Nonconvex Sparse Regularization and Convex Optimization for Bearing Fault Diagnosis

Shibin Wang, *Member, IEEE*, Ivan Selesnick, *Fellow, IEEE*, Gaigai Cai, *Member, IEEE*, Yining Feng, Xin Sui, and Xuefeng Chen, *Member, IEEE*

***Abstract*—Vibration monitoring is one of the most effective ways for bearing fault diagnosis, and a challenge is how to accurately estimate bearing fault signals from noisy vibration signals. In this paper, a nonconvex sparse regularization method for bearing fault diagnosis is proposed based on the generalized minimax-concave (GMC) penalty, which maintains the convexity of the sparsity-regularized least squares cost function, and thus the global minimum can be solved by convex optimization algorithms. Furthermore, we introduce a k-sparsity strategy for the adaptive selection of the regularization parameter. The main advantage over conventional filtering methods is that GMC can better preserve the bearing fault signal while reducing the interference of noise and other components; thus, it can significantly improve the estimation accuracy of the bearing fault signal. A simulation study and two run-to-failure experiments verify the effectiveness of GMC in the diagnosis of localized faults in rolling bearings, and the comparison studies show that GMC provides more accurate estimation results than L1-norm regularization and spectral kurtosis.**

***Index Terms*—Bearing fault diagnosis, convex optimization, condition monitoring, generalized minimax-concave (GMC) penalty, nonconvex sparse regularization (NSR).**

I. Introduction

CONDITION monitoring is the process of monitoring physical quantities related to machinery during its operation, in order to identify changes that indicate developing faults, which is the basis of condition-based maintenance. Diagnosis and prognosis are two important aspects of a condition-based maintenance program. Condition monitoring techniques are used to investigate a wide range of machinery, such as wind turbines [1], aero-engines [2]–[4], and high-speed trains [5]. Rolling bearings are extensively used in all types of rotating machinery and often operate under nonideal conditions; thus, minor problems sometimes cause bearings to fail quickly [6]–[13]. Vibration-based techniques have proved to be one of the most effective ways for bearing fault diagnosis among various techniques [14], [15], and thus have become popular.

Bearing fault signals with localized faults can be modeled as pseudo-cyclostationary [16], and measured vibration signals with localized bearing faults always contain series of short transients, which occur periodically (at least quasi-periodically with random slip) [15]. Moreover, measured vibration signals always combine with interference components or noise in the time domain and frequency domain. Therefore, bearing fault diagnosis requires fault feature extraction techniques, and an effective method is signal filtering. After bearing fault signals are extracted, an envelope spectrum is calculated to indicate bearing faults by comparing the fault characteristic frequencies [17], [18]. However, a challenge is how to accurately estimate bearing fault signals from noisy vibration signals. Several signal filtering methods have recently been studied for bearing fault diagnosis, such as spectral kurtosis (SK) [19]–[21] and wavelet transforms [22]. In the frequency domain, the noise is distributed over the whole frequency range, while bearing fault signals are in a specific frequency range, which motivates the envelope technique for bearing fault diagnosis. SK is a statistical tool that can indicate the presence of bearing fault signals and their locations in the frequency domain, even when they are buried in a strong noise [19]. Similarly, in the wavelet domain, the noise is distributed throughout, while bearing fault signals are concentrated in a few large-amplitude wavelet coefficients (i.e., wavelet transform provides sparse representation for bearing fault signals). Therefore, the bearing fault feature can be extracted or identified by these wavelet coefficients. These studies provide insights on how to extract fault features for bearing fault diagnosis. However, the denoising performance of conventional filtering methods heavily depends on the characteristics of measured signals. If the measured signals are complex or the noise is too strong, then the denoising performance is hindered. For example, some interference or noise may be present after denoising,

Manuscript received September 19, 2017; revised November 16, 2017 and December 19, 2017; accepted December 27, 2017. Date of publication January 15, 2018; date of current version May 1, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 51605366 and Grant 51421004, in part by the National Key Basic Research Program of China under Grant 2015CB057400, in part by the China Postdoctoral Science Foundation under Grant 2016M590937 and Grant 2017T100740, and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding author: Xuefeng Chen.)*

S. Wang is with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 10003 USA (e-mail: wangshibin2008@gmail.com).

I. Selesnick, Y. Feng, and X. Sui are with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 10003 USA (e-mail: selesi@nyu.com; yf889@nyu.edu; xs405@nyu.edu).

G. Cai is with the School of Rail Transportation, Soochow University, Suzhou 215131, China (e-mail: gaigaicai@gmail.com).

X. Chen is with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: chenxf@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2018.2793271

or the signal of interest is attenuated or annihilated, which may cause missed alarms or may underestimate the fault severity for bearing fault diagnosis, especially in an early stage of faults [23], [24]. How to improve the denoising performance of filtering methods to provide better noise filtering capability and accurately estimate bearing fault signals is a key problem for bearing fault diagnosis.

Sparsity-assisted filtering methods have been recently studied for condition monitoring and machine fault diagnosis [25]–[30]. L1-norm regularization methods have become widely used in various applications due to the convexity of the L1 norm. However, L1-norm solutions often underestimate the high-amplitude components, yet these comprise the signal of interest in most cases. Thus, L1-norm solutions are not ideal for certain applications, including bearing fault diagnosis. Moreover, one of the key problems is how to adaptively set the regularization parameter to balance data fidelity and sparsity in the sparsity-regularized least squares cost function: a larger value of the regularization parameter makes the optimal solution sparser and less noisy, but reduces the coefficient amplitudes; a smaller value increases the coefficient amplitudes of the optimal solution, but makes the solution less sparse and noisier.

In order to enhance sparsity and overcome the underestimation characteristic of L1-norm regularization, many nonconvex sparse regularization (NSR) methods have been proposed and widely used for compressive sensing and signal processing [31]–[34]. However, most NSR methods forgo the convexity of the cost function; thus, the cost function to be minimized generally has extraneous local minimizers. NSR that maintains the convexity of the cost function has been recently studied, to capture the advantages of both nonconvex regularization and convex optimization [35], [36].

In this paper, we propose an NSR method for bearing fault diagnosis, which uses a nonconvex penalty and simultaneously maintains the convexity of the least squares cost function to be minimized. Thus, we solve convex optimization problems for their global minimizers to obtain sparse solutions. The main advantage compared to conventional filtering methods is that the proposed method can better preserve the bearing fault signal while reducing noise and other interference components; thus, it can significantly improve the estimation accuracy of the bearing fault signal. The contribution of this paper is twofold. First, we use the generalized minimax-concave (GMC) penalty [36] as a nonconvex penalty that maintains the convexity of the sparsity-regularized cost function, which we minimize using a forward–backward splitting algorithm. Second, we introduce a k-sparsity strategy for the adaptive selection of the regularization parameter. We present a simulation study and two experiment studies to validate the effectiveness of the proposed method in the diagnosis of localized faults in rolling bearings.

This paper is organized as follows. Section II recalls the sparsity formulation. In Section III, we propose the NSR method based on the GMC penalty (the GMC method for short) and introduce the k-sparsity strategy for adaptive parameter selection. Section IV presents the simulation study, and Section V presents application verifications for localized bearing fault diagnosis. Finally, Section VI presents the conclusion.

II. Review of Sparsity Formulation

In this section, we briefly review the formulation of the sparse signal model, L1-norm regularization, and the tunable Q-factor wavelet transform (TQWT).

A. Sparse Signal Model and L1-Norm Penalty

Vibration signals measured from machinery with a faulty bearing generally contain bearing fault signals and noise [15]. Thus, the noisy vibration signal $y \in \mathbb{R}^M$ can be modeled as

$$y = y_0 + n \tag{1}$$

where $y_0 \in \mathbb{R}^M$ represents the bearing fault signal, which is repetitive transients caused by localized bearing faults, and $n \in \mathbb{R}^M$ is white Gaussian noise and other unwanted components. A challenge of bearing fault diagnosis is how to accurately estimate the bearing fault signal y_0 from the noisy vibration signal y.

