


## Article

# Weak Fault Detection for Gearboxes Using Majorization–Minimization and Asymmetric Convex Penalty Regularization

Qing Li <sup>1,\*</sup>  and Steven Y. Liang <sup>1,2</sup>

<sup>1</sup> College of Mechanical Engineering, Donghua University, Shanghai 201620, China; steven.liang@me.gatech.edu

<sup>2</sup> George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405, USA

\* Correspondence: suesliqing@163.com

Received: 5 June 2018; Accepted: 16 June 2018; Published: 26 June 2018



**Abstract:** It is a primary challenge in the fault diagnosis community of the gearbox to extract the weak fault features under heavy background noise and nonstationary conditions. For this purpose, a novel weak fault detection approach based on majorization–minimization (MM) and asymmetric convex penalty regularization (ACPR) is proposed in this paper. The proposed objective cost function (OCF) consisting of a signal-fidelity term, and two parameterized penalty terms (i.e., one is an asymmetric nonconvex penalty regularization term, and another is a symmetric nonconvex penalty regularization term). To begin with, the asymmetric and symmetric penalty functions are established on the basis of an L1-norm model, then, according to the splitting idea, the majorizer of the symmetric function and the majorizer of the asymmetric function are respectively calculated via the MM algorithm. Finally, the MM is re-introduced to solve the proposed OCF. As examples, the effectiveness and reliability of the proposed method is verified through simulated data and gearbox experimental real data. Meanwhile, a comparison with the state-of-the-art methods is illustrated, including nonconvex penalty regularization (NCPR) and L1-norm fused lasso optimization (LFLO) techniques, the results indicate that the gear chipping characteristic frequency 13.22 Hz and its harmonic (2f, 3f, 4f and 5f) can be identified clearly, which highlights the superiority of the proposed approach.

**Keywords:** sparse regularization; majorization minimization (MM); asymmetric convex penalty regularization (ACPR); gearbox weak; gearbox

## 1. Introduction

Gearboxes are widely used in transmission systems with rotating machinery, due to its advantages, such as a high transmission ratio and large load capacity, which plays an important role in modern industrial applications. Unexpected bearing failure caused by a harsh working environment and intricate running conditions may result in catastrophic incidents, significant economic losses, and even human casualties. Thus, timely and precise condition-based monitoring (CBM) and health assessments of rotating machines are of great significance in practice, performed online or offline, to avoid a mechanical breakdown, especially in the early stages of failure [1–3].

The vibration signal generated by localized faults is a typical non-linear and non-stationary signal; thus, fault feature extraction based on vibration signal has become the mainstream technique in the fault diagnosis field. However, the observed vibration signals are usually corrupted by severe background noise and interference components. For the purpose of detecting fault from rotating machines, over the past decades, numerous diagnostic techniques have been developed. For example, some methods

are used via signal transforms, such as wavelet/wavelet packet transform [4–6], short-time Fourier transform [7,8], or tunable Q-factor wavelet transform (TQWT) [9,10]. Some methods are used via signal adaptive decomposition, such as empirical mode decomposition (EMD) [11,12], local mean decomposition (LMD) [13,14], and variational mode decomposition (VMD) [15,16]; in addition, some methods are used by intelligent supervised learning techniques, such as artificial neural network (ANN) [17,18], deep learning (DL) [19,20], just to mention a few. These methods have achieved successful applications in fault diagnosis of rotating machinery to some extent.

Compared to the above conventional fault feature extraction approaches, most recently, sparse coding techniques (SCT) or sparse representation (SR) as a new fault detection method was introduced for fault feature extraction and condition monitoring [21,22]. By applying the SR approach, the observed vibration signals can be represented as a linear combination of a few sparse bases (e.g., wavelet base) or sparse atoms generated by an over-complete dictionary (OCD), so the instantaneous fault impulses can be substituted by a few sparse coefficients. Zhou et al. [23] proposed shift-invariant dictionary learning (SIDL) and a hidden Markov model (HMM) to address bearing fault diagnosis and identify fault types. Further, Yang et al. [24] and Feng et al. [25] proposed a shift-invariant K-singular value decomposition (SI-K-SVD) method combined with a sparse representation algorithm to extract the fault impulse of wind turbine generator bearing and planetary gearbox fault, respectively. Qin [26] developed the model-based impulsive wavelet basis and Fourier basis for the weak repetitive transient feature extraction of a rolling bearing. The adaptive impulse dictionary, double-dictionary matching pursuit (DDMP), and step-impulse dictionary were introduced for detecting the physical defect size of the inner and outer race of a rolling bearing in [27–29]. Ding et al. [30] employed the time-frequency (TF) dictionary and orthogonal matching pursuit (OMP) to extract the fault feature information of single row cylindrical roller bearings based on acoustic signals. Li et al. [31] introduced a novel dictionary learning method called the impulse-step impact dictionary, based on a nonconvex optimization approach; the results demonstrate the method's superiority in weak fault feature extraction via accelerated lifetime testing experiment, compared with OMP, L1-norm convex penalty regularization, and the spectral Kurtogram (SK) method.

Although the above SR and its extension, combined with the redundant dictionary, have achieved satisfactory applications in fault diagnosis, those algorithms have some common drawbacks:

- (1) Unique dictionaries' atoms and optimal wavelet basis cannot simultaneously match the natural structure of every real vibration signal well;
- (2) A large number of observed signals should be collected to form a training dictionary before diagnosis, which is always infeasible in practical applications;
- (3) Computational complexity and time-consuming problems occur simultaneously in dictionary training, such as with the K-SVD training and SI-K-SVD dictionaries training [31].

In an attempt to overcome the above shortcomings of an over-complete dictionary (OCD) and estimate the fault characteristics more accurately, the family of low-rank matrix approximation (LRMA) methods were introduced in recent years. Thus the reconstruction problem of dictionary atoms is translated into an optimization or inverse regularization problem. In [32], Ding et al. propose an algorithm that combines alternating the direction method of multipliers (ADMM) and the majorization–minimization (MM) method to compute a low-rank matrix optimization model. In [33] and [34], He et al. develop nonconvex penalty regularization and a sparsity-based group-sparse signal denoising (OGS) approach to extract the single and compound features of outer and inner race defects. In [35], Zhang et al. proposed a novel convex penalty regularization method called Kurtosis-based weighted sparse decomposition (Kur-WSD) on the basis of the L1-norm method to detect the inner fault of an alternating current motor. In [36], Du et al. propose a rigorously weighted low-rank sparse detection framework to explore the physical/internal mechanisms of bearing faults and implement fault diagnosis of a wind turbine.

Unfortunately, the above low-rank matrix approximation approach still suffers from the following challenges, which have restricted its application development:

- (1) Generally, the penalty functions that are established in a low-rank matrix approximation (LRMA) model are symmetric functions, e.g., the absolute value function (AVF); the common drawback is that this penalty function is non-differentiable at the zero point, which can lead to some numerical issues, such as a local optimum and early termination of algorithm.
- (2) In a conventional low-rank matrix approximation (LRMA) method, the convex regularizer, such as the L1-norm, usually underestimates the sparse signal when the absolute value function (AVF) is used as a sparsity regularizer; the nonconvex regularizer suffers from several issues, such as a strict convexity problem of objective cost function (OCF), a non-convergence problem, etc. Additionally, both the convex and nonconvex regularizers shrink all the coefficients equally and remove too much energy from the useful signal, resulting in the estimation of the fault signal becoming more challenging.
- (3) When the useful fault characteristics signals are very weak but additive noise extremely strong, the conventional LRMA method cannot estimate low-dimensional feature distribution accurately.

Aiming at the issue of estimating the weak fault signal from its noisy observations, and enlightened by the LRMA method, a novel detection approach based on asymmetric convex penalty regularization (ACPR) and majorization–minimization (MM) is proposed in this paper, using the weak chipping fault of a gearbox as a research object. The proposed ACPR method neither relies on the reconstruction of dictionary atoms, nor requires a large number of sample data to form a certain training dictionary; meanwhile, the numerical issues, such as being non-differentiable at zero point, are addressed by introducing the asymmetric convex penalty function. In the modeling solution process, MM is implemented to solve a relatively complicated sparsity-based optimization problem. Finally, the proposed method is validated via the weak fault feature of the simulation and the gearbox dataset collected in the lab experiment; the results of fault frequency and its harmonics indicate that the ACPR method is superior to the other LRMA methods, such as nonconvex penalty regularization (NCPR) and L1-norm fused lasso optimization (LFLO) techniques.

The layout of this work is organized as follows. Section 2 introduces the MM algorithm. Section 3 introduces the algorithm and theoretical derivation of asymmetric convex penalty regularization (ACPR). The simulation evaluation of the proposed method is presented in Section 4. In Section 5, the practical diagnosis results and the discussion of the gearbox using the proposed approach with other LRMA methods are presented. Conclusions are drawn in Section 6.

## 2. Majorization–Minimization Algorithm

Before the derivation of the proposed algorithm, a review to the majorization–minimization (MM) algorithm is given. Usually, the MM algorithm is used to reduce the computational complexity—in other words, for a minimization problem that is too complex to solve directly, the MM method can simplify it into a series of simpler problems by constructing the majorization functions [37–39].

Suppose  $F(u)$  is the minimization problem, which is too complex to solve directly. The MM algorithm constructs a simple function  $G(u, u^k)$ ,  $G(u, u^k)$ , which is called the majorization function of  $F(u)$  at the  $k$ -th iteration point  $u^k$ . More specifically, consider the minimization problem

$$u^{\text{opt}} = \underset{u}{\operatorname{argmin}} F(u) \quad (1)$$

By using the MM method, the above problem can be described by the following iteration:

$$u^{(k+1)} = \underset{u}{\operatorname{argmin}} G(u, u^{(k)}) \quad (2)$$

where  $k$  is the iteration number and  $G(u, u^k)$  is a continuously differentiable function, satisfying

$$\begin{aligned} G(u, v) &\geq F(u), \text{ for all } u, v, \\ G(u, u) &= F(u) \end{aligned} \quad (3)$$

Note that the majorizer  $G(u, v)$  reaches  $F(u)$  for  $u = v$ . According to Equations (2) and (3), we have

$$F(u^{k+1}) \leq G(u^{k+1}, u^k) \leq G(u^k, u^k) = F(u^k) \quad (4)$$

This iteration briefly describes the MM approach for minimizing a minimization problem function. The MM algorithm consists of two processes: the construction of the majorization function and the optimization of the majorization function. The main steps of the MM algorithm can be summarized as follows:

- (1) Initialize  $u^0$  and  $k = 0$ ;
- (2) Construct a majorization function  $G(u, u^k)$ ;
- (3) Operate the iteration  $u^{(k+1)} = \underset{u}{\operatorname{argmin}} G(u, u^{(k)})$ ;
- (4) If the stopping criterion is satisfied, then output  $u^{\text{opt}}$ ; otherwise,  $k = k + 1$ , and go to step (2);
- (5) Output  $u^{\text{opt}}$ .

