Multi-task neural network blind deconvolution and its application to bearing fault feature extraction

Jing-Xiao Liao[a], Hang-Cheng Dong[a], Lei Luo[b], Jinwei Sun[a] and Shiping Zhang[a,*]

[a]School of Instrumentation Engineering, Harbin Institute of Technology, Harbin, 150001, China
[b]Key Laboratory of Optoelectronic Technology & Systems, Education Ministry of China, Chongqing University, Chongqing, 400044, China

ARTICLE INFO

Keywords:
Bearing fault feature extraction
Blind deconvolution(BD)
Multi-task optimization
Convolutional neural network

ABSTRACT

Blind deconvolution (BD) is one of the effective methods that help pre-process vibration signals and assist in bearing fault diagnosis. Currently, most BD methods design an optimization criterion and use frequency or time domain information independently to optimize a deconvolution filter. It recovers weak periodic impulses related to incipient faults. However, the random noise interference may cause the optimizer to overfit. The time-domain-based BD methods tend to extract fault-unrelated single peak impulse, and the frequency-domain-based BD methods tend to retain the maximum energy frequency component, which will lose the fault-related harmonics frequency components. To solve the above issue, we propose a hybrid criterion that combines the kurtosis for time domain optimization and the $G - l_1/l_2$ norm for the frequency domain. These two criteria are monotonically increasing and decreasing, so they mutually constrain to avoid overfitting. After that, we design a multi-task one-dimensional convolutional neural network with time and frequency branches to achieve an optimal solution for this hybrid criterion. The multi-task neural network realizes the simultaneous optimization of two domains. Experimental results show that our proposed method outperforms other state-of-the-art methods.

1. Introduction

Rolling element bearings (REBs) are the core component of the widely-employed rotating machinery. Due to working in high load, high speed, high temperature and other harsh environments, they are easily prone to damage [1, 2]. Several statistics show that bearing faults are responsible for 40 to 70 percent of rotating machinery failures [3]. These accidental failures may cause damage to economic property and personal safety. Hence, it is valuable to detect incipient faults and determine the type of faults. A commonly-used approach sets accelerometers on the rotating machinery housing and measures the vibration signal. Nevertheless, due to the complexity of mechanical systems, the multi-transmission paths and background noise affect identifying fault from the measured signal [4]. Because the noise drowns out the fault signature, it is hard to observe fault-related repetitive transient impulses from the measured signal.

In this context, researchers proposed many efficient approaches for fault feature extraction. The representation of them were wavelet transform (WT) [5–8], empirical mode decomposition (EMD) [9, 10], variational mode decomposition (VMD) [11, 12], singular value decomposition (SVD) [13, 14], spectral kurtosis (SK) [15–17] and blind deconvolution (BD) [2, 18, 19]. In particular, the BD methods optimized finite impulse response (FIR) filter coefficients, and the filter recovered the fault-related repetitive transient impulses, which helped identify incipient faults. The methods counteracted the effect of the transmission path, which tried to address this main challenge of fault signal feature extraction [1]. Moreover, the BD methods optimized the best FIR filter adaptively, so the bandwidth and center frequency unrestricted them. At last, the BD methods had a wide range of applications. They can act as a stand-alone feature extraction method for fault signals denoising and interpretation or serve as a pre-processing tool to assist in downstream faults diagnosis tasks [20–24].

The traditional BD methods used time domain signals for optimization. Minimum entropy deconvolution (MED) was the first BD method proposed in 1978 [25]. It used the kurtosis [26] as the BD optimization criterion and updated filter coefficients by maximizing the kurtosis value. However, the MED was susceptible to the random impulse interference that misguided to recover a fault-unrelated individual impulse [27]. To overcome the shortcomings, McDonald et al. [28] proposed so-called maximum correlated kurtosis deconvolution (MCKD). MCKD integrated

*Corresponding author
thumbnails/casemail.jpeg (J. Liao); spzhang@hit.edu.cn (S. Zhang)
ORCID(s): 0000-0002-1880-2621 (J. Liao); 0000-0001-9329-8894 (S. Zhang)

the periodicity of the faulty signal and used correlated kurtosis (CK) as an optimization criterion. Along this direction, sparse maximum harmonics-to-noise-ratio deconvolution (SHMD) [29], multipoint optimal MED adjusted (MOMEDA) [30], and second-order cyclostationarity blind deconvolution (CYCBD) [19] recovered repetitive transient impulses by maximizing harmonics-to-noise-ratio (HNR), multi-D-norm, and cyclostationary criterion, respectively. Nonetheless, these methods required a cyclic period beforehand to recover the fault-related signal. This limited BD methods to deploy in real industrial scenarios.

Several priori-unknown BD methods aimed to address the cyclic period requirement were twofold. The first category of methods used periodicity detection techniques (PDTs) to estimate the cyclic period in the frequency domain, then fed it to a period-dependent time domain optimization criterion [31]. PDTs utilized the squared envelope spectrum (ES) of the signal and identified fault period by some criteria, such as autocorrelation function (AF) [32], periodic modulation intensity function (PMIF) [33], and envelope harmonic product spectrum (EHPS) [27, 34] etc. However, PDTs may obtain a wrong cyclic period when encountering complex signal frequency components, then affect the downstream BD methods.