Research works show that the bearing fault signal y_0 admits a sparse representation with respect to an invertible linear transformation [25]–[27], i.e.,

$$y_0 = Ax \quad \text{and} \quad x = A^{\mathrm{T}} y_0 \tag{2}$$

where the matrix $A \in \mathbb{R}^{M \times N}$ represents the linear transformation, and $x \in R^N$ represents a sparse set of transform coefficients. Here, A^{T} represents the transpose of the matrix A. We usually have $M < N$, i.e., the transform coefficients outnumber the signal values.

Sparse regularization is used to find the sparse approximate solution x from the noisy signal y. The L1 norm is usually used as a penalty (regularizer) for sparse approximation and representation, since it induces sparsity most effectively among convex penalties. The L1-norm-regularized least squared problem is

$$\min_x \left\{ F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda \|x\|_1 \right\} \tag{3}$$

where $\|x\|_1 = \sum_n |x_n|$ is the L1 norm of $x \in \mathbb{R}^N$ and $\lambda > 0$ is the regularization parameter, which balances data fidelity and sparsity. Problem (3) is known as basis pursuit denoising [37] and lasso [38], which is a convex optimization problem, and extensive optimization methods have been developed to solve such problems [39], including a forward–backward splitting algorithm, alternating direction method of multipliers, etc. Algorithm 1 is the example of the forward–backward splitting algorithm for solving problem (3), where the notation $\|A\|_2$ denotes the square root of the maximum eigenvalue of $A^{\mathrm{T}}A$, and the soft-thresholding operator is defined as

$$\mathrm{soft}(x; \lambda) = x \cdot \frac{\max\{|x| - \lambda, 0\}}{\max\{|x|, \lambda\}} \tag{4}$$

which is element-by-element for any vector input x.

There are two problems when applying the L1-norm method for bearing fault diagnosis and other sparse filtering purposes. First, L1-norm penalty always underestimates of the signal of interest. Second, how to adaptively select the regularization parameter to balance data fidelity and sparsity is a key problem. In

Algorithm 1: Iterative algorithm for (3).

Initialize $x^{(0)}$, $0 < \mu < \|A\|_2^{-2}$
for $i = 0, 1, 2, \ldots$ **do**
  $w^{(i)} = x^{(i)} - \mu A^{\mathrm{T}}(Ax^{(i)} - y)$,
  $x^{(i+1)} = \mathrm{soft}(w^{(i)}; \lambda\mu)$,
end for
return $x^{(i+1)}$

the next section, we propose the GMC method to avoid the underestimation of the L1-norm method and provide a k-sparsity strategy for adaptive parameter selection.

B. TQWT for Sparse Representation

The linear transformation A is important for sparse representation. An effective transformation can promote the sparsity of coefficients. For bearing fault diagnosis, we use TQWT as the linear transformation A in this paper. The reason is threefold. First, TQWT can effectively represent a periodic or quasi-periodic impulse response, as demonstrated by recent research on the application of the TQWT to machine fault diagnosis [28], [40]. Second, TQWT is a type of discrete wavelet transform that is well-suited for the fast solution of sparse regularized inverse problems, because it is a tight frame, easily invertible, and efficiently implemented using radix-2 fast Fourier transforms (FFTs). Third, the advantage of TQWT over classic discrete wavelet transforms is that the Q-factor Q and asymptotic redundancy (over-sampling rate r) of the transform are easily and independently specified. The oscillatory behavior of the wavelet can be chosen to match the oscillatory behavior of the signal of interest, so as to enhance the sparsity of a sparse signal representation. The detailed description of TQWT can be found in [41].

For the sake of consistency with sparse formulations, we denote TQWT by A^{T} and inverse TQWT by A. It is noted that TQWT is not a "uniform" tight frame, which means that the wavelet functions do not all have the same L2 norm. That is, the rows of A do not all have the same L2 norm. Thus, considering this fact, we set the parameter λ as

$$\lambda_j = \theta\|\psi_j\|_2, \quad j = 1, 2, \ldots, J+1 \tag{5}$$

where $\psi_j, j = 1, 2, \ldots, J$ is the wavelet function in scale j, and ψ_{J+1} is the scaling function in scale J.

III. GMC Regularization Method and k-Sparsity Strategy

In order to enhance sparsity, many NSR methods have been proposed, and a general way to model the problem is the nonconvex regularization least squares problem:

$$\min_x \left\{ F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\psi(x) \right\} \tag{6}$$

where $\psi : \mathbb{R}^N \to \mathbb{R}$ is a nonconvex sparsity-inducing penalty. In this section, we use the nonconvex GMC penalty [36] and analyze the convexity of the GMC-regularized cost function. Moreover, we introduce a k-sparsity strategy to solve a key problem of sparse regularization, i.e., how to adaptively set the regularization parameter λ.

A. GMC Penalty

We describe a particular nonconvex sparsity-inducing penalty called GMC penalty, which maintains the convexity of the cost function, thereby avoiding the presence of spurious local minima in the cost function (6).

The minimax-concave penalty [42] with parameter $\gamma > 0$ can be defined as

$$\phi(x) = \begin{cases} |x| - \frac{\gamma}{2}x^2, & |x| < \frac{1}{\gamma} \\ \frac{1}{2\gamma}, & |x| \geqslant \frac{1}{\gamma}. \end{cases} \tag{7}$$

A separable multivariate minimax concave penalty may be defined as $\psi_{\mathrm{sep}}(x) = \sum_{n=1}^{N} \phi(x_n)$. We can rewrite the penalty as a Moreau envelope [43]

$$\psi_{\mathrm{MC}}(x) = \|x\|_1 - \min_v \left\{ \|v\|_1 + \frac{\gamma}{2}\|x - v\|_2^2 \right\}. \tag{8}$$

Using ψ_{MC} as the penalty in (6) gives

$$\begin{aligned} F(x) &= \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 - \lambda \min_v \left\{ \|v\|_1 + \frac{\gamma}{2}\|(x - v)\|_2^2 \right\} \\ &= \frac{1}{2}x^{\mathrm{T}}(A^{\mathrm{T}}A - \lambda\gamma I)x + \lambda\|x\|_1 + \max_v\{g(x, v)\} \end{aligned} \tag{9}$$

where the function g is affine in x. The convexity of F is maintained if $A^{\mathrm{T}}A - \lambda\gamma I \succeq 0$. However, for TQWT, A is a matrix such that $A^{\mathrm{T}}A$ is singular. Consequently, we must set $\gamma = 0$ to maintain the convexity of the function F. Yet, when $\gamma = 0$, the penalty in (8) reduces to the L1 norm. In this case, we achieve no benefit in comparison to a conventional sparse regularization using L1-norm regularization.

In order to maintain the convexity of the cost function F in (6) to be minimized, we use the GMC penalty ψ_{GMC}, which is defined as

$$\psi_{\mathrm{GMC}}(x) = \|x\|_1 - \min_v \left\{ \|v\|_1 + \frac{\gamma}{2\lambda}\|A(x - v)\|_2^2 \right\} \tag{10}$$

where the matrix A is introduced to maintain the convexity of the cost function F and the parameter γ controls the nonconvexity of the GMC penalty.

Using the GMC penalty in (10), the cost function F in (6) can be expressed as

$$\begin{aligned} F(x) &= \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 \\ &\quad - \min_v \left\{ \lambda\|v\|_1 + \frac{\gamma}{2}\|A(x - v)\|_2^2 \right\} \\ &= \max_v \left\{ \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 - \lambda\|v\|_1 \right. \\ &\quad \left. - \frac{\gamma}{2}\|A(x - v)\|_2^2 \right\} \\ &= \max_v \left\{ \frac{1}{2}(1 - \gamma)\|Ax\|_2 + \lambda\|x\|_1 + g(x, v) \right\} \\ &= \frac{1}{2}(1 - \gamma)\|Ax\|_2 + \lambda\|x\|_1 + \max_v\{g(x, v)\} \end{aligned} \tag{11}$$

where the function g is affine in x. The last term is convex as it is the pointwise maximum of a set of convex functions (see [43, Prop. 8.14]). Thus, the cost function F is convex if $0 < \gamma \leqslant 1$.