### 3. Asymmetric Convex Penalty Regularization Algorithm

#### 3.1. Sparse Representation and Filter Banks

Generally, the noisy noisy vibration signal  $y$  of rotating machinery, such as a rolling bearing, contains three parts: the fault transient impulses  $x$ , the systematic natural vibration signal  $f$ , and the additive noise  $w$ , i.e.,

$$y = x + f + w \quad (5)$$

The core work of the fault diagnosis is to extract the fault transient impulses  $x$  from the noisy observation  $y$ . Assuming that fault transient component impulses and the systematic natural vibration signal are estimated, we have

$$\hat{x} \approx x, \hat{f} \approx f \quad (6)$$

Assuming that the estimation of fault transient impulses  $x$  is obtained as  $\hat{x}$ , we can estimate signal  $f$  as follows:

$$\hat{f} = \text{LPF}(y - \hat{x}) \quad (7)$$

where  $\text{LPF}$  is a specified low-pass filter. Substituting Equation (7) into  $\hat{f} \approx f$ , we have

$$\text{LPF}(y - \hat{x}) \approx f \quad (8)$$

Substituting  $y = x + f + w$  into  $\text{LPF}(y - \hat{x}) \approx f$ , we have

$$\begin{aligned} y &\approx x + \text{LPF}(y - \hat{x}) + w \\ \Rightarrow \text{LPF}(y - \hat{x}) &\approx y - x - w \end{aligned} \quad (9)$$

Substituting  $\hat{x} \approx x$  into  $\text{LPF}(y - \hat{x}) \approx y - x - w$ , we have

$$\text{LPF}(y - \hat{x}) \approx y - \hat{x} - w \Leftrightarrow (I - \text{LPF})(y - \hat{x}) \approx w \quad (10)$$

Defining  $HPF = I - LPF = H$ , thus, we get

$$HPF(y - \hat{x}) \approx w \Rightarrow H(y - \hat{x}) \approx w \quad (11)$$

On the other hand, it should be noted that Equation (5) belongs to a highly underdetermined equation, i.e., an ill-posed or  $N$ - $P$  hard problem [40–42], for which there are infinite sets of solutions because the number of unknowns is greater than the number of equations. Usually, convex optimization approaches are commonly used to estimate a transient component from the noisy signal; based on the aforementioned work, the estimation of  $x$  can be formulated as the constrained optimization problem, i.e.,

$$\hat{x} = \operatorname{argmin}_x \left\{ \frac{1}{2} \|H(y - x)\|_2^2 + \lambda \|Dx\|_1 \right\} \quad (12)$$

where  $H$  is a specified high-pass filter (i.e., the  $HPF$  in Equation (11)),  $\lambda$  is a regularization parameter,

and  $D$  is a matrix defined as  $D = \begin{bmatrix} -1, 1 & & & \\ & \cdots & & \\ & & \cdots & \\ & & & -1, 1 \end{bmatrix}$ , which controls the sparsity of the approximating value of  $x$ . If  $x$  is a sparse signal, i.e., most of the amplitude values in  $x$  tend to zero, then the problem in Equation (12) can also be solved by the L1-norm fused lasso optimization (LFLO) algorithm, i.e.,

$$\hat{x} = \operatorname{argmin}_x \left\{ \frac{1}{2} \|H(y - x)\|_2^2 + \lambda_0 \|x\|_1 + \lambda_1 \|Dx\|_1 \right\} \quad (13)$$

where  $\lambda_0$  and  $\lambda_1$  are regularization parameters. The solution of the LFLO algorithm can be given by a soft-threshold function [43]. In that case, we have

$$x = \operatorname{soft}(\operatorname{tvd}(y, \lambda_2), \lambda_1) \quad (14)$$

where function  $\operatorname{tvd}(\cdot, \cdot)$  is the total variation de-noising (TVD) algorithm [44–46], and the soft-threshold function is defined as follows

$$\operatorname{soft}(x, \lambda) = \begin{cases} x - \lambda \frac{x}{|x|}, & |x| > \lambda \\ 0, & |x| \leq \lambda \end{cases} \quad (15)$$

In addition, the high-pass filter  $H$  described above could be formulated as follows [45–47]:

$$H = A^{-1}B \quad (16)$$

where  $A$  and  $B$  are Toeplitz matrices.

### 3.2. Asymmetric Convex Penalty Regularization Model

Compared with optimization algorithms Equations (12) and (13), in order to estimate the fault transient impulses  $x$  precisely, this work introduces a novel penalty regularization method, i.e., an asymmetric and symmetric nonconvex penalty regularization model:

$$\begin{aligned} \hat{x} = \operatorname{argmin}_x \{ & F(x) = \frac{1}{2} \|H(y - x)\|_2^2 + \lambda_0 \sum_{n=0}^{N-1} \theta_\varepsilon(x_n; r) \\ & + \sum_{i=1}^M \lambda_i \sum_{n=0}^{N-1} \phi([D_i x]_n) > \} \end{aligned} \quad (17)$$

where  $F(x)$  is the proposed objective cost function (OCF), the penalty function  $\theta_\varepsilon(x_n; r)$  is a asymmetric and differentiable function, and  $\phi([D_i x]_n)$  is a symmetric and differentiable function, if the term

$M = 2$ , the matrix  $D_1$  is defined as  $D_1 = \begin{bmatrix} -1, 1 & & & \\ & \dots & & \\ & & \dots & \\ & & & -1, 1 \end{bmatrix}$ , and the matrix  $D_2$  is defined as  $D_2 = \begin{bmatrix} -1, & 2, & -1 \\ & -1, & 2, & -1 \\ & & \dots & \\ & & -1, & 2, & -1 \end{bmatrix}$ . The innovations of the novel compound regularizer model are as follows:

- (I) The  $M$ -term compound regularizers estimate the fault transient impulses;
- (II) The compound regularizers model consists of symmetric and asymmetric penalty functions, wherein the symmetric penalty function is a differentiable function, compared with the nondifferentiable function  $\|x_i\|$  at  $i = 0$ .
- (III) The MM algorithm is introduced for the solution of the proposed compound regularization method.

Based on this, therefore, the core tasks of the proposed algorithm are (1) how to construct a symmetric and differentiable penalty function; (2) how to construct an asymmetric and differentiable penalty function; and (3) how to solve the proposed method based on the MM algorithm and the estimation  $x$ , and make diagnosis results more accurately than the traditional LFLO and nonconvex penalty regularization approaches.

For the first task, traditional LFLO regularization approach uses the absolute value function (AVF)  $\phi_A(x) = \|x\|$  as the penalty function; however, the common drawback of  $\phi_A(x) = \|x\|$  is that this function is non-differentiable at zero point, which can cause some numerical problems. To address this issue, some non-linear approximation functions of  $\phi_A(x) = \|x\|$  are proposed, i.e.,

$$\phi_B(x) = \sqrt{|x|^2 + \varepsilon} \quad (18)$$

Or

$$\phi_c(x) = |x| - \varepsilon \log(|x| + \varepsilon) \quad (19)$$

Note that when  $\varepsilon = 0$ , then the  $\phi_B(x)$  and  $\phi_c(x)$  degrade into the AVF  $\phi_A(x)$ , and while the  $\varepsilon > 0$ , the  $\phi_B(x)$  and  $\phi_c(x)$  are differentiable at the zero point. The functions  $\phi_A(x)$ ,  $\phi_B(x)$ ,  $\phi_c(x)$ , and their first-order derivatives are listed in Table 1.

**Table 1.** Symmetric penalty functions and their derivatives.

Functions	$\phi(x)$	$\phi'(x)$
$\phi_A(x)$	$\ x\ $	Signal(x)
$\phi_B(x)$	$\sqrt{ x ^2 + \varepsilon}$	$x / \sqrt{ x ^2 + \varepsilon}$
$\phi_c(x)$	$ x  - \varepsilon \log( x  + \varepsilon)$	$x / ( x  + \varepsilon)$

In order for the non-linear approximation functions to maintain the reliable sparsity-inducing behavior of the original LFLO algorithm, the parameter  $\varepsilon$  should be set to an adequately small positive value. For example, the parameters  $\varepsilon = 10^{-5}$  and  $\varepsilon = 10^{-6}$  are small enough that numerical issues can be avoided, and their impact on the sparsity promoting can also be ignored.

For the second task, inspired by the AVF  $\phi_A(x) = \|x\|$ , and in contrast to the symmetric and differentiable penalty function  $\phi_B(x)$  and  $\phi_c(x)$ , here a segmented function is proposed as follows:

$$\theta_\varepsilon(x) = \begin{cases} x, & x > \varepsilon \\ f(x), & \|x\| \leq \varepsilon \\ -rx, & x < -\varepsilon \end{cases} \quad (20)$$

where  $r > 0$  is a positive constant. Therefore, the main problem of Equation (20) will transform into a task about how to construct the intermediate function  $f(x)$ ,  $-\varepsilon \leq x \leq \varepsilon$ . To address this issue, we seek a majorizer as the approximation function of  $f(x)$ ,  $-\varepsilon \leq x \leq \varepsilon$ ; here, in order to eliminate the issue that the penalty function is non-differentiable at zero point, a simple quadratic equation (QE) is introduced accordingly,

$$g(x, v) = ax^2 + bx + c \quad (21)$$

According to the theory of the majorization–minimization (MM), we have,

$$\begin{aligned} g(v, v) &= \theta(v, r), \quad g'(v, v) = \theta'(v, r) \\ g(s, v) &= \theta(s, r), \quad g'(s, v) = \theta'(s, r) \end{aligned} \quad (22)$$

The parameters  $a, b, c$  and  $s$  are all function of  $v$ , and solving for them gives

$$a = \frac{1+r}{4|v|}, b = \frac{1-r}{2}, c = \frac{(1+r)|v|}{4}, s = -v \quad (23)$$

Substituting Equation (23) into  $g(x, v) = ax^2 + bx + c$ , we have

$$g(x, v) = \frac{1+r}{4|v|}x^2 + \frac{1-r}{2}x + \frac{(1+r)|v|}{4} \quad (24)$$

Similarly, the numerical issue of Equation (24) will appear if the parameter  $v$  approaches zero. To address this issue, the sufficiently small positive value  $\varepsilon$  is used instead of  $|v|$ ; thus, the segmented function Equation (20) can be rewritten as,

$$\theta_\varepsilon(x) = \begin{cases} x, & x > \varepsilon \\ \frac{1+r}{4\varepsilon}x^2 + \frac{1-r}{2}x + \frac{(1+r)\varepsilon}{4}, & \|x\| \leq \varepsilon \\ -rx, & x < -\varepsilon \end{cases} \quad (25)$$

Hence, the new function  $\theta_\varepsilon(x)$  is a continuously differentiable function.

