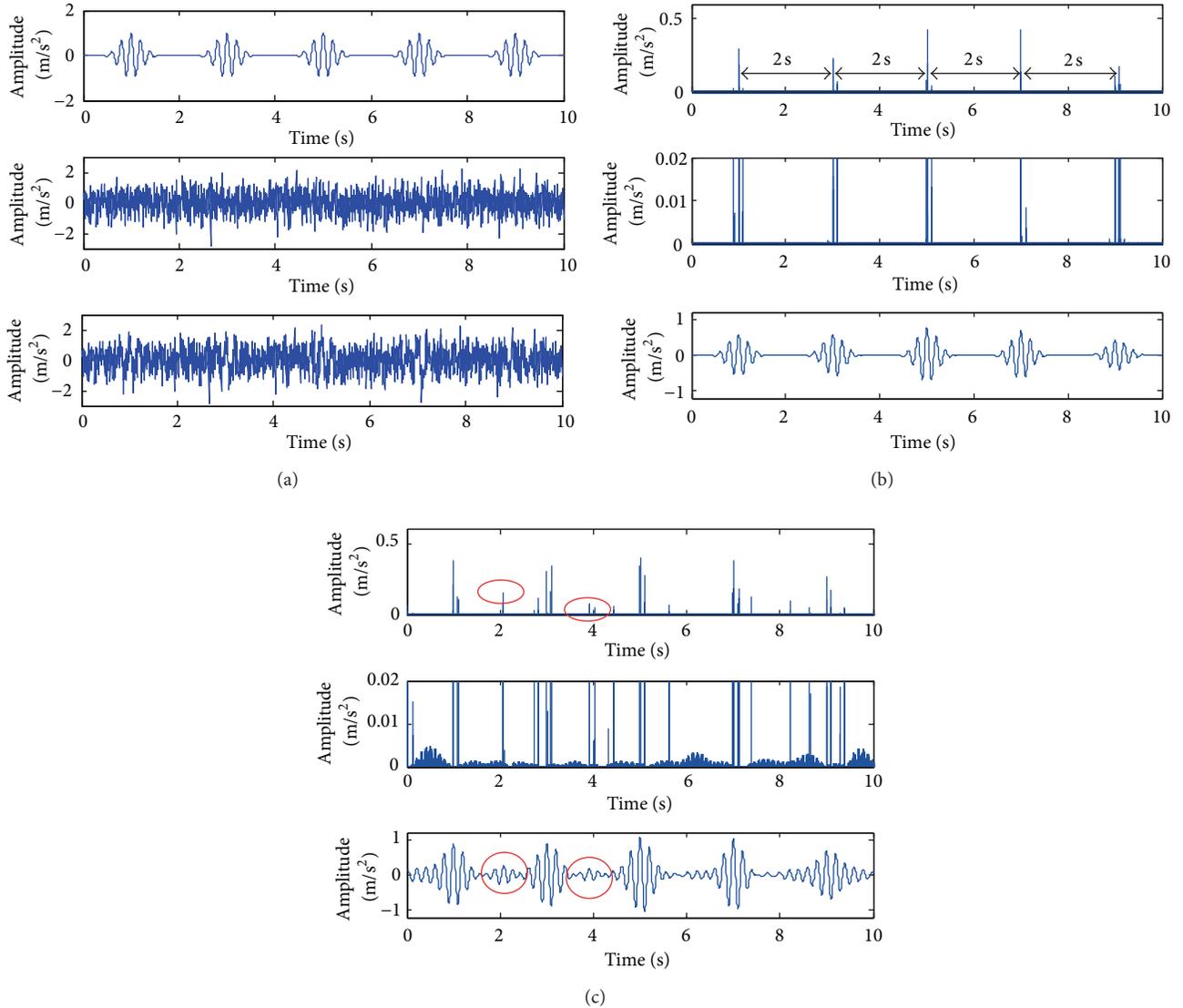


FIGURE 7: Detection results of these two methods: (a) the simulated signal with much noise; (b) the proposed method, and (c) the SALS method.

4. Experimental Verification

To verify the effectiveness of the proposed method in practical engineering applications, defective gearbox data is analyzed. The experimental data were acquired from an automobile transmission gearbox, which has five forward speeds and one backward speed, as shown in Figure 8. At constant rotating speed, localized faults in gearbox tend to result in periodic shocks and thereby arouse periodic transients in vibration signals, so the transients can be represented as certain wavelet basis sparsely; thus the fault features can be extracted sparsely by the proposed method. To further demonstrate the effectiveness of the proposed method, the results of fault feature detection by the proposed method are compared with the detection results of SALS method.