B. Solving the GMC Regularization Problem

Because the GMC penalty does not have a closed-form formula and so cannot be directly evaluated, it may appear necessary to use an iterative algorithm comprising a double nested loop: an inner loop to solve for the optimal solution v^{opt} and an outer loop to iteratively solve for the optimal solution x^{opt}. Fortunately, because F can be expressed as a saddle function convex in x and concave in v, a global minimizer can be readily calculated using proximal algorithms. It is not necessary to explicitly evaluate the GMC penalty or its gradient.

The minimization of the GMC-regularized least squares problem can be written as a saddle-point problem:

$$(x^{\text{opt}}, v^{\text{opt}}) = \arg\min_{x} \max_{v} F(x, v) \tag{12}$$

where

$$F(x,v) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 - \lambda\|v\|_1 - \frac{\gamma}{2}\|A(x-v)\|_2^2$$

with $0 < \gamma < 1$. This saddle-point problem can be solved by the forward–backward splitting algorithm, as listed in Algorithm 2, where the constant $\mu > 0$ should be smaller than the constant $2/\max\{1, \gamma/(1-\gamma)\}\|A\|_2^{-2}$, which ensures that the iterative algorithm converges.

It can be observed from the w-update in Algorithms 1 and 2 that the main difference between the GMC method and the L1-norm method is that the term $\mu\gamma A^{\mathrm{T}}A(x^{(i)} - v^{(i)})$ is added in the w-update in the GMC algorithm. This is the reason why the GMC regularization avoids underestimating high-amplitudes values, in contrast to the L1-norm method.

C. K-Sparsity Strategy for Adaptive Threshold

In this section, we introduce a k-sparsity strategy for the adaptive selection of the regularization parameter λ.

It can be observed in Algorithm 2 that the parameter λ affects only the threshold value in the soft thresholding operator. Therefore, the parameter λ determines how many coefficients at each iteration will be set to zero and how many coefficients will be preserved in the soft thresholding step. Moreover, the value of λ does not directly affect the convergence of the algorithm or the convergence speed of the algorithm. Thus, according to the k-sparsity prior, we can set λ adaptively such that, in each iteration, an adaptive threshold $T^{(i)}$ is set as the kth largest coefficient in the vector $w^{(i)}$, i.e., $T^{(i)} = w^{(i)}_{[k]}$, where the notation $w^{(i)}_{[k]}$ represents the kth largest coefficient (in absolute value) of the vector $w^{(i)}$. This adaptive threshold $T^{(i)}$ will preserve only the k largest coefficients via soft thresholding and set the other coefficients to zero. The iterative algorithm with an adaptive parameter for solving the saddle point problem in (12) is summarized as Algorithm 3.

Algorithms 2 and 3 involve only simple computational steps, i.e., soft thresholding and the operator A and A^{T}. Moreover, the operation of TQWT is matrix-free and efficiently implemented using radix-2 FFTs; hence, Algorithms 2 and 3 are also matrix-free. Thus, the algorithm is efficient for minimizing the cost function F in (6). The computational cost of radix-2 TQWT is $O(rM\log_2 M)$, where M is the length of the input signal, and r is the redundancy factor [41]. Thus, the total computational cost of Algorithms 2 and 3 is $O(rLM\log_2 M)$, where L is the number of iterations.

Algorithm 2: Iterative algorithm for GMC problem.

Initialize $x^{(0)}$, $v^{(0)}$,
$0 < \mu < 2/\max\{1, \gamma/(1-\gamma)\}\|A\|_2^{-2}$
for $i = 0, 1, 2, \ldots$ **do**
  $w^{(i)} = x^{(i)} - \mu A^{\mathrm{T}}(Ax^{(i)} - y) + \mu\gamma A^{\mathrm{T}}A(x^{(i)} - v^{(i)})$,
  $u^{(i)} = v^{(i)} + \mu\gamma A^{\mathrm{T}}A(x^{(i)} - v^{(i)})$,
  $x^{(i+1)} = \text{soft}\,(w^{(i)}; \lambda\mu)$,
  $v^{(i+1)} = \text{soft}\,(u^{(i)}; \lambda\mu)$,
end for
return $x^{(i+1)}$

Algorithm 3: K-sparsity iterative algorithm for GMC problem.

Initialize $x^{(0)}$, $v^{(0)}$,
$0 < \mu < 2/\max\{1, \gamma/(1-\gamma)\}\|A\|_2^{-2}$
for $i = 0, 1, 2, \ldots$ **do**
  $w^{(i)} = x^{(i)} - \mu A^{\mathrm{T}}(Ax^{(i)} - y) + \mu\gamma A^{\mathrm{T}}A(x^{(i)} - v^{(i)})$,
  $u^{(i)} = v^{(i)} + \mu\gamma A^{\mathrm{T}}A(x^{(i)} - v^{(i)})$,
  $T^{(i)} = w^{(i)}_{[k]}$,
  $x^{(i+1)} = \text{soft}\,(w^{(i)}; T^{(i)})$,
  $v^{(i+1)} = \text{soft}\,(u^{(i)}; T^{(i)})$,
end for
return $x^{(i+1)}$

It is noted that, in Algorithm 3, an assumption is that the matrix consists of uniform column vectors, i.e., $\|A_i\|_2 = \|A_j\|_2$ for any $1 \leqslant i \neq j \leqslant N$, where A_j represents column j of matrix A. However, TQWT used in this paper is not a "uniform" tight frame. Thus, TQWT should be normalized beforehand such that $\|A_i\|_2 = \|A_j\|_2$ for any $1 \leqslant i \neq j \leqslant N$; or, alternatively, the wavelet coefficients should be normalized before setting the kth largest coefficient as the threshold for soft thresholding and using a nonuniform threshold for soft thresholding, i.e.,

$$\left[\tilde{w}^{(i)}\right]_j = \left[w^{(i)}\right]_j / \|A_j\|_2, \quad j = 1, 2, \ldots, N, \tag{13}$$

$$\left[T^{(i)}\right]_j = \tilde{w}^{(i)}_{[k]} \cdot \|A_j\|_2, \quad j = 1, 2, \ldots, N. \tag{14}$$

IV. Simulation Study

In this section, we use a simulation study to verify the effectiveness of the GMC method and the k-sparsity strategy in bearing fault diagnosis. First, we analyze the simulation signal using Algorithm 2 with a fixed regularization parameter λ to verify the performance of the GMC method for the estimation

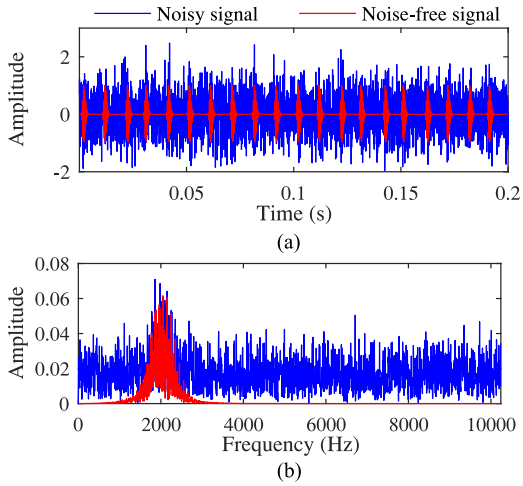

Fig. 1. Simulation signal with and without noise: (a) the waveform in the time domain, and (b) its spectrum.

of a bearing fault signal. Then, we analyze the same signal using Algorithm 3 to verify the effectiveness of the k-sparsity strategy for adaptive parameter selection.