The third task will be solved and derived in Section 3.3 using the MM algorithm.

### 3.3. The Solution of the Proposed Model Based on the Majorization–Minimization Algorithm

In this paper, the majorization–minimization (MM) algorithm is implemented to derive an iterative solution procedure for the proposed approach [47]. The function  $G(x, v)$  is chosen as the majorizer of  $F(x)$ . Specifically, the iterative solution procedure can be divided into three phases:

- The majorizer of the symmetric and differentiable function  $\phi([D_i x]_n)$  based on the MM algorithm.
- The majorizer of the asymmetric and differentiable function  $\theta_\varepsilon(x_n; r)$  based on the MM algorithm.
- The majorizer of the objective cost function  $F(x)$  based on the MM algorithm.

For problem (a), we first seek the majorizer  $g(x, v)$  for  $\phi(x)$ , i.e.,

$$\begin{aligned} g(v, v) &= \phi(v), \text{ for all } x, v \\ g(x, v) &\geq \phi(x) \end{aligned} \quad (26)$$

Since  $\phi(x)$  is symmetric function, we set  $g(x, v)$  to be an even second-order polynomial, i.e.,

$$g(x, v) = mx^2 + b \quad (27)$$

Thus, according to Equation (26), as well as  $g(v, v) = \phi(v)$  and  $g'(v, v) = \phi'(v)$ , we have

$$mx^2 + b = \phi(v) \text{ and } 2mv = \phi'(v) \quad (28)$$

The parameters  $m$  and  $b$  can be computed as

$$m = \frac{\phi'(v)}{2v} \text{ and } b = \phi(v) - \frac{v}{2}\phi'(v) \quad (29)$$

Substituting Equation (29) into Equation (27), we have

$$g(x, v) = \frac{\phi'(v)}{2v}x^2 + \phi(v) - \frac{v}{2}\phi'(v) \quad (30)$$

In summation, we obtain

$$\begin{aligned} \sum_n g(x_n, v_n) &= \sum_n \left[ \frac{\phi'(v_n)}{2v_n} x_n^2 + \phi(v_n) - \frac{v_n}{2}\phi'(v_n) \right] \\ &= \frac{1}{2}x^T \frac{\phi'(v_n)}{v_n} x + \sum_n [\phi(v_n) - \frac{v_n}{2}\phi'(v_n)] \\ &= \frac{1}{2}x^T [\Lambda(v)]_n x + c(v) \\ &\geq \sum_{n=0}^{N-1} \phi(x_n) \end{aligned} \quad (31)$$

where  $[\Lambda(v)]_n = \frac{\phi'(v_n)}{v_n}$  is diagonal matrix and  $c_i(D_i v) = \sum_n [\phi([D_i v]_n) - \frac{[D_i v]_n}{2}\phi'([D_i v]_n)]$ . Therefore, based on Equation (31), we obtain

$$\begin{aligned} &\sum_{i=0}^M \lambda_i \sum_{n=0}^{N-1} g([D_i x]_n, [D_i v]_n) \\ &= \sum_{i=0}^M \left[ \frac{\lambda_i}{2} (D_i x)^T \left[ \frac{\phi'(D_i v)}{D_i v} \right] (D_i x) + \sum_n [\phi([D_i v]_n) - \frac{[D_i v]_n}{2}\phi'([D_i v]_n)] \right] \\ &= \sum_{i=0}^M \left[ \frac{\lambda_i}{2} (D_i x)^T [\Lambda(D_i v)] (D_i x) + c_i(D_i v) \right] \\ &\geq \sum_{i=0}^M \lambda_i \sum_{n=0}^{N-1} \phi([D_i x]_n) \end{aligned} \quad (32)$$

where  $[\Lambda(D_i v)]_n = \frac{\phi'([D_i v]_n)}{[D_i v]_n}$  is a diagonal matrix and  $c(D_i v) = \sum_n [\phi([D_i v]_n) - \frac{[D_i v]_n}{2}\phi'([D_i v]_n)]$ .

For problem (b), we assume that  $g_0(x, v)$  is the majorizer of the asymmetric and differentiable function  $\theta_\varepsilon(x_n; r)$ ; since  $f(x) = \frac{1+r}{4\varepsilon}x^2 + \frac{1-r}{2}x + \frac{(1+r)\varepsilon}{4}$ ,  $\|x\| \leq \varepsilon$ , we have

$$\begin{cases} g_0(x, v) = \frac{1+r}{4v}x^2 + \frac{1-r}{2}x + \frac{(1+r)v}{4} \geq f(x), & v > \varepsilon \\ g_0(x, v) = -\frac{1+r}{4v}x^2 + \frac{1-r}{2}x - \frac{(1+r)v}{4} \geq f(x), & v < -\varepsilon \end{cases} \quad (33)$$

when  $v < -\varepsilon$ , then

$$\begin{aligned} g_0(x, v) - f(x) &= \left( \frac{1+r}{4v} - \frac{1+r}{4\varepsilon} \right) x^2 + (v - \varepsilon) \frac{1+r}{4} \\ &= \frac{(1+r)(v-\varepsilon)(v\varepsilon-x^2)}{4v\varepsilon} > 0 \end{aligned} \quad (34)$$

when  $v < \varepsilon$ , then

$$\begin{aligned} g_0(x, v) - f(x) &= \left( -\frac{1+r}{4v} - \frac{1+r}{4\varepsilon} \right) x^2 - (v + \varepsilon) \frac{1+r}{4} \\ &= -\frac{(1+r)(v+\varepsilon)(v\varepsilon+x^2)}{4v\varepsilon} > 0 \end{aligned} \quad (35)$$

Therefore, the majorizer of the asymmetric and differentiable function  $\theta_\varepsilon(x_n; r)$  is obtained:

$$\begin{cases} g_0(x, v) = \frac{1+r}{4|v|}x^2 + \frac{1+r}{2}x + |v| \frac{1+r}{4}, & |v| > \varepsilon \\ g_0(x, v) = \frac{1+r}{4\varepsilon}x^2 + \frac{1+r}{2}x + \varepsilon \frac{1+r}{4}, & |v| \leq \varepsilon \end{cases} \quad (36)$$



In summation, we obtain

$$\begin{aligned} \sum_{n=0}^{N-1} g_0(x_n, v_n) &= x^T [\Gamma(v)]x + b^T x + c(v) \\ &\geq \sum_{n=0}^{N-1} \theta_\varepsilon(x_n, r) \end{aligned} \quad (37)$$

where  $\Gamma(v)$  is a diagonal matrix, i.e.,  $[\Gamma(v)]_n = (1+r)/4|v_n|, |v_n| \geq \varepsilon$  and  $[\Gamma(v)]_n = (1+r)/4\varepsilon, |v_n| \leq \varepsilon$ , and  $[b]_n = (1-r)/2$ .

For problem (c), based on Equations (32) and (37), the majorizer of  $F(x)$  based on the MM algorithm is given by

$$\begin{aligned} G(x, v) &= \frac{1}{2} \|H(y - x)\|_2^2 + \lambda_0 x^T [\Gamma(v)]x + \lambda_0 b^T x \\ &+ \sum_{i=1}^M \left[ \frac{\lambda_i}{2} (D_i x)^T [\Lambda(D_i v)] (D_i x) \right] + c(v) \end{aligned} \quad (38)$$

Minimizing  $G(x, v)$  with respect to  $x$  yields

$$x = H^T H + 2\lambda_0 \Gamma(v) + \sum_{i=1}^M \lambda_i D_i^T [\Lambda(D_i v)] D_i^{-1} (H^T H y - \lambda_0 b) \quad (39)$$

Here, substituting  $H = BA^{-1}$  into Equation (39), we have

$$\begin{aligned} x &= A \{ B^T B + A^T (2\lambda_0 \Gamma(v) + \sum_{i=1}^M \lambda_i D_i^T [\Lambda(D_i v)] D_i) A \}^{-1} \times (B^T B A^{-1} y - \lambda_0 A^T b) \\ &= A (B^T B + A^T M A)^{-1} (B^T B A^{-1} y - \lambda_0 A^T b) \\ &= A Q^{-1} (B^T B A^{-1} y - \lambda_0 A^T b) \end{aligned} \quad (40)$$

where matrix  $M = 2\lambda_0 \Gamma(v) + \sum_{i=1}^M \lambda_i D_i^T [\Lambda(D_i v)] D_i$  and matrix  $Q = B^T B + A^T M A$ .

Finally, by using the above formulas, the fault transient impulses  $x$  can be obtained by the following iterations:

$$M^{(k)} = 2\lambda_0 \Gamma(x^{(k)}) + \sum_{i=1}^M \lambda_i D_i^T [\Lambda(D_i x^{(k)})] D_i \quad (41)$$

$$Q^{(k)} = B^T B + A^T M^{(k)} A \quad (42)$$

$$x^{(k+1)} = A [Q^{(k)}]^{-1} (B^T B A^{-1} y - \lambda_0 A^T b) \quad (43)$$

In conclusion, the complete steps of the proposed algorithm are summarized as follows:

- (1) **Input:** signal  $y$ ,  $r \geq 1$ , matrix  $A$ , matrix  $B$ ,  $\lambda_i, i = 0, 1, \dots, M, k = 0$ ;
- (2)  $E = B^T B A^{-1} y - \lambda_0 A^T b$
- (3) Initialize  $x = y$ ;
- (4) Repeat the following iterations:

$$[\Gamma(v)]_n = (1+r)/4|v_n|, |v_n| \geq \varepsilon$$

$$[\Gamma(v)]_n = (1+r)/4\varepsilon, |v_n| \leq \varepsilon;$$

$$[\Lambda(D_i v)]_n = \frac{\phi'([D_i v]_n)}{[D_i v]_n}, i = 0, 1, 2, \dots, M;$$

$$M^{(k)} = 2\lambda_0 \Gamma(x^{(k)}) + \sum_{i=1}^M \lambda_i D_i^T [\Lambda(D_i x^{(k)})] D_i;$$

$$\begin{aligned} Q^{(k)} &= B^T B + A^T M^{(k)} A; \\ x^{(k+1)} &= A[Q^{(k)}]^{-1} E; \end{aligned}$$

- (5) If the stopping criterion is satisfied, then output signal  $x$ —otherwise,  $k = k + 1$ , and go to step (4).  
 (6) **Output:** signal  $x$ .