During the test, a broken-tooth fault occurred on the driving gear of the third speed. The vibration signal was acquired by an accelerometer mounted on the outer case of

the gearbox when it is loaded on the third gearbox. For a gear transmission, the meshing frequency f_m is calculated by

$$f_m = \frac{nz}{60i}, \quad (27)$$

where z is the number of the gear teeth, n is the rotating speed of the input shaft, and i is the transmission ratio. The working parameters are shown in Table 3, the meshing frequency here is 500 Hz, and thus the sampling frequency is set at 3000 Hz.

A measured vibration signal caused by one driving gear teeth broken with a length of 900 and its frequency spectrum are shown in Figures 9(a) and 9(b). From Figure 9(a), the impulse period cannot be identified as the noise corruption; from Figure 9(b), the main frequency component can be identified as 500 Hz.

The proposed transient sparse representation method is then applied to the signal. The optimal wavelet function is first calculated by the correlation filtering rule, and the associated

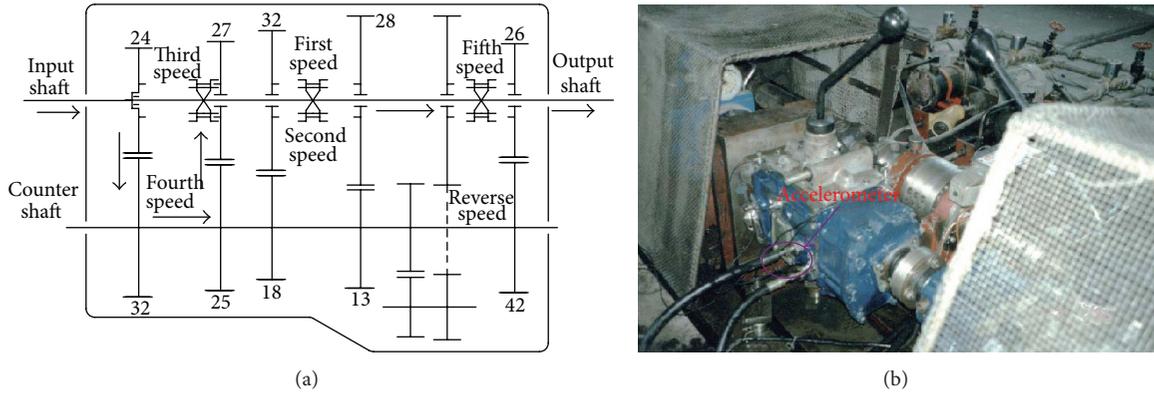


FIGURE 8: The automobile transmission gearbox: (a) structure of gearbox, and (b) gearbox setup.

TABLE 3: Working parameters of the third speed gears.

	Number of teeth	Rotating period (s)	Rotating frequency (Hz)	Meshing frequency (Hz)
Driving gear	25	0.050	20	500
Driven gear	27	0.054	18.5	

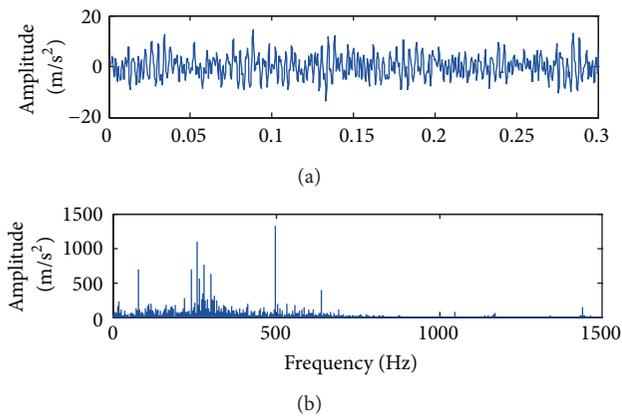


FIGURE 9: (a) The measured gearbox defective vibration signal, and (b) its Fourier spectrum.