We consider the following noisy signal to simulate the bearing fault signal:

$$y[i] = s[i] + n[i] = \sum_k h(i\Delta_t - kT - \Delta T_k - \tau_0) + n[i]$$

where $i = 1, 2, \ldots, M$, s is quasi-periodic, i.e., the period $T = 0.01$ with random fluctuation ΔT_k drawn from the uniform distribution on the open interval $(-0.001, 0.001)$, and consists of a series of double-side asymmetric transients $h(t)$

$$h(t) = \begin{cases} e^{\frac{-\zeta_L}{\sqrt{1-\zeta_L^2}}(2\pi f_1 t)^2} \cos(2\pi f_1 t), & t < 0 \\ e^{\frac{-\zeta_R}{\sqrt{1-\zeta_R^2}}(2\pi f_1 t)^2} \cos(2\pi f_1 t), & t \geqslant 0 \end{cases} \quad (15)$$

with $f_1 = 2000$, $\zeta_L = 0.02$, $\zeta_R = 0.005$, and $\tau_0 = 0.002$. The symbol Δ_t is the sampling interval, the sampling frequency is 20.48 kHz, and the signal length is $M = 4096$. The noise n is white Gaussian noise $\mathcal{N}(0, \sigma^2)$ with standard deviation $\sigma = 0.6$. In this simulation, the random fluctuation of the time interval between adjacent transients is considered to simulate speed fluctuations or random slipping of rolling elements [15], [16]. The simulation signal and its spectrum are shown in Fig. 1. It can be seen that the true signal is heavily corrupted by noise. The SNR of the noisy signal is -8.43 dB, and the kurtosis value of the noisy signal is 3.046, which is very near to the kurtosis value of white Gaussian noise.

A. Fixed Regularization Parameter λ

We first use Algorithm 2 with a fixed regularization parameter λ to analyze the simulation signal. In order to explore the effect of λ on the performance of the sparse regularized method, we vary λ from 0.5 to 3.5 (with increment 0.1) and use 100 realizations of the random parameter of the simulation signal, including the random fluctuation ΔT_k and random noise n_0. The

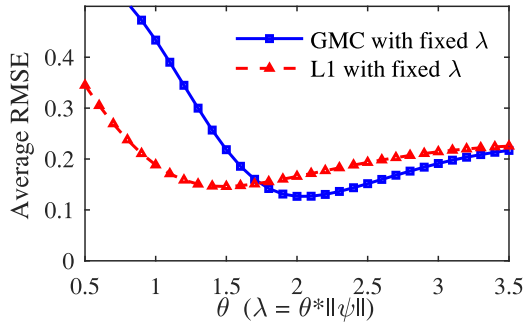

Fig. 2. Average RMSE values of the denoised signal for 100 random realizations obtained by GMC regularization and L1-norm regularization with fixed regularization parameter.

performance is evaluated by the average root-mean-squared error (RMSE), as shown in Fig. 2, where the RMSE of the L1-norm method is also included for comparison. In this case, the GMC nonconvexity parameter is set to $\gamma = 0.8$, the TQWT parameters are $Q = 2$, $r = 5$, and $J = 10$ (leading to $N = 24\,576$ wavelet coefficients). We find that the performance of both methods depends on the parameter λ [i.e., θ according to relation (5)]. The minimal average RMSE of the GMC and L1 methods is 0.1267 and 0.1460, respectively, and the "best" parameter is $\theta = 2.0$ for the GMC method and $\theta = 1.5$ for the L1 method, respectively. The improvement of GMC regularization over the L1 regularization is 13.2%.

The tradeoff between data fidelity and sparsity is realized by adjusting the parameter λ. Generally speaking, a larger value makes the optimal solution sparser but reduces the coefficient amplitudes; a smaller value increases the coefficient amplitudes of the optimal solution but makes the solution less sparse and noisier. How to set the regularization parameter λ effectively in practical applications is a key problem for sparsity regularization methods.

B. Adaptive Thresholding Via a k-Sparsity Strategy

We use the GMC method with the k-sparsity strategy (i.e., Algorithm 3) to analyze the simulation signal. In this case, except the k-sparsity parameter, other parameters are the same as the parameters used in the previous section for the method with fixed regularization parameter. That is, the GMC nonconvexity parameter is set to $\gamma = 0.8$, the TQWT parameters are $Q = 2$, $r = 5$, and $J = 10$. We also use 100 realizations of the random parameter (i.e., the random fluctuation ΔT_k and random noise n) of the simulation signal to evaluate the performance of the GMC method and vary the k-sparsity parameter k from 40 to 600 (with increment 10). The performance is evaluated by the average RMSE. For comparison purposes, we also use the L1-norm method with the k-sparsity strategy to analyze the simulation signal. Fig. 3 compares the average RMSE of GMC and L1-norm regularization methods. The minimal average RMSE of the GMC and L1-norm methods is 0.1277 and 0.1459, and the "best" k-sparsity parameter is $k = 100$ for GMC and $k = 240$ for L1 norm, respectively. The improvement of GMC over L1 norm is 12.5%. The comparison results show that the GMC

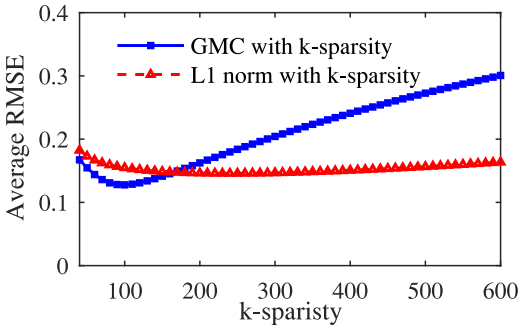

Fig. 3. Average RMSE values of the denoised signal obtained by GMC regularization and L1-norm regularization for 100 random realizations.

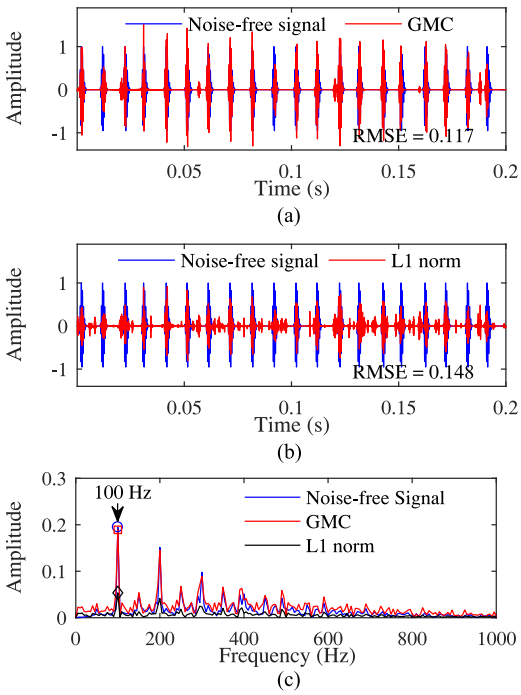

Fig. 4. Results by GMC regularization and L1-norm regularization with k-sparsity strategy: (a) the GMC denoised signal, (b) the L1-norm denoised signal, and (c) their envelope spectra.

is more accurate and uses fewer coefficients and yields more accurate denoising result. Moreover, it can be observed that the accuracy of the GMC and L1-norm methods is basically the same as Algorithm 2 with a fixed regularization parameter when each method has the "best" parameter to minimize the average RMSE. However, we find in Fig. 3 that the k-sparsity strategy is more intuitive to illustrate the performance of the GMC method that uses fewer coefficients and yields more accurate denoising result compared to the L1-norm method.

In order to further illustrate the effectiveness of the method with the k-sparsity strategy, the analysis results are shown in Fig. 4, including the denoised signal using GMC and L1 norm, and their squared envelope spectra (SES). In this case, the parameter k is set to minimize the average RMSE, i.e., $k = 100$ for the GMC method and $k = 240$ for the L1-norm method. It can be observed that, compared to L1 norm, GMC preserves the signal's amplitude while reducing noise, i.e., the amplitude-preserving performance of GMC is much better than the L1-norm method.