### 3.4. Parameter Selection

In this section, when  $M = 2$ , the regularization parameters  $\lambda_i$  are set as:

$$\begin{aligned} \lambda_0 &= \beta_0 \sigma \\ \lambda_1 &= \beta_1 \sigma = \gamma(1 - \beta_0) \sigma \\ \lambda_2 &= \beta_0 \sigma \end{aligned} \quad (44)$$

where  $\sigma$  is the standard deviation (SD) of the additive noise,  $\beta_0$  and  $\beta_1$  are the constants, so as to maximize the signal-to-noise (SNR); here, parameters  $\beta_0$  and  $\gamma$  are typically set up to be constant value, i.e.,  $\beta_0 = [0.5, 1]$ ,  $\gamma = [7.5, 8]$ . In practice, the SD of the background noise in Equation (44) can be computed using both the fault signal and healthy data under same operating environment. Moreover, when the healthy data is not available or is unknown, the standard deviation of the background noise can still be estimated by the following formula:

$$\hat{\sigma} = MAD(y) / 0.6745 \quad (45)$$

which is a traditional estimator of the noise level that used for wavelet denoising [48], where  $MAD(y)$  represents the median absolute deviation (MAD) of signal  $y$ , i.e.,

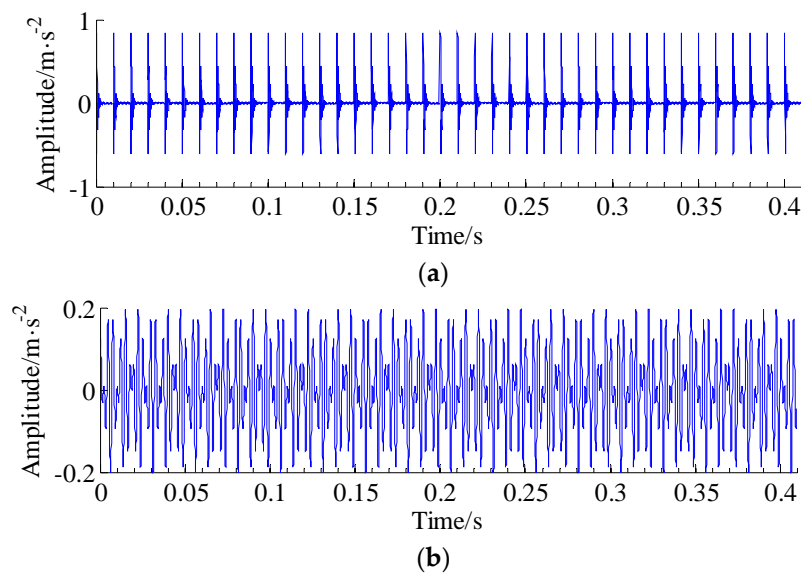
$$MAD(y) = \text{median}(|y_i - \text{median}(y)|), i = 1, 2, \dots, N \quad (46)$$

## 4. Numerical Simulation

A simulation experiment was utilized to investigate the effectiveness of the proposed approach for extracting fault characteristic frequencies. Generally, the localized gear fault signal consists of three typical parts: periodic impulses caused by the localized fault, natural modulated signals due to systematic components, and additive background noise. Take the gearbox structure as an under-damped second-order nonlinear mass-spring-damper (UD-SO-NMSD) system, and its synthetic response function can be described by following formula:

$$\begin{cases} x(t) = x_1(t_0) + x_2(t) + x_3(t) \\ x_1(t_0) = A_0 \exp(-a \times 2\pi f_n t_0) \times \sin(2\pi f_n \times \sqrt{1 - a^2} t_0) \\ x_2(t) = A_1 \sin(2\pi f_1 t) \times \cos(2\pi f_2 t) \\ x_3(t) = \text{sigma} \times \text{randn}(1, N) \end{cases} \quad (47)$$

where  $A_0 = 1 \text{ m/s}^2$  is the intensity of the fault impulse impact,  $A_1 = 0.1 \text{ m/s}^2$  is the intensity of the systematic vibration signal, and the damping ratio is  $a = 0.1$ ; in addition,  $f_n = 2000 \text{ Hz}$  represents the natural frequency of excited structure, the length of vibration signal is  $N = 8192$ , the rotating frequencies are  $f_1 = 280 \text{ Hz}$  and  $f_2 = 400 \text{ Hz}$ , and the sampling frequency  $f_s = 20 \text{ KHz}$ . Additionally, in this simulated case, the additive white noise  $x_3(t)$  with SDs  $\text{sigma} = 0.5$ ,  $\text{sigma} = 0.7$ , and  $\text{sigma} = 0.9$  are respectively added to the simulated signal, in order to test the noise tolerance of the proposed ACPR method. Therefore, it can be calculated that the gear fault frequency is  $100 \text{ Hz}$  (because the repetition interval  $T = 0.01$ , the sampling point of a single period is  $NT = \text{round}(f_s \times T)$ , and the sampling time series is  $t_0 = (0:NT-1)/f_s$ ; therefore, the resonant frequency is  $100 \text{ Hz}$ ). Figure 1a,b depict the obtained periodical impulse of the gear fault and natural modulated signal, respectively.



**Figure 1.** Simulated synthetic signal. (a) Faulty periodical transient impulses of a gearbox; and (b) the systematic natural modulated signal.

As a benchmark approach for gearbox fault detection, the Hilbert envelope spectrum has been effectively applied to fault diagnosis of rotating machinery, such as gears and rolling bearings. In this section, comparisons are made between the raw envelope spectrum method and envelope spectrums based on the proposed ACPR, on nonconvex penalty regularization (NCPR), and on the LFLO method, by analyzing simulated signals. In order to compare the diagnosis effect of the ACPR, NCPR, and LFLO methods, the evaluating parameters below are considered in Table 2. Here, according to Equation (44), we will take the constant parameters to be  $\beta_0 = 0.7$  and  $\gamma = 7.5$ . The parameters of the proposed ACPR can be obtained as follows: the standard deviation  $\sigma = 0.5$ , regularization term parameters  $\lambda_0 = 0.7 \times 0.5 = 0.35$ ,  $\lambda_1 = 7.5 \times (1 - 0.7) \times 0.5 = 1.125$ , and  $\lambda_2 = 0.7 \times 0.5 = 0.35$ .

**Table 2.** Parameter settings of the proposed asymmetric convex penalty regularization (ACPR) method for gear fault detection.

Regularization Parameter $\lambda_0$	Regularization Parameter $\lambda_1$	Regularization Parameter $\lambda_2$	M-Term	Iteration Times
$\lambda_0 = 0.35$	$\lambda_1 = 1.125$	$\lambda_2 = 0.35$	2	50

The simulated synthetic signal under noise a standard deviation sigma of 0.5, obtained from Equation (47), is shown in Figure 2a. From Figure 2a, it can be seen that the periodic impulses are completely buried in background noise. The Hilbert envelope spectrum of the raw synthetic signal is illustrated in Figure 2b, in which the fault frequencies 100 Hz and 200 Hz can be recognized from the envelope spectrum; however, those frequencies are clouded by the interference components. The resulting fault signal extracted through ACPR and its envelope spectrums are depicted in Figure 2c,d, revealing frequency information related to faults. Meanwhile, the resulting impulses components extracted through NCPR and LFLO and their envelope spectrums are exhibited in Figure 2e–h, respectively. As observed, in the case of SD = 0.5, all these results proved that the above three approaches can provide excellent performance in extracting the fault information, in contrast to the results from direct envelope spectrum analysis.

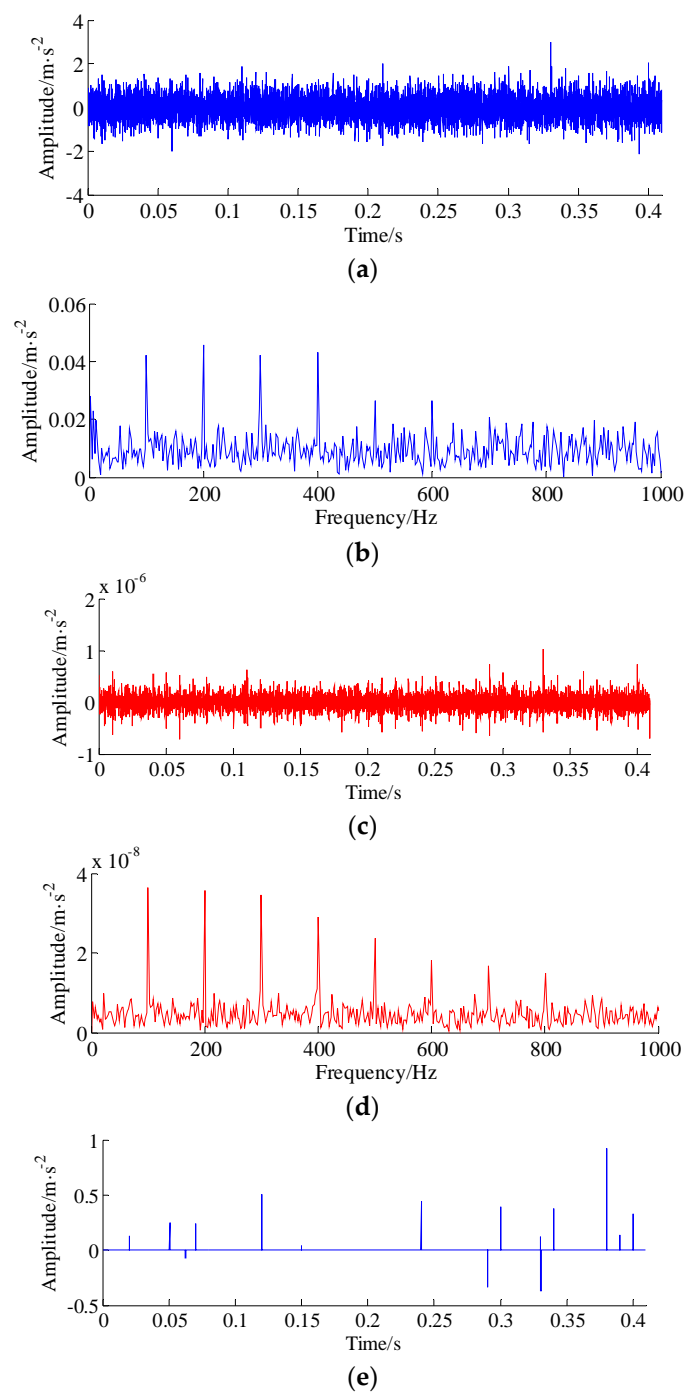
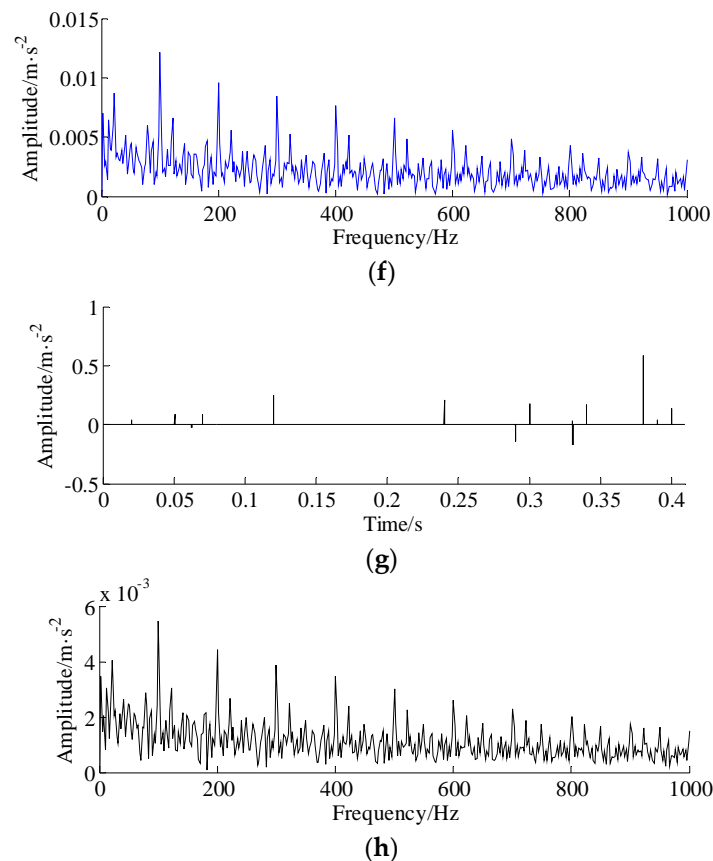