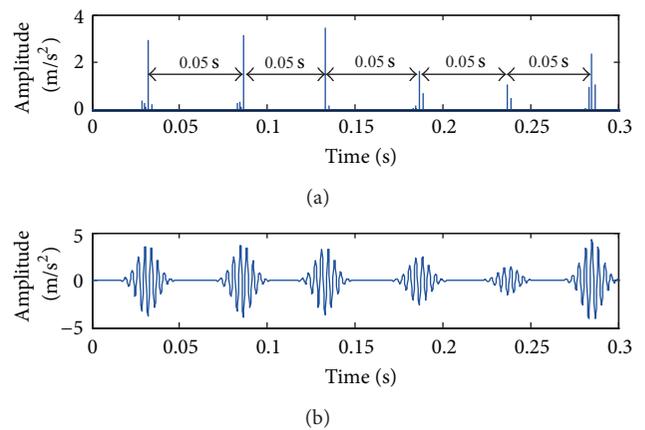


FIGURE 10: The analysis results of vibration signal by using the proposed method: (a) sparse coefficients, and (b) the reconstructed signal.

parameters are $\bar{f} = 272 \text{ Hz}$, $\bar{\zeta} = 0.0074$, and $\bar{\tau} = 0.0633 \text{ s}$. Then the optimal wavelet basis $\text{Mor}(t, \tau)$ with parameters \bar{f} and $\bar{\zeta}$ is founded so as to realize the feature detection sparsely. Figure 10(a), obtained by the proposed method, gives the sparse coefficients, which represents a series of periodic impulses. The average time period of these impulses is around 0.0505 s, which is very close to the theoretical value of 0.050 s. Figure 10(b) gives the reconstructed signal, whose periodical features represent the localized fault existing in the driving gear of the third speed. The analysis results confirm that the proposed transient sparse representation method can detect the transients clearly and reduce the noise effectively; thus the fault features can be identified.

For the purpose of comparison, the same signal is also processed by using SALSA method, and the results are displayed in Figure 11.

The sparse coefficients obtained by SALSA method are illustrated in Figure 11(a), which theoretically represent a series of periodic impulses. However, from Figure 11(a), the fault period cannot be identified clearly. The reconstructed signal is shown in Figure 11(b). The analysis result in Figure 11(b) has almost no periodical features, thus it is difficult for gearbox fault diagnosis. By comparing with the analysis results in Figure 10, it can be found that the proposed transient sparse representation method is obviously superior to SALSA method.

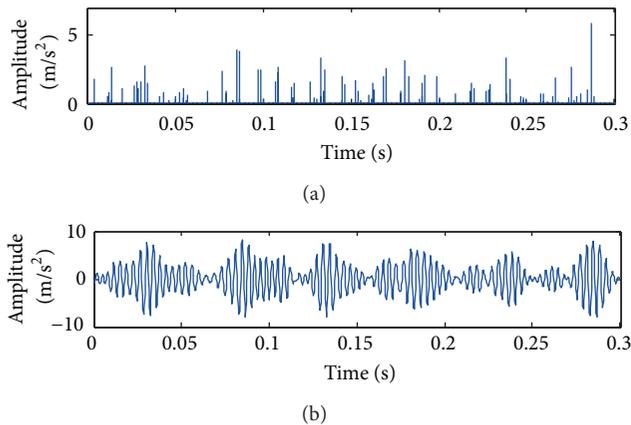


FIGURE 11: The analysis results of vibration signal by using SALS method: (a) sparse coefficients, and (b) the reconstructed signal.

5. Conclusions

This paper presents a novel method for transient sparse representation by employing MM algorithm combining wavelet basis to cope with the ill-posed inverse problems. Based on the discussion above, the following conclusions can be drawn.

- (1) The proposed method inherits the merits of MM algorithm in noise suppression and sparsity properties to represent the transient features sparsely. Not only is the strong background noise reduced by this method, but also the transients are represented as a series of sparse coefficients, which are of great importance for transient feature detection.
- (2) Identification of impulse time and the period parameter based on the proposed method is effective for feature detection, which is demonstrated by the simulated study. Furthermore, the performance of the proposed method has been verified by the application in the defective gearbox data in comparison with the SALS method. Both the simulation and the application show the proposed method is reasonable and effective in transient detection and thus for machinery fault diagnosis.

Due to the flexibility in the design of the sequence of simpler optimization problems in MM algorithm, the further research may be mainly concentrated on the improvement of sparse algorithm to decrease the computational complexity. Moreover, future research directions will also include repeating the experiments with a wider variety of bases, test signals and other sparsity methods.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China (nos. 51375322, and 61373098) and also partly supported by the Open Fund of State Key Laboratory for Manufacturing System Engineering (Xi'an Jiaotong University) (no. Sklms2011006).