It should be noted that, even though the GMC method can remove the most of noise from the noisy signal, a low level of noise cannot be removed, which causes the phenomenon that the amplitude of some denoised signals using GMC is larger than that of the noise-free signal, and some smaller, which may cause a misunderstanding that the GMC may amplify the signal of interest. But the fact is that the noise is distributed throughout, even though the bearing fault signals are concentrated in a few large-amplitude wavelet coefficients, these coefficients are also polluted by noise: some coefficients are larger than the true value and some coefficients are smaller than the true value. When the signal of interest is reconstructed from these coefficients, the amplitude of some transients may be larger than that of the noise-free signal.

For comparison, SK is also used to analyze the signal. As is well-known, if the resonance frequency band can be determined, the SES is useful for bearing fault diagnosis. For example, in this simulation case, if we manually set the bandpass filter with passband (1000, 3000) Hz, which covers the central frequency 2000 Hz of the simulation signal, the SES can identify the characteristic frequency (100 Hz) of the simulated bearing fault signal, as shown in Fig. 5. In practical applications, SK is useful to determine the filter passband. The SK results are also shown in Fig. 5, which can also indicate the characteristic frequency, where the filtering band is selected adaptively by the maximum kurtosis. However, the amplitude of the signal is significantly reduced and the denoised signal is also noisy. The GMC result shown in Fig. 4 indicates that the proposed method can preserve the amplitude of the signal and provide more accurate estimation result.

The favorable behavior of GMC with the k-sparsity strategy is maintained over a range of noise levels, as shown in Fig. 6, which shows the average RMSE a function of k-sparsity for values σ in the interval $0.3 \leqslant \sigma \leqslant 0.6$. We calculate the average RMSE using 20 noise realizations for each noise level. The behavior of the k-sparsity strategy is robust to noise, and GMC compares favorably to L1 norm. The improvements of GMC over L1 norm in different noise levels are 24.1% ($\sigma = 0.3$), 24.0% ($\sigma = 0.4$), 20.6% ($\sigma = 0.5$), 12.5% ($\sigma = 0.6$). Moreover, the noise level has a negligible effect on the selection of the "best" parameter k for GMC and L1-norm regularization. Thus, considering that the signal length is different in the practical applications, we can prescribe a straightforward strategy by setting the k-sparsity parameter such that only 0.5–2% of the coefficients in the wavelet domain are preserved (the other coefficients are set to zero by thresholding).

V. Applications in Bearing Fault Diagnosis

We present two case studies to verify the effectiveness of the GMC method in bearing fault diagnosis.

A. Case Study I

In the first case study, we analyze vibration signals collected during a bearing run-to-failure experiment. In this experiment,

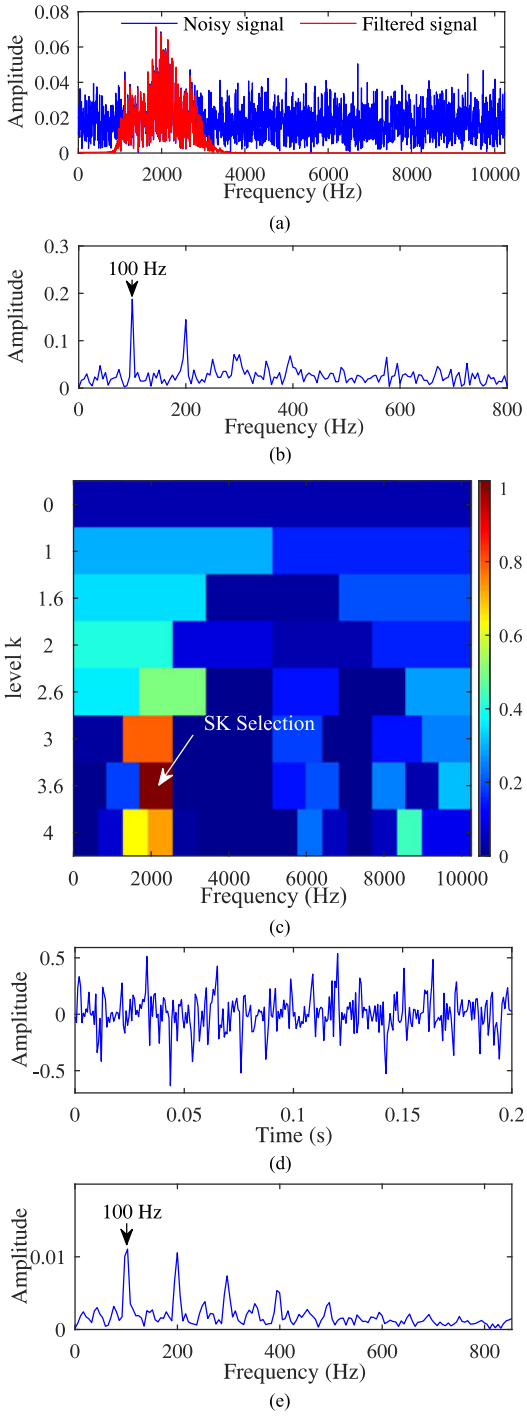

Fig. 5. SES and SK result. (a) The spectrum of the noisy signal and the filtered signal (the specified frequency band [1000, 3000] Hz), (b) the SES of the filtered signal, (c) the kurtogram (the filter band indicated by the arrow is determined by the maximum kurtosis), (d) the filtered signal via SK, and (e) its SES.

vibration data were collected every 10 min at the sampling frequency of 20.48 kHz; the data length in each sampling is 20 480 samples. At the end of the run-to-failure experiment, an outer race failure occurred on one bearing (the fault characteristic frequency being 236.4 Hz), and 984 files were acquired during the experiment with a total duration of 163.8 h. For more detailed information about this experiment, refer to [44].

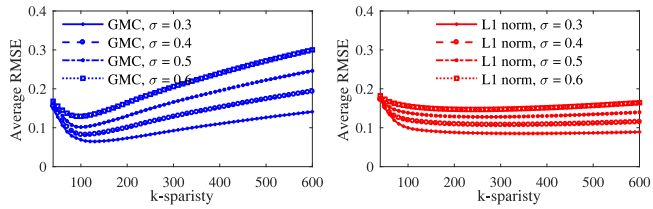

Fig. 6. Average RMSE values of 20 realizations of GMC and L1 norm for different noise levels.

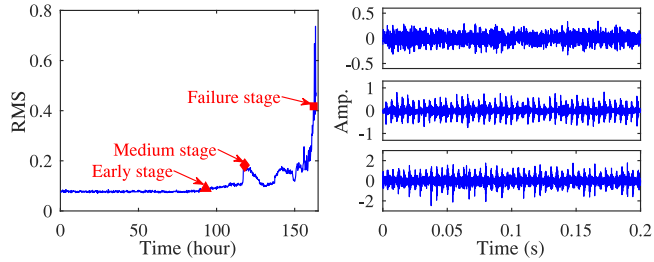

Fig. 7. Root-mean-square (RMS) of the vibration in the run-to-failure experiment (left) and the vibration signal in different stages: early-stage defect (right top), medium-stage defect (right middle), and last failure (right bottom).

Fig. 7 depicts the root-mean-square feature of vibration signals for the entire life cycle, and three segments of vibration signals during different stages: the early-stage defect (file no. 558, about 70 h before the bearing failure), the medium-stage defect (file no. 703, about 45 h before the bearing failure), and the last failure stage (file no. 973). It can be observed that the vibration signals collected at the medium stage and last failure stage exhibit strong periodic impulses, while the amplitude of impulses at the medium stage is smaller than the amplitude at the last failure stage. However, during the early stage, obvious periodic impulses in the vibration signal do not exist, and the transient feature generated by the early defect is masked by noise.