Figure 2. Cont.



**Figure 2.** A simulated synthetic signal, its detected impulses, and its envelope spectrum, using the proposed non-convex penalty regularization and L1-norm fused lasso optimization (LFLO) method, when noise standard deviation is 0.5. (a) Simulated synthetic signal; (b) envelope spectrum of the simulated synthetic signal; (c) detected impulses using the proposed ACPR method; (d) envelope spectrum of detected impulses using proposed ACPR method; (e) detected impulses using the nonconvex penalty regularization (NCPR) method; (f) envelope spectrum of detected impulses using the NCPR method; (g) detected impulses using the LFLO method; (h) envelope spectrum of detected impulses using the LFLO method.

Next, with the purpose of testifying to the noise tolerance performance of the proposed ACPR, the vibration signals with different standard deviations (i.e., 0.7 and 0.9) are simulated and compared. The NCPR and LFLO methods are also used to process the simulated heavy noisy vibration signals, for comparison. The extraction results with different methods are illustrated in Figures 3 and 4, respectively. As shown in Figure 3e,g and Figure 4e,g, most of noise harmonic interferences are removed; meanwhile, the energy of the useful fault periodic impulses is also reduced dramatically. It can be seen that, with the standard deviations (SD) of sigma increasing, the characteristic frequencies and their harmonic components cannot be accurately and clearly extracted by the NCPR and LFLO methods. Hence, the NCPR and LFLO methods are no longer effective for such a signal. Thankfully, the proposed ACPR is an appealing solution for the problem. It can be concluded that for the extraction results shown in Figures 3b–d and 4b–d, the ACPR method is superior to the NCPR and LFLO methods in terms of increasing noise. Table 3 shows the running time of the proposed ACPR algorithm and other benchmarking methods. In this work, the simulation and experimental data were carried out by MATLAB, on a computer with Windows 10, quad-core processors at 2.9 GHz CPU, and 16 GB RAM.

In conclusion, when applied to the noisy signals, the proposed ACPR is more stable with regards to noise perturbation than the other two methods, at least when the noise standard deviation equals 0.7 and 0.9.

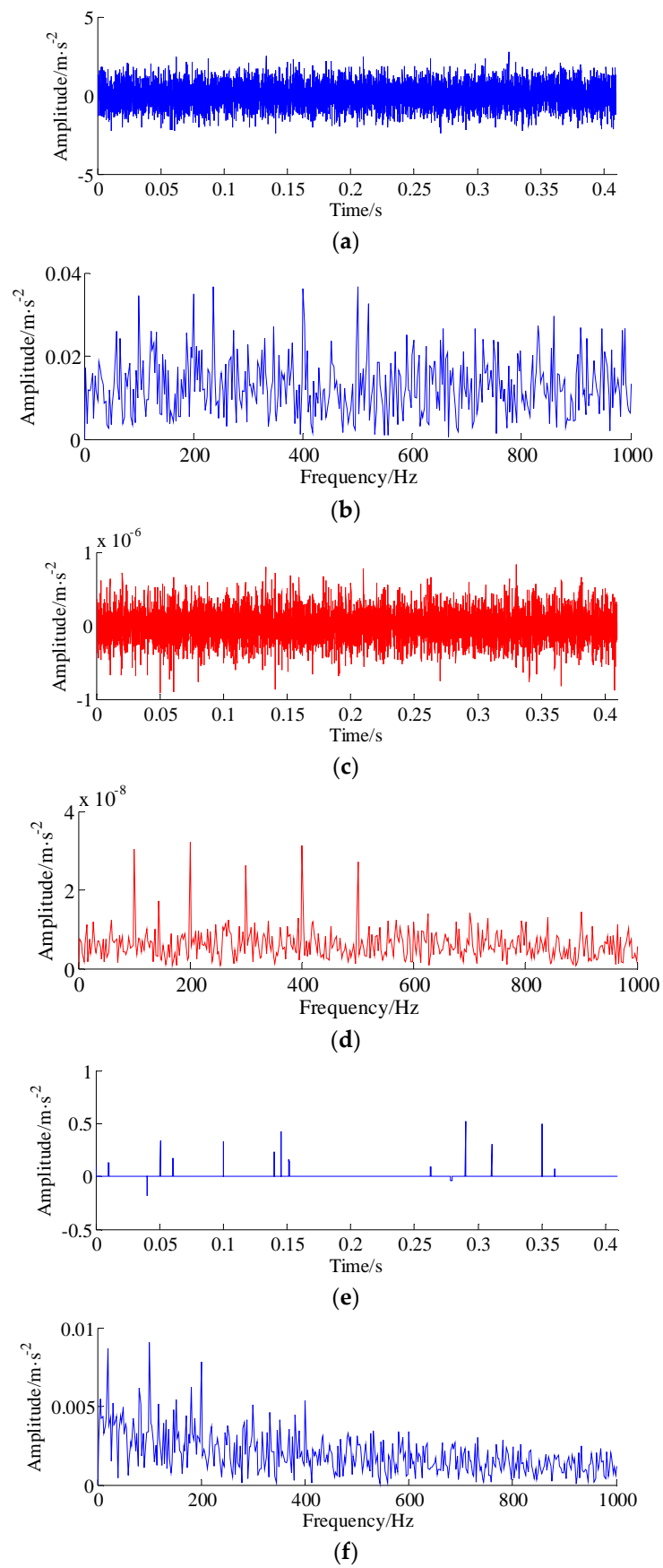
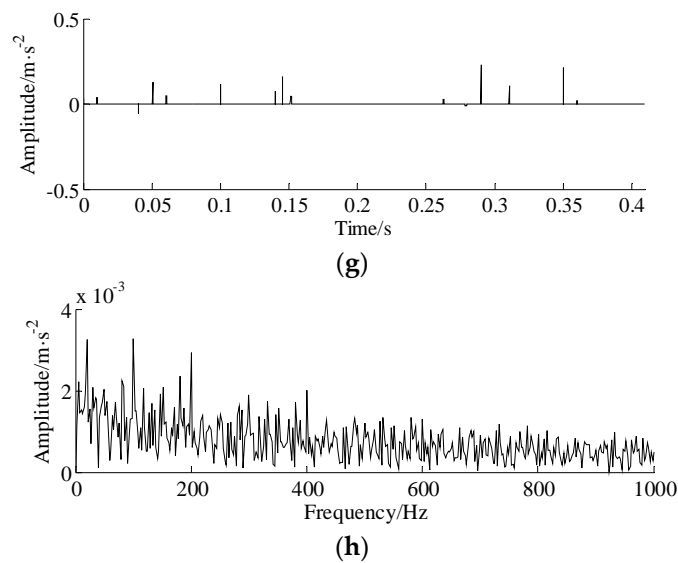
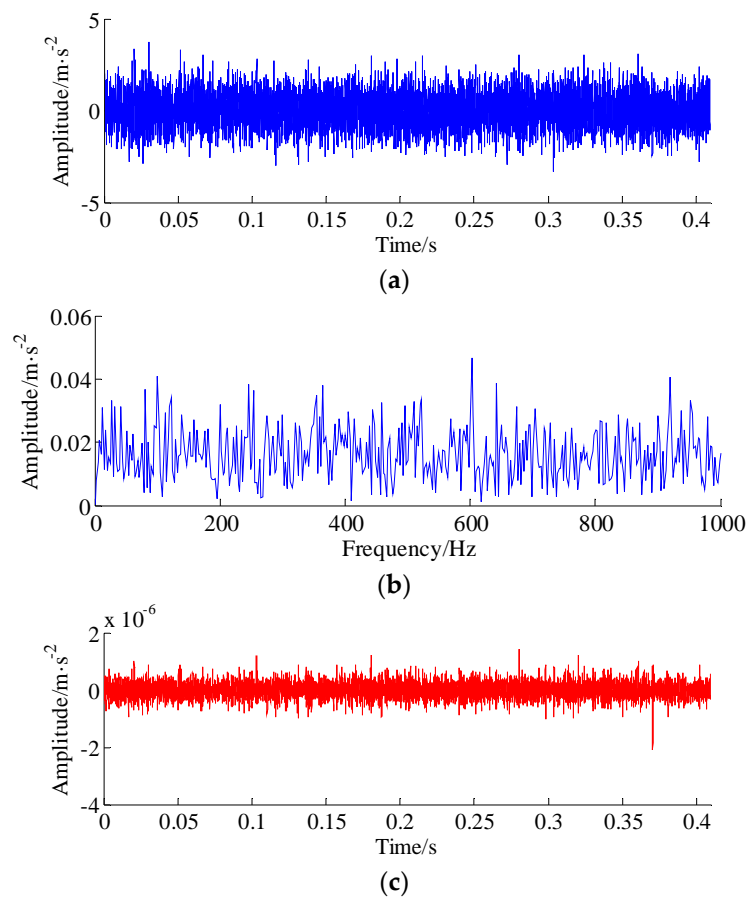