References

- [1] M. Nixon and A. S. Aguado, *Feature Extraction & Image Processing*, Academic Press, Oxford, UK, 2008.
- [2] R. X. Gao and R. Q. Yan, "Non-stationary signal processing for bearing health monitoring," *International Journal of Manufacturing Research*, vol. 1, no. 1, pp. 18–40, 2006.
- [3] J. Lin and L. S. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of Sound and Vibration*, vol. 234, no. 1, pp. 135–148, 2000.
- [4] Z. K. Zhu, R. Q. Yan, L. Luo, Z. H. Feng, and F. R. Kong, "Detection of signal transients based on wavelet and statistics for machine fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1076–1097, 2009.
- [5] D. Wang, W. T. Peter, W. Guo, and Q. Miao, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis," *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
- [6] H. Liu, J. Wang, and C. Lu, "Rolling bearing fault detection based on the teager energy operator and elman neural network," *Mathematical Problems in Engineering*, vol. 2013, Article ID 498385, 10 pages, 2013.
- [7] Y. G. Lei, J. Lin, Z. J. He, and D. Kong, "A method based on multi-sensor data fusion for fault detection of planetary gearboxes," *Sensors*, vol. 12, no. 2, pp. 2005–2017, 2012.
- [8] R. Q. Yan, R. X. Gao, and X. F. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing A*, vol. 96, pp. 1–15, 2014.
- [9] He, Z. j, Y. Y. Zi, Q. F. Meng, and J. Zhao, *Fault Diagnosis Principle of Non-Stationary Signal and Applications to Mechanical Equipment*, Higher Education Press, Beijing, China, 2001.
- [10] Q. B. He, X. X. Wang, and Q. Zhou, "Vibration sensor data denoising using a time-frequency manifold for machinery fault diagnosis," *Sensors*, vol. 14, no. 1, pp. 382–402, 2013.
- [11] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics-A tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.
- [12] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [13] X. F. Chen, X. Li, S. B. Wang, Z. B. Yang, B. Q. Chen, and Z. J. He, "Composite damage detection based on redundant second-generation wavelet transform and fractal dimension tomography algorithm of lamb wave," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 1354–1363, 2013.
- [14] S. B. Wang, X. F. Chen, and G. G. Cai, "Matching demodulation transform and synchrosqueezing in time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 69–84, 2013.

- [15] S. B. Wang, X. F. Chen, G. Y. Li, X. Li, and Z. J. He, "Matching demodulation transform with applications to feature extraction of rotor rub-impact fault," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1372–1383, 2013.
- [16] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [17] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [18] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "A fast algorithm for the constrained formulation of compressive image reconstruction and other linear inverse problems," in *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 4034–4037, March 2010.
- [19] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.
- [20] I. Selesnick, "Penalty and shrinkage functions for sparse signal processing," 2014, <http://cnx.org/content/m45134>.
- [21] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.
- [22] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980–2991, 2007.
- [23] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [24] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [25] G. Cai, X. Chen, and Z. He, "Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox," *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 34–53, 2013.
- [26] I. Selesnick, "Introduction to sparsity in signal processing," 2014, <http://cnx.org/content/m43545/1.3>.
- [27] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
- [28] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Adaptive total variation image deconvolution: a majorization-minimization approach," in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO '06)*, pp. 1–4, September 2006.
- [29] M. A. T. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: a new majorization-minimization algorithm and an experimental comparison with wavelet denoising," in *Proceedings of the 2006 IEEE International Conference on Image Processing (ICIP '06)*, pp. 2633–2636, October 2006.
- [30] P. Y. Chen and I. W. Selesnick, "Translation-invariant shrinkage/thresholding of group sparse signals," *Signal Processing*, vol. 94, no. 1, pp. 476–489, 2014.
- [31] J. Lin and M. J. Zuo, "Gearbox fault diagnosis using adaptive wavelet filter," *Mechanical Systems and Signal Processing*, vol. 17, no. 6, pp. 1259–1269, 2003.
- [32] S. Wang, W. Huang, and Z. K. Zhu, "Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1299–1320, 2011.
- [33] Q. Miao, C. Tang, W. Liang, and M. Pecht, "Health assessment of cooling fan bearings using wavelet-based filtering," *Sensors*, vol. 13, no. 1, pp. 274–291, 2013.
- [34] L. C. Freudinger, R. Lind, and M. J. Brenner, "Correlation filtering of modal dynamics using the Laplace wavelet," in *Proceedings of the 1998 16th International Modal Analysis Conference (IMAC '98)*, vol. 2, pp. 868–877, February 1998.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