First, we apply the GMC method with the k-sparsity strategy to analyze the vibration signal with an early-stage defect. We set the GMC nonconvexity parameter to $\gamma = 0.8$, the TQWT parameters to $Q = 2$, $r = 5$, and $J = 10$ (there are $M = 4096$ signal samples in the time domain and thus $N = 24\,576$ coefficients in the wavelet domain). We set the k-sparsity strategy to preserve only 2% of the total coefficients in the wavelet domain, i.e., $k = 492$. The GMC denoising results are shown in Fig. 8. It can be observed that the bearing fault signal is extracted from the vibration signal. The zoomed-in comparison shows that the amplitude of the bearing fault signal is almost preserved, and the period is consistent with the outer race fault, as marked with the red arrow. The SES also shows that the characteristic frequency is consistent with the outer race fault.

For comparison purposes, we also apply the L1-norm method (see Fig. 8) and SK (see Fig. 9) to analyze the same signal. It can be observed that the L1-norm result is under-estimated; it is not as accurate as the GMC result. The SES of SK-filtered signal can also indicate the outer race fault. However, it is affected by the heavy noise, and the amplitude is much smaller than the GMC result.

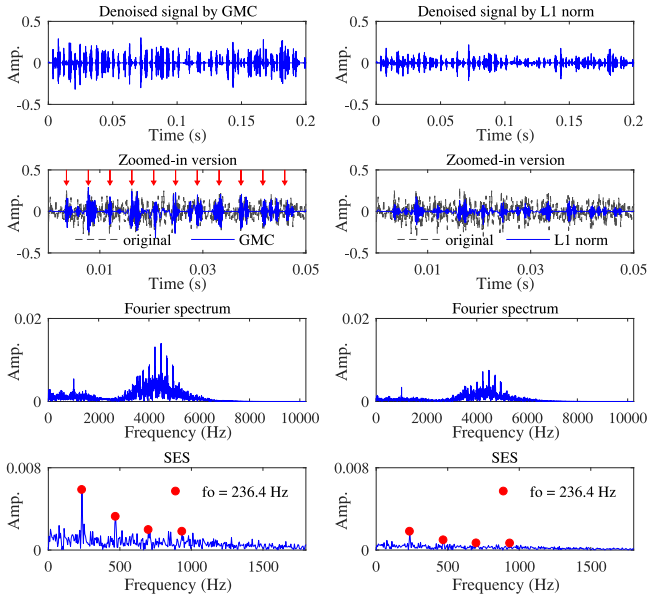

Fig. 8. Denoised result for vibration with the early-stage defect obtained by the GMC (left) and L1 norm (right), including the denoising signal and its zoomed-in version, the spectrum and SES.

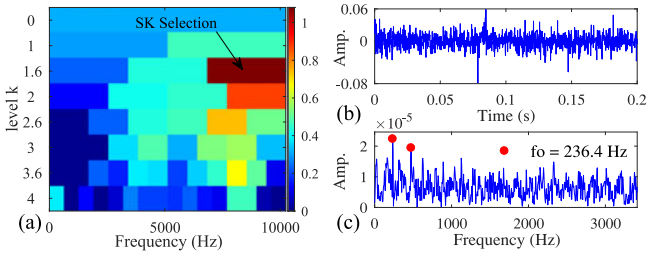

Fig. 9. SK result for vibration with the early-stage defect: (a) the kurtogram (the filter band indicated by the black arrow is determined by the maximum kurtosis), (b) the filtered signal via SK, and (c) its SES.

Then, the proposed method is used to further analyze the vibration signals collected during the later stages in this experiment. If we use the GMC method with fixed parameter λ (i.e., Algorithm 2), the regularization parameter λ should be adjusted for different stages to rebalance the signal of interest and noise. Since the simulation study shows that the noise level has a negligible effect on the selection of the best k-sparsity parameter, when we use the GMC method with the k-sparsity strategy, the same parameters used for the early-stage defect are used here to analyze the vibration signals collected during the later stages. The GMC results of the vibration signals with medium-stage defect and with failure bearing (shown in Fig. 7) are shown in Fig. 10. Even though the k-sparsity parameter is not adjusted, the GMC method with the k-sparsity strategy can also preserve the bearing fault signal while reducing noise. Hence, the adaptive k-sparsity strategy is more easily used in practical applications. For comparison purposes, SK is also used to analyze the same signals, and the results are shown in Fig. 11. It can be observed that the SK results significantly reduce the amplitude of the vibration signal while reducing the interference. Even the amplitude of the SK result for failure bearing is much less than that of the GMC result for the early-stage defect. The

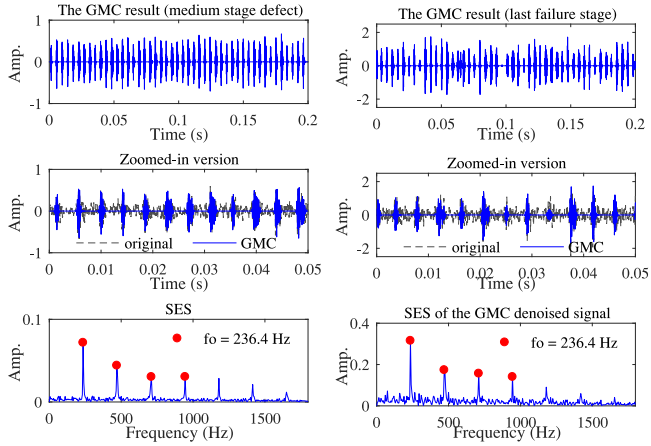

Fig. 10. GMC results for vibration at different stages, i.e., medium-stage defect (left) and last failure (right), including the denoising signal, its zoomed-in plot, and SES.

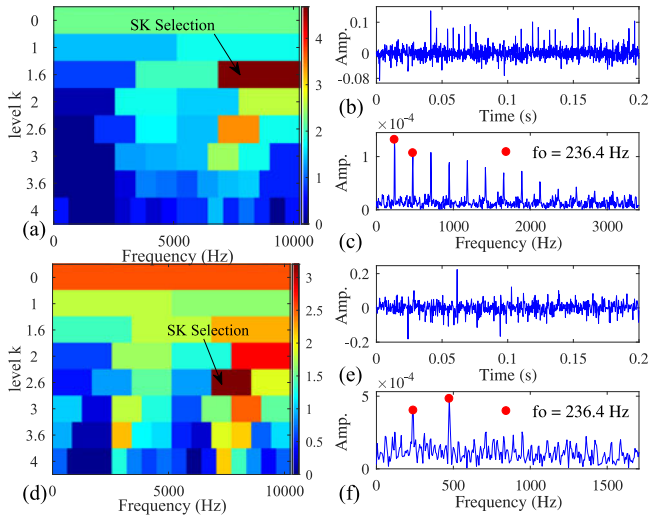

Fig. 11. SK result for vibration at different stages. (a) Kurtogram of the vibration signal with the medium-stage defect, (b) the filtered signal, and (c) its SES; (d) kurtogram of the vibration signal with failure bearing, (e) the filtered signal, and (f) its SES.

application for different stages of defects and the comparisons further demonstrates that the proposed method preserves the amplitude of the signal and improves the estimation accuracy, which is a promising tool for providing effective features to assess the severity of bearing faults.

B. Case Study II

In this section, we validate the GMC method using another run-to-failure experiment. The test rig consists of a mechanical system, a loading system, a lubrication system, an electrical control system, and an industrial PC, as shown in Fig. 12. A radial load of 11 kN and an axial load of 2 kN were applied to the test bearing by the load system. Two accelerometers were mounted on the sleeve that was connected to the outer ring of the tested bearing. Vibration signals were sampled at 20.48 kHz by a data acquisition system.

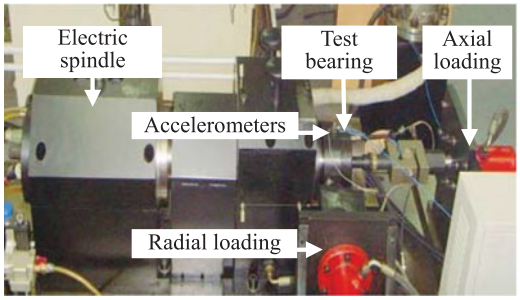

Fig. 12. Bearing test rig.