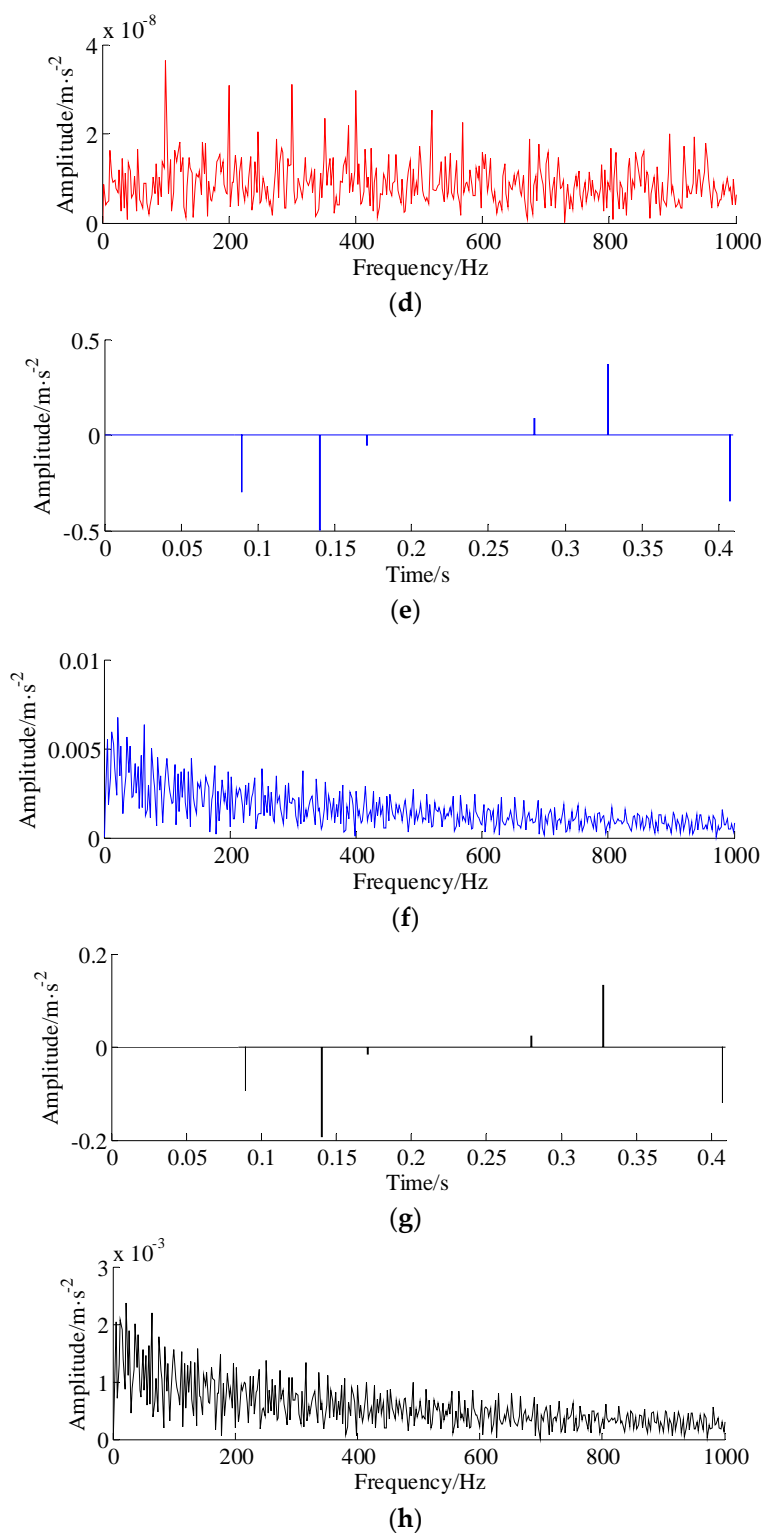
Figure 3. Cont.



**Figure 3.** A simulated synthetic signal, its detected impulses, and its envelope spectrum, using the proposed non-convex penalty regularization and LFLO methods, when noise standard deviation sigma is 0.7. (a) Simulated synthetic signal; (b) envelope spectrum of simulated synthetic signal; (c) detected impulses using the proposed ACPR method; (d) envelope spectrum of detected impulses using the proposed ACPR method; (e) detected impulses using the NCPR method; (f) envelope spectrum of detected impulses using the NCPR method; (g) detected impulses using the LFLO method; and (h) envelope spectrum of detected impulses using the LFLO method.



**Figure 4.** Cont.



**Figure 4.** The simulated synthetic signal, its detected impulses, and its envelope spectrum, using the proposed non-convex penalty regularization and LFLO methods, when the noise standard deviation sigma is 0.9. (a) Simulated synthetic signal; (b) envelope spectrum of the simulated synthetic signal; (c) detected impulses using the proposed ACPR method; (d) envelope spectrum of detected impulses using the proposed ACPR method; (e) detected impulses using the NCPR method; (f) envelope spectrum of detected impulses using the NCPR method; (g) detected impulses using the LFLO method; and (h) envelope spectrum of detected impulses using the LFLO method.



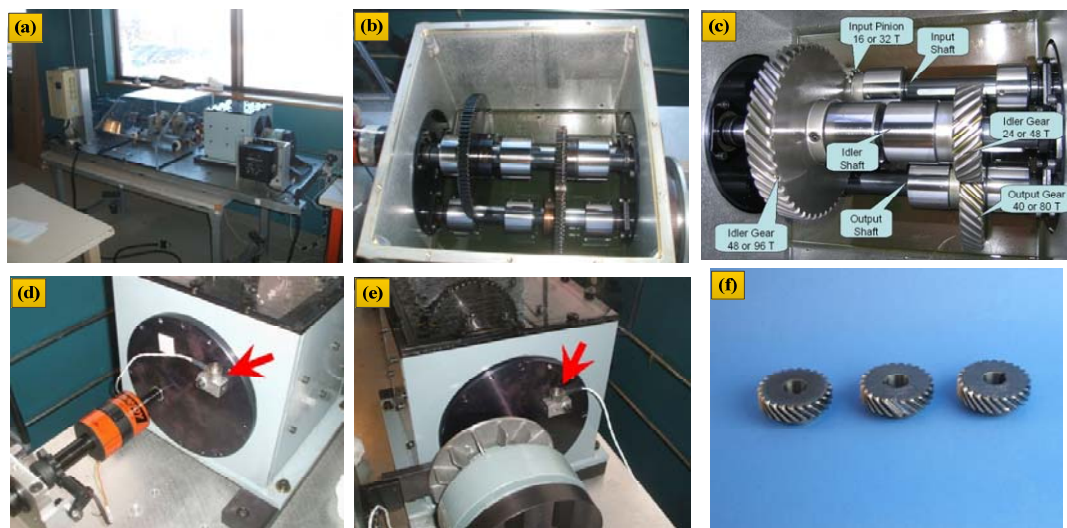
**Table 3.** The running time of the proposed algorithm and benchmarking methods.

Noise Standard Deviation	ACPR Algorithm	NCPR Algorithm	LFLO Algorithm
sigma = 0.5	0.288113 s	0.05391 s	0.000486 s
sigma = 0.7	0.283184 s	0.073986	0.000743 s
sigma = 0.9	0.308288 s	0.052860 s	0.000524 s

## 5. Experimental Validation

The experimental vibration data were collected via a two-class standard cylinder spur gear reducer [49,50]. The gearbox consists of an input shaft, an idler shaft, and an output shaft. The physical experimental rig and schematic diagram of the gearbox are shown in Figures 5 and 6, respectively. There are 16 and 40 teeth of helical gear in the input and output shafts, respectively, and the two gears on the idler shaft have 48 teeth and 24 teeth. The first- and second-stage reduction gear ratios are 1.5 and 1.667, respectively. In this experiment, two triaxial accelerometers with three channels were mounted to the housing with magnetic bases, near the idler shaft (see Figure 5d,e), and the acquisition system was Endevco with 10 mv/g. An artificial fault (chipping failure with a small area) was produced on the tooth of a helical gear (i.e., 24 teeth in the idler shaft), and the vibration signal was picked up from the bearing pedestal with a sample frequency of 66.667 kHz. The shaft rotational speed was 2400 rpm (i.e., 40 Hz) with a low load, the sampling time was set to 4 seconds, and the characteristic defect frequency was 13.33 Hz ( $40 \text{ Hz} \times 16/48 = 13.33 \text{ Hz}$ ). It should be pointed out that the collected signals were not pre-filtered, and hence contains many interference factors and heavy background noise, which render more challenges for the implementation of weak fault detection.

More details about the experimental description and experiment parameters can be found at the website [49].



**Figure 5.** Experimental setup for the gearbox multiple faults. (a) Overview of the experimental apparatus; (b) the internal structure the gearbox; (c) the internal structure of the gear meshing; (d) the installation location of input shaft accelerometer; (e) the installation location of the output shaft accelerometer; and (f) the gears with failure [49,50].

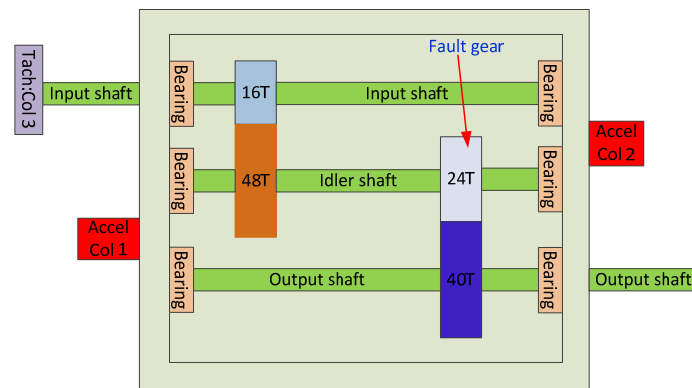


Figure 6. Schematic of the experimental apparatus [49,50].

The raw vibration signal, (here in order to get a more complex vibration signal, a Gaussian white noise with amplitude sigma = 0.04 was added to the original signals) with a length of 4 s, is shown in Figure 7a, in which the periodic impulsive symptoms cannot be unraveled due to high level of noise and harmonic interference. Furthermore, no sign of the gear fault is revealed by its Hilbert envelope demodulation shown in Figure 7b—that is, the weak gear chipping characteristic frequency is completely submerged in the strong noise interference. Thus, an advanced signal processing method for weak fault detection should be developed.

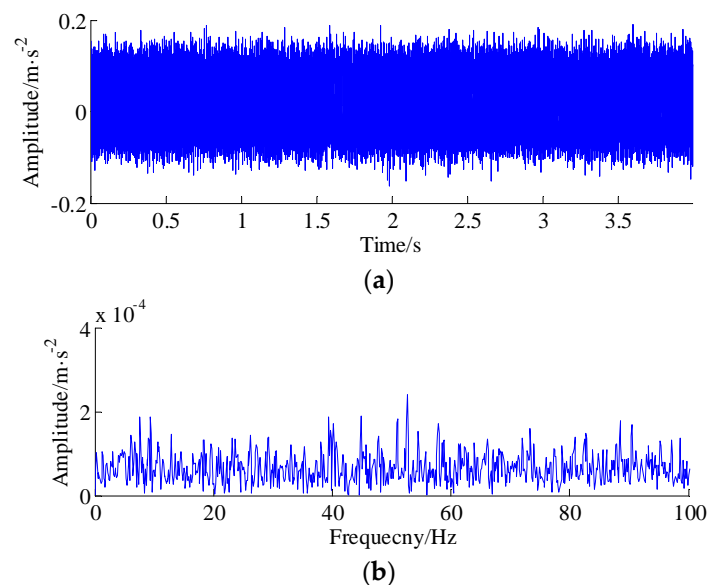
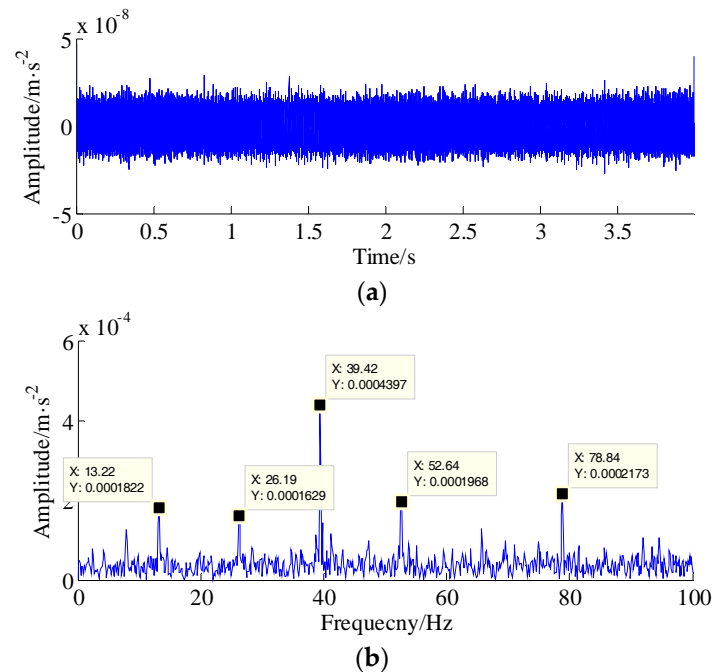


Figure 7. The raw vibration signal and its Hilbert envelope spectrum. (a) The raw vibration signal and (b) the Hilbert envelope spectrum of that raw vibration signal.

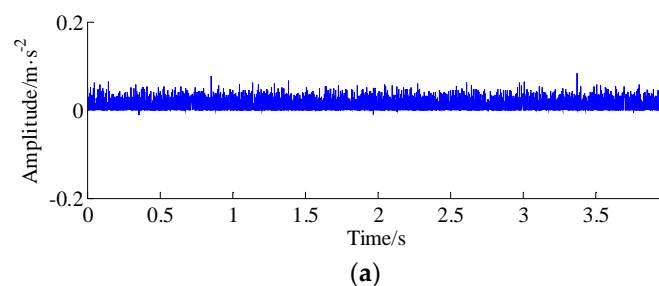
The proposed ACPR method was adopted to process the raw vibration signal for the detected feature information. Since the standard deviation (SD) of the background noise was unknown, it could be calculated by the equation  $\hat{\sigma} = MAD(y)/0.6745 = 0.0409$ . Therefore, the related parameter specification for the ACPR approach is summarized in Table 4. Figure 8 depicts the extracted fault signal and its envelope spectrum. Compared with the original envelope spectrum shown in Figure 7b, it can be clearly seen that the gear chipping characteristic frequency was 13.22 Hz (close to the theoretical fault frequency of gear chipping, which was 13.33 Hz), and its higher orders (2f, 3f, 4f, and 5f) were obvious and had a satisfying distribution pattern. Therefore, it can be concluded that a localized fault exists on the helical gear with 24 teeth in the idler shaft.

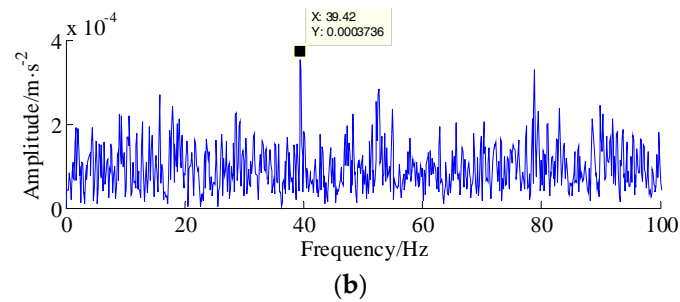
**Table 4.** Parameter settings of the proposed ACPR method for gear fault detection.

Regularization Parameter $\lambda_0$	Regularization Parameter $\lambda_1$	Regularization Parameter $\lambda_2$	M-Term	Iteration Times
$\lambda_0 = 0.02863$	$\lambda_1 = 0.09203$	$\lambda_2 = 0.02863$	2	50

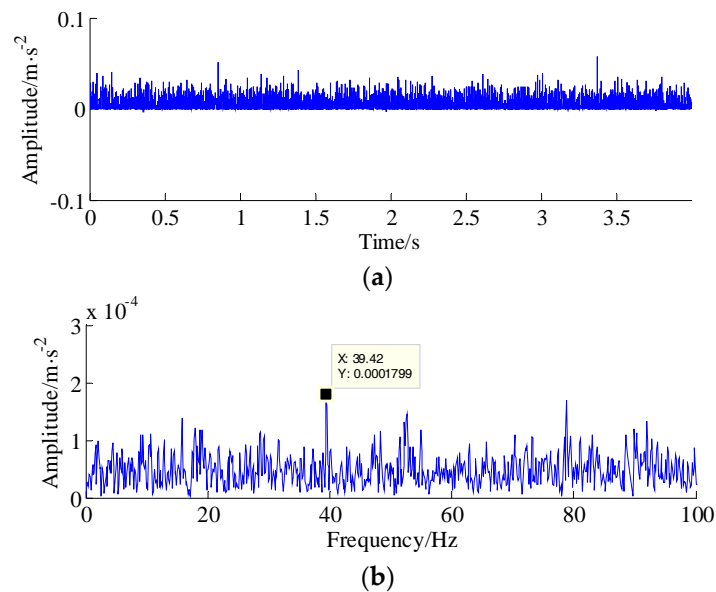
**Figure 8.** The fault information extracted through the proposed approach. (a) The time-domain waveform of the extracted fault signal and (b) the envelope spectrum of the extracted fault signal.

For the sake of comparing the diagnosis results obtained by the NCPR and LFLO methods, the parameters like regularization coefficients and penalty parameters were also calculated. The parameters of the algorithm are also given in Table 4. The detected fault components and related spectrum by the NCPR and LFLO methods are demonstrated in Figures 9 and 10. From Figures 9 and 10, however, only one higher order line (39.42 Hz) associated with the gear chipping fault can be observed, so the fundamental frequency of failure of 13.33 Hz cannot be reliably identified. This may be because of the fact that the gear chipping fault, in this case, was quite weak and was submerged by strong noise. Therefore, it can be verified that the proposed ACPR method has an excellent performance with regards to recognizing the fault information of weak gearbox failure, in comparison with the NCPR and LFLO methods.

**Figure 9.** Cont.



**Figure 9.** The fault information extracted through the NCPR approach. (a) The time-domain waveform of extracted fault signal and (b) the envelope spectrum of the extracted fault signal.



**Figure 10.** The fault information extracted through the LFLO approach. (a) The time-domain waveform of the extracted fault signal and (b) the envelope spectrum of the extracted fault signal.

## 6. Conclusions

Aiming at the issue of estimating the sparse low-rank matrix (i.e., the weak fault signal) from its noise observation, this paper proposes a novel weak fault feature extraction approach for a gearbox, combining majorization–minimization (MM) and asymmetric convex penalty regularization (ACPR) algorithms. In this work, the observed noisy signal was modeled as the sum of the fault's transient impulses, the systematic natural vibration signal, and the additive noise; thus, the ill-posed problem of sparse representation was translated into a regularization inverse problem. The proposed objective function better estimates sparse low-rank matrices (SLRM) than the nonconvex and convex methods, which utilize the sum of the signal-fidelity term and the two parameterized penalty terms.

In terms of an algorithm solution, the asymmetric penalty function was established on the basis of the L1-norm model and simple quadratic equation forms, and a symmetric penalty function was established on the basis of the L1-norm model, respectively. Then, the majorizers of the symmetric and asymmetric functions were calculated via the majorization–minimization (MM) algorithm. Afterward, the efficient iterative algorithm was derived by the MM algorithm. Meanwhile, we show how to set the model parameters of the proposed objective cost function (OCF), in order to ensure that the objective function is strictly convex. Finally, the effectiveness of the proposed ACPR method is demonstrated by the simulation signal and practical gearbox fault data, compared with two state-of-art methods, especially for the measured signal with heavy background noise. Specifically, the gear

chipping characteristic frequency 13.22 Hz and its harmonic (2f, 3f, 4f and 5f) can be detected obviously. The diagnosis results show that the fault frequency and its harmonics could be detected clearly by the proposed method.

The proposed approach has great potential for online or offline detection for gearboxes, especially in the early stages of failure or when the background noise is heavy and uncorrelated interference is very strong. The proposed ACPR approach uses symmetric and asymmetric penalty functions to promote the positivity of the estimated fault signal; however, it should be noted that the periodic transient characteristics of fault impulses cannot be revealed by the current proposed methodology. Hence, future work will focus on formulating an improved algorithm to address this issue, which makes periodic transient impulses as similar to the signal's physical structure as possible.

For the applicability of the proposed algorithm, this algorithm might be applied to other rotating machinery, such as a rolling bearing or rotor, and even mechanical processing (e.g., turning chatter, milling chatter, tool wear etc.).

**Author Contributions:** Algorithm improvement, programming, experimental analysis, and paper writing were done by Q.L. Review and suggestions were provided by S.Y.L. All authors have read and approved the final manuscript.