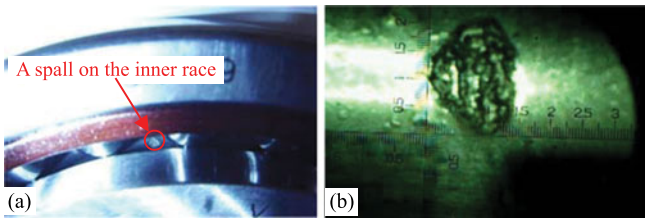

Fig. 13. (a) A spall on inner race of the failure bearing and (b) the area of the spall.

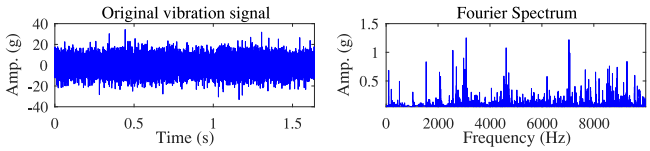

Fig. 14. Original vibration signal and its spectrum.

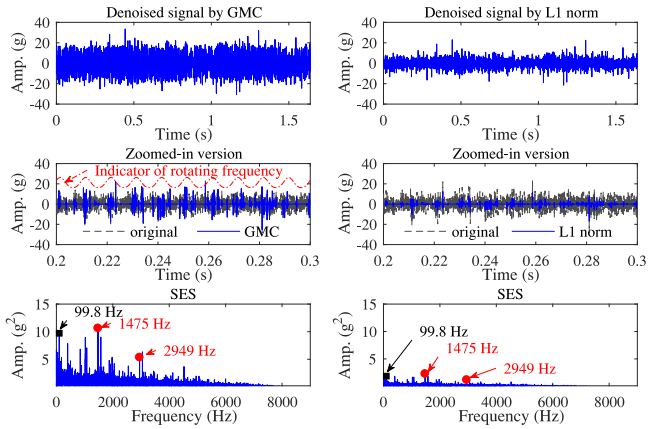

Fig. 15. Denoised result obtained by the GMC and L1 norm, including the denoising signal and its zoomed-in version, the spectrum, and SES.

After about 146 working hours, a spall formed in the inner race (see Fig. 13), and the area of the spall was measured to be about 3 mm^2. The tested bearing was an H7018C angular contact ball bearing. The bearing has 27 balls, a pitch diameter of 117 mm, ball diameter of 11.12 mm, and a contact angle of 15°. The shaft rotation speed was about 6000 r/min. The ball-passing frequency on the inner raceway was about 1475 Hz.

We use the GMC method to analyze the vibration signal (see Fig. 14) collected during the experiment. The signal is of length $M = 32\,768$. The GMC results are shown in Fig. 15. In this case, we set the GMC nonconvexity parameter to $\gamma = 0.8$, the

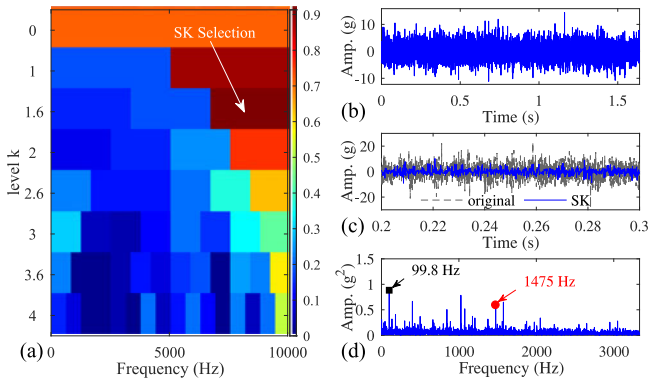

Fig. 16. SK result: (a) the kurtogram (the filter band indicated by the white arrow is determined by the maximum kurtosis), (b) the filtered signal via SK, (c) zoomed-in version for comparison, and (d) the SES of the filtered signal.

TQWT parameters to $Q = 1$, $r = 5$, and $J = 10$ (then there are $N = 208\,896$ coefficients in the wavelet domain). We set the k-sparsity parameter such that each iteration preserves 2% of the total coefficients in the wavelet domain, i.e., $k = 4178$. The result shows that the GMC method preserves the bearing fault signal while reducing noise. Moreover, because the fault occurs on the inner race, the periodic change of the load leads to a clustering/grouping phenomenon of the transients, as illustrated in the zoomed-in comparison in Fig. 15. The fault characteristic frequency (1475 Hz) and the rotating frequency (99.8 Hz) are dominant in the SES, which verifies that the GMC method can effectively extract the bearing fault feature.

For comparison, the L1 norm (see Fig. 15) and SK (see Fig. 16) are also applied to analyze the same vibration signal. The SES of the L1 norm and SK also indicate the inner race fault. However, the amplitude of results of L1 norm is much less than that of the original vibration signal. It can be observed that, besides the amplitude difference of the filtered signal between GMC and SK, the fault characteristic frequency is more dominant in the GMC result than the SK result. If we further observe the waveform of the filtered signal via the GMC and the SK method, it can be observed that the main part of the bearing fault signal is preserved and the noise is better suppressed in the GMC filtered signal. However, in the SK-filtered signal, it can be observed that we cannot observe the clustering/grouping phenomenon of the transients caused by the periodic change of the load, and the bearing fault signal is heavily reduced and more of the noise remains. The proposed method behaves better than these methods.

VI. Conclusion

In this paper, we presented the NSR method for bearing fault diagnosis based on the GMC penalty [defined in (10)]. The main advantage of the GMC regularization over conventional filtering methods is that GMC can better preserve the bearing fault signal while reducing the interference of noise and other components; thus, it can significantly improve the estimation accuracy of the bearing fault signal. The GMC penalty itself is a nonconvex function; thus, the GMC regularization can enhance

sparsity compared to convex penalty. Moreover, the GMC regularization can maintain the convexity of the cost function to be minimized, thus allowing convex optimization to be used to find the global optimal solution (e.g., using the forward–backward splitting algorithm). The k-sparsity strategy was introduced for the adaptive selection of the regularization parameter, which is useful for practical applications, where the k-sparsity parameter can be prescribed simply, e.g., 0.5–2% of the total number of coefficients in the sparsity domain. The simulation example and two application case studies verified the effectiveness of the GMC method in the diagnosis of localized faults in rolling bearings. The comparisons to L1 norm and SK demonstrated the advantage of GMC regularization in preserving the signal and improving the estimation accuracy.

The GMC method is useful for the feature extraction of repetitive transients, and the paper focused on the application of single-point bearing fault diagnosis. However, the application of the method is not limited to this topic. It could also be useful for other applications whose vibration signals have a similar feature, for example, gearbox fault diagnosis, compound fault diagnosis. Moreover, because of the performance in preserving the signal of interest while reducing the interference, it is a promising preprocessing method to assess fault condition of bearings and gearboxes.

Because the GMC method provides a unified framework of NSR and convex optimization, several extensions of the GMC method are of interest. For example, the idea may admit extension to more signal models and more general convex regularizers for machinery fault diagnosis or other purposes, such as morphological component analysis [28] or union of multidictionary [29], weighted sparse model [40], group sparsity [27], and cosparse regularization [45].

REFERENCES

[1] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis—Part II: Signals and signal processing methods," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6546–6557, Oct. 2015.

[2] A. J. Volponi, "Gas turbine engine health management: Past, present, and future trends," *J. Eng. Gas Turbines Power*, vol. 136, no. 5, 2014, Art. no. 051201.

[3] S. Wang, X. Chen, C. Tong, and Z. Zhao, "Matching synchrosqueezing wavelet transform and application to aeroengine vibration monitoring," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 2, pp. 360–372, Feb. 2017.

[4] S. Wang, X. Chen, I. W. Selesnick, Y. Guo, C. Tong, and 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.*, vol. 100, pp. 242–288, 2018.

[5] J. Guzinski, M. Diguet, Z. Krzeminski, A. Lewicki, and H. Abu-Rub, "Application of speed and load torque observers in high-speed train drive for diagnostic purposes," *IEEE Trans. Ind. Electron.*, vol. 56, no. 1, pp. 248–256, Jan. 2009.

[6] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.

[7] L. Frosini, C. Harlişca, and L. Szabó, "Induction machine bearing fault detection by means of statistical processing of the stray flux measurement," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1846–1854, Mar. 2015.

[8] M. Hamadache, D. Lee, and K. C. Veluvolu, "Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6486–6495, Oct. 2015.

[9] M. Amar, I. Gondal, and C. Wilson, "Vibration spectrum imaging: A novel bearing fault classification approach," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 494–502, Jan. 2015.

[10] H.-T. Yau, S.-Y. Wu, C.-L. Chen, and Y.-C. Li, "Fractional-order chaotic self-synchronization-based tracking faults diagnosis of ball bearing systems," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3824–3833, Jun. 2016.

[11] Y. Li, M. Xu, X. Liang, and W. Huang, "Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6506–6517, Aug. 2017.

[12] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[13] W. Ahmad, S. A. Khan, and J.-M. Kim, "A hybrid prognostics technique for rolling element bearings using adaptive predictive models," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1577–1584, Feb. 2018.

[14] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. New York, NY, USA: Wiley, 2011.

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

[16] J. Antoni and R. Randall, "Differential diagnosis of gear and bearing faults," *J. Vib. Acoust.*, vol. 124, no. 2, pp. 165–171, 2002.

[17] V. C. Leite *et al.*, "Detection of localized bearing faults in induction machines by spectral kurtosis and envelope analysis of stator current," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1855–1865, Mar. 2015.

[18] M. Kang, J. Kim, L. M. Wills, and J.-M. Kim, "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7749–7761, Dec. 2015.

[19] J. Antoni, "The spectral kurtosis: A useful tool for characterising non-stationary signals," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 282–307, 2006.

[20] E. Fournier, A. Picot, J. Régnier, M. T. Yamdeu, J.-M. Andréjak, and P. Maussion, "Current-based detection of mechanical unbalance in an induction machine using spectral kurtosis with reference," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1879–1887, Mar. 2015.

[21] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.

[22] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, pp. 1–15, 2014.

[23] H. Hong and M. Liang, "Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform," *J. Sound Vib.*, vol. 320, no. 1, pp. 452–468, 2009.

[24] X. Chen, G. Cai, H. Cao, and W. Xin, "Condition assessment for automatic tool changer based on sparsity-enabled signal decomposition method," *Mechatronics*, vol. 31, pp. 50–59, 2015.

[25] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 558–574, 2011.

[26] L. Cui, J. Wang, and S. Lee, "Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis," *J. Sound Vib.*, vol. 333, no. 10, pp. 2840–2862, 2014.

[27] W. He, Y. Ding, Y. Zi, and I. W. Selesnick, "Sparsity-based algorithm for detecting faults in rotating machines," *Mech. Syst. Signal Process.*, vol. 72, pp. 46–64, 2016.

[28] G. Cai, X. Chen, and Z. He, "Sparsity-enabled signal decomposition using tunable q-factor wavelet transform for fault feature extraction of gearbox," *Mech. Syst. Signal Process.*, vol. 41, no. 1, pp. 34–53, 2013.

[29] Z. Du, X. Chen, H. Zhang, and R. Yan, "Sparse feature identification based on union of redundant dictionary for wind turbine gearbox fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6594–6605, Oct. 2015.

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

[31] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$l_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.

[32] Y. Wang, J. Zeng, Z. Peng, X. Chang, and Z. Xu, "Linear convergence of adaptively iterative thresholding algorithms for compressed sensing," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2957–2971, Jun. 2015.
[33] R. Chartrand and W. Yin, "Nonconvex sparse regularization and splitting algorithms," in *Splitting Methods in Communication, Imaging, Science, and Engineering*. New York, NY, USA: Springer, 2016, pp. 237–249.
[34] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Probl.*, vol. 32, no. 7, pp. 75004–75028, 2016.
[35] I. Selesnick and M. Farshchian, "Sparse signal approximation via nonseparable regularization," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2561–2575, May 2017.
[36] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.
[37] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
[38] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B (Methodol.)*, vol. 58, pp. 267–288, 1996.
[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[40] H. Zhang, X. Chen, Z. Du, and R. Yan, "Kurtosis based weighted sparse model with convex optimization technique for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 80, pp. 349–376, 2016.
[41] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3560–3575, Aug. 2011.
[42] C.-H. Zhang *et al.*, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
[43] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408. New York, NY, USA: Springer, 2011.
[44] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vib.*, vol. 289, no. 4, pp. 1066–1090, 2006.
[45] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal.*, vol. 34, no. 1, pp. 30–56, 2013.

Shibin Wang (M'15) received the B.S. and M.S. degrees in electrical engineering from Soochow University, Suzhou, China, in 2008 and 2011, respectively, and the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2015.

He then joined the School of Mechanical Engineering, Xi'an Jiaotong University, where he is a Lecturer. In 2017, he was a Visiting Scholar at the Tandon School of Engineering, New York University, Brooklyn, NY, USA. His research interests include time-frequency analysis and sparsity-assisted signal processing for machine condition monitoring and fault diagnosis.

Ivan Selesnick (S'91–M'98–SM'08–F'16) received the B.S., M.E.E., and Ph.D. degrees in electrical engineering from Rice University, Houston, TX, USA, in 1990, 1991, and 1996, respectively.

In 1997, he was a Visiting Professor with the University of Erlangen-Nurnberg, Erlangen, Germany. He then joined the Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn, NY, USA (now the Tandon School of Engineering, New York University), where he is currently a Professor. His research interests include signal and image processing, wavelet-based signal processing, sparsity techniques, and biomedical signal processing.

Dr. Selesnick has been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Gaigai Cai (M'17) received the B.S. degree from Xidian University, Xi'an, China, in 2008, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2013, both in mechanical engineering.

She then joined the School of Rail Transportation, Soochow University, Suzhou, China, where she is currently an Associate Professor. In 2017, he was a Visiting Scholar at the Tandon School of Engineering, New York University, Brooklyn, NY, USA. Her current research interests include machinery fault diagnosis, sparse techniques for machinery signal processing, and machinery failure evolution mechanisms.

Yining Feng received the B.S. degree from Binghamton University, State University of New York, New York, NY, USA, in 2013, and the M.S. degree from New York University, Brooklyn, NY, USA, in 2016, both in electrical engineering.

Currently, he is a Research Assistant with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University.

Xin Sui received the B.S. degree in electrical engineering from Peking University, Beijing, China, in 2012, and the M.S. degree in electrical engineering from New York University, Brooklyn, NY, USA, in 2014, where he is working toward the Ph.D. degree in electrical engineering.

His research interests include optimization algorithms and convex analysis with applications in signal processing, biomedical imaging, and digital filter design.

Xuefeng Chen (M'12) received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2004.

He is a Full Professor and the Dean of the School of Mechanical Engineering, Xi'an Jiaotong University. He has authored more than 100 SCI publications in areas such as composite structure, aero-engines, wind power equipment, etc.

Dr. Chen works as the Executive Director of the Fault Diagnosis Branch in China Mechanical Engineering Society. He is also a Member of ASME, and the Chair of the IEEE Xian and Chengdu Joint Section Instrumentation and Measurement Society Chapter. He was the recipient of the National Excellent Doctoral Thesis Award in 2007, the First Technological Invention Award of Ministry of Education in 2008, the Second National Technological Invention Award in 2009, the First Provincial Teaching Achievement Award in 2013, the First Technological Invention Award of Ministry of Education in 2015, and the Science and Technology Award for Chinese Youth in 2013. Additionally, he hosted a National Key 973 Research Program of China as a Principal Scientist in 2015.