**Funding:** This research was funded by the Fundamental Research Funds for the Central Universities (Grant No. CUSF-DH-D-2017059 and BCZD2018013) and the Research Funds of Worldtech Transmission Technology (Grant No. 12966EM).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jing, L.Y.; Zhao, M.; Li, P.; Xu, X.Q. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* **2017**, *111*, 1–10. [[CrossRef](#)]
2. Li, Y.B.; Li, G.Y.; Yang, Y.T.; Liang, X.H.; Xu, M.Q. A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy. *Mech. Syst. Signal Process.* **2018**, *105*, 319–337. [[CrossRef](#)]
3. Kia, S.H.; Henao, H.; Capolino, G.A. Fault index statistical study for gear fault detection using stator current space vector analysis. *IEEE Trans. Ind. Appl.* **2016**, *52*, 781–788. [[CrossRef](#)]
4. Teng, W.; Ding, X.; Zhang, X.L.; Liu, Y.B.; Ma, Z.Y. Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform. *Renew. Energy* **2016**, *93*, 591–598. [[CrossRef](#)]
5. Hemmati, F.; Orfali, W.; Gadala, M.S. Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation. *Appl. Acoust.* **2016**, *104*, 101–118. [[CrossRef](#)]
6. Qu, J.X.; Zhang, Z.S.; Gong, T. A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion. *Neurocomputing* **2016**, *171*, 837–853. [[CrossRef](#)]
7. Wang, L.H.; Zhao, X.P.; Wu, J.X.; Xie, Y.Y.; Zhang, Y.H. Motor fault diagnosis based on short-time Fourier transform and convolutional neural network. *Chin. J. Mech. Eng.* **2017**, *30*, 1357–1368. [[CrossRef](#)]
8. Bouchikhi, E.H.E.; Choqueuse, V.; Benbouzid, M.E.H. Current frequency spectral subtraction and its contribution to induction machines bearings condition monitoring. *IEEE Trans. Energy Convers.* **2013**, *28*, 135–144. [[CrossRef](#)]
9. Wang, H.C.; Chen, J.; Dong, G.M. Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform. *Mech. Syst. Signal Process.* **2014**, *48*, 103–119. [[CrossRef](#)]
10. Li, Q.; Liang, S.Y. Bearing incipient fault diagnosis based upon maximal spectral kurtosis TQWT and group sparsity total variation de-noising approach. *J. Vibroeng.* **2018**, *20*, 1409–1425. [[CrossRef](#)]
11. Li, R.Y.; He, D. Rotational machine health monitoring and fault detection using EMD-based acoustic emission feature quantification. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 990–1001. [[CrossRef](#)]
12. Yuan, J.; Ji, F.; Gao, Y.; Zhu, J.; Wei, C.J.; Zhou, Y. Integrated ensemble noise-reconstructed empirical mode decomposition for mechanical fault detection. *Mech. Syst. Signal Process.* **2018**, *104*, 323–346. [[CrossRef](#)]

13. Wang, L.; Liu, Z.W.; Miao, Q.; Zhang, X. Time-frequency analysis based on ensemble local mean decomposition and fast kurtogram for rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* **2018**, *103*, 60–75. [\[CrossRef\]](#)
14. Li, Q.; Liang, S.Y.; Song, W.Q. Revision of bearing fault characteristic spectrum using LMD and interpolation correction algorithm. *Procedia CIRP* **2016**, *56*, 182–187. [\[CrossRef\]](#)
15. Li, Z.X.; Jiang, Y.; Guo, Q.; Hu, C.; Peng, Z.X. Multi-dimensional variational mode decomposition for bearing-crack detection in wind turbines with large driving-speed variations. *Renew. Energy* **2018**, *116*, 55–73. [\[CrossRef\]](#)
16. Li, Q.; Ji, X.; Liang, S.Y. Incipient fault feature extraction for rotating machinery based on improved AR-minimum entropy deconvolution combined with variational mode decomposition approach. *Entropy* **2017**, *19*, 7. [\[CrossRef\]](#)
17. Zhang, R.; Tao, H.Y.; Wu, L.F.; Guan, Y. Transfer learning with neural networks for bearing fault diagnosis in changing working conditions. *IEEE Access* **2017**, *5*, 14347–14357. [\[CrossRef\]](#)
18. Masud, A.A.; Albarracín, R.; Rey, J.A.A.; Sukki, F.M.; Illias, H.A.; Bani, N.A.; Munir, A.B. Artificial Neural Network Application for Partial Discharge Recognition: Survey and Future Directions. *Energies* **2016**, *9*, 574. [\[CrossRef\]](#)
19. He, M.; He, D. Deep learning based approach for bearing fault diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065. [\[CrossRef\]](#)
20. Jia, F.; Lei, Y.G.; Guo, L.; Lin, J.; Xing, S.B. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* **2018**, *272*, 619–628. [\[CrossRef\]](#)
21. Li, Q.; Liang, S.Y. Multiple faults detection for rotating machinery based on Bi-component sparse low-rank matrix separation approach. *IEEE Access* **2018**, *6*, 20242–20254. [\[CrossRef\]](#)
22. Tang, H.F.; Chen, J.; Dong, G.M. Sparse representation based latent components analysis for machinery weak fault detection. *Mech. Syst. Signal Process.* **2014**, *46*, 373–388. [\[CrossRef\]](#)
23. Zhou, H.T.; Chen, J.; Dong, G.M.; Wang, R. Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model. *Mech. Syst. Signal Process.* **2016**, *72–73*, 65–79. [\[CrossRef\]](#)
24. Yang, B.Y.; Liu, R.N.; Chen, X.F. Fault diagnosis for a wind turbine generator bearing via sparse representation and shift-invariant K-SVD. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1321–1331. [\[CrossRef\]](#)
25. Feng, Z.P.; Liang, M. Complex signal analysis for planetary gearbox fault diagnosis via shift invariant dictionary learning. *Measurement* **2016**, *90*, 382–395. [\[CrossRef\]](#)
26. Qin, Y. A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2018**, *65*, 2716–2726. [\[CrossRef\]](#)
27. Cui, L.L.; Wang, J.; Lee, S. Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis. *J. Sound Vib.* **2014**, *333*, 2840–2862. [\[CrossRef\]](#)
28. Cui, L.L.; Gong, X.Y.; Zhang, J.Y.; Wang, H.Q. Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel-Ziv complexity. *J. Sound Vib.* **2016**, *385*, 372–388. [\[CrossRef\]](#)
29. Cui, L.L.; Wu, N.; Ma, C.Q.; Wang, H.Q. Quantitative fault analysis of roller bearings based on a novel matching pursuit method with a new step-impulse dictionary. *Mech. Syst. Signal Process.* **2016**, *68–69*, 34–43. [\[CrossRef\]](#)
30. Ding, X.X.; He, Q.B. Time-frequency manifold sparse reconstruction: A novel method for bearing fault feature extraction. *Mech. Syst. Signal Process.* **2016**, *80*, 392–413. [\[CrossRef\]](#)
31. Li, Q.; Liang, S.Y. Incipient fault diagnosis of rolling bearings based on impulse-step impact dictionary and re-weighted minimizing nonconvex penalty Lq regular technique. *Entropy* **2017**, *19*, 8. [\[CrossRef\]](#)
32. Ding, Y.; He, W.P.; Chen, B.Q.; Zi, Y.Y.; Selesnick, I.W. Detection of faults in rotating machinery using periodic time-frequency sparsity. *J. Sound Vib.* **2016**, *382*, 357–378. [\[CrossRef\]](#)
33. He, W.P.; Ding, Y.; Zi, Y.Y.; Selesnick, I.W. Sparsity-based algorithm for detecting faults in rotating machines. *Mech. Syst. Signal Process.* **2016**, *72–73*, 46–64. [\[CrossRef\]](#)
34. He, W.P.; Ding, Y.; Zi, Y.Y.; Selesnick, I.W. Repetitive transients extraction algorithm for detecting bearing faults. *Mech. Syst. Signal Process.* **2017**, *84*, 227–244. [\[CrossRef\]](#)
35. Zhang, H.; Chen, X.F.; Du, Z.H.; Li, X.; Yan, R.Q. Nonlocal sparse model with adaptive structural clustering for feature extraction of aero-engine bearings. *J. Sound Vib.* **2016**, *368*, 223–248. [\[CrossRef\]](#)
36. Du, Z.H.; Chen, X.F.; Zhang, H.; Yang, B.Y.; Zhai, Z.; Yan, R.Q. Weighted low-rank sparse model via nuclear norm minimization for bearing fault detection. *J. Sound Vib.* **2017**, *400*, 270–287. [\[CrossRef\]](#)



37. Mourad, N.; Reilly, J.P.; Kirubarajan, T. Majorization minimization for blind source separation of sparse sources. *Signal Process.* **2017**, *131*, 120–133. [[CrossRef](#)]
38. Qiu, T.Y.; Palomar, D.P. Undersampled sparse phase retrieval via majorization minimization. *IEEE Trans. Signal Process.* **2017**, *65*, 5957–5969. [[CrossRef](#)]
39. Ndoye, M.; Anderson, J.M.M.; Greene, D.J. An MM-based algorithm for  $\ell_1$ -regularized least-squares estimation with an application to ground penetrating radar image reconstruction. *IEEE Trans. Image Process.* **2016**, *25*, 2206–2221. [[CrossRef](#)] [[PubMed](#)]
40. Donoho, D.L.; Elad, M.; Temlyakov, V.N. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory* **2006**, *52*, 6–18. [[CrossRef](#)]
41. Donoho, D.L. Compressed sensing. *IEEE Trans. Inform. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
42. Donoho, D.L.; Tsai, Y. Fast solution of  $\ell_1$ -norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory* **2008**, *54*, 4789–4812. [[CrossRef](#)]
43. Donoho, D.L. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [[CrossRef](#)]
44. Selesnick, I.W. Total variation denoising via the Moreau envelope. *IEEE Signal Process. Lett.* **2017**, *24*, 216–220. [[CrossRef](#)]
45. Selesnick, I.W.; Parekh, A.; Bayram, I. Convex 1-D total variation denoising with non-convex regularization. *IEEE Signal Process. Lett.* **2015**, *22*, 141–144. [[CrossRef](#)]
46. Selesnick, I.W.; Graber, H.L.; Pfeil, D.S.; Barbour, R.L. Simultaneous low-pass filtering and total variation denoising. *IEEE Trans. Signal Process.* **2014**, *62*, 1109–1124. [[CrossRef](#)]
47. Ning, X.R.; Selesnick, I.W.; Duval, L. Chromatogram baseline estimation and denoising using sparsity (BEADS). *Chemom. Intell. Lab.* **2014**, *139*, 156–167. [[CrossRef](#)]
48. Donoho, D.L.; Johnstone, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **1993**, *81*, 425–455. [[CrossRef](#)]
49. Annual Conference of the Prognostics and Health Management Society 2009. Available online: <https://www.phmsociety.org/events/conference/phm/09> (accessed on 5 June 2018).
50. Atat, H.A.; Siegel, D.; Lee, J. A systematic methodology for gearbox health assessment and fault classification. *Int. J. Progn. Health Manag.* **2011**, *2*, 1.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).