Another category of methods optimized target function using the envelope spectrum. In the frequency domain, the feature of the fault signal consists of a series of cyclic frequencies, and its envelope spectrum is sparse. Therefore, optimising the sparsity criterion such as envelope spectra l_2/l_1 [35] and envelope spectra kurtosis (ESK) [36] can recover a sparse envelope spectrum, which represents cyclic frequencies of the fault signal. Recently, Liu He et al. [37] proposed a priori-unknown blind deconvolution by optimising the generalized l_p/l_q of the envelope spectrum (Mini-blp-lplq). The envelope spectrum-based optimization method showed competitive results. However, the characteristics of the ball and inner race fault consist of multiple frequency components, and experimental results show that optimizing with only frequency domain features resulted in frequency component loss.

Intuitively, if we integrate time and frequency optimization methods: i) This method is sensitive to extracting repetitive transient impulses in the time domain. ii) This method recovers fault-related cyclic frequencies in the frequency domain. We aim to achieve complementary optimization in the time and frequency domains and avoid overfitting caused by independent optimization in the time or frequency domains. To achieve this, we optimize the time and frequency domains using two criteria with opposite monotonicity so that they constrain mutually. Moreover, the multi-task convolutional neural network [38] has successfully dealt with multi-task learning problems. We introduce it into this multi-task BD optimization issue. Such that we can optimize the time and frequency criterion simultaneously. The convolutional neural network acts as an end-to-end optimizer without manual derivation of the gradients from the target function to filter coefficients, and neural networks showed excellent performance in solving non-smooth non-convex optimization problems [39–41]. Therefore, using the convolutional neural network to solve the BD problem is a promising option.

Main contributions. i) We propose a hybrid optimization criterion that combines the time domain criterion kurtosis and the frequency domain criterion $G - l_1/l_2$ norm, then derive the characteristics of this criterion. ii) We design a multi-task one-dimensional convolutional neural network (1DCNN) with two branches to achieve the joint optimization of two criteria. To our best knowledge, it is the first time we use a multi-task convolutional neural network for BD problems. iii) Experiments on simulated and real-world bearing fault signals show that our method outperforms other state-of-the-art methods.

2. Methodology

2.1. Review of blind deconvolution

Assuming that the fault source and the noise pass multiple transmission paths and blend into the measured signal [2, 18]. Formally, denote raw signal $\boldsymbol{y} \in \mathbb{R}^n$, fault impulse component $\boldsymbol{d} \in \mathbb{R}^n$ and additive noise $\boldsymbol{n} \in \mathbb{R}^n$, the relationship between them is expressed as

$$\boldsymbol{y} = \boldsymbol{d} * \boldsymbol{h}_d + \boldsymbol{n} * \boldsymbol{h}_n, \tag{1}$$

where, $\boldsymbol{h}_d$, $\boldsymbol{h}_n$ are transfer functions and $*$ is the convolution operation. Hereafter, we use bold italic letters to represent the matrix. The idea of blind deconvolution methods is to search a FIR filter $\boldsymbol{f} \in \mathbb{R}^L$ for recovering the fault-related signal $\boldsymbol{x} \in \mathbb{R}^n$. We have,

$$\boldsymbol{x} = \boldsymbol{y} * \boldsymbol{f} \approx \boldsymbol{d}. \tag{2}$$

The deconvolution filter expects to extract fault-related signals and removes the interference of additive noise, random pulses, etc. Because the transfer function is unknown and we can only estimate filter coefficients without actual labels, the optimization of the deconvolution filter is "blind".

 doi:10.20944/preprints202211.0

The BD problem is an optimization problem. By maximizing the optimization criterion $K(\cdot)$, the optimizer finds the best deconvolution filter coefficients, which can be used to recover the fault signal. The criterion is as follows,

$$\begin{aligned} &\max_{\boldsymbol{f}} K(\boldsymbol{x}) \\ &s.t. \quad \boldsymbol{x} = \boldsymbol{y} * \boldsymbol{f}, \ \|\boldsymbol{f}\|_{l_2} = 1. \end{aligned} \tag{3}$$

The optimizer iterates and stops when the criterion converges,

$$K(\boldsymbol{x}_j) - K(\boldsymbol{x}_i) \leq \epsilon \quad j > i \in \mathbb{N}^+, \tag{4}$$

with ϵ being a small constant. Some effective optimization methods are applied to the BD problem, such as matrix operations [19, 25], particle swarm optimization (PSO) [42, 43], backpropagation [41, 44], etc. Different optimization methods affect the performance of the BD methods.

The core of the BD methods is the optimization criterion. These criteria describe the essential characteristics of the fault signal. The kurtosis is the optimization criterion that used in the first-proposed MED,

$$\text{Kurtosis}: \quad K_1 = \frac{\sum_{n=1}^{N} x(n)^4}{(\sum_{n=1}^{N} x(n)^2)^2}. \tag{5}$$

The kurtosis is defined by the fourth standardized moment of the real-valued random variable divided by its standard deviation [26]. As the kurtosis value increases, the data distribution is further away from the standard normal distribution. Because the characteristic of bearing fault signals is repetitive transient impulses, the higher kurtosis value means it has more outliers (peaks), which implies that it has more significant fault signal characteristics. In brief, as an optimization criterion, the kurtosis guides the filter to achieve the sparsity-promoted solution.

Other sparsity criteria were developed for solving the BD problem, including GaryâĂŹs variable [45], skewness [46], l_2/l_1 norm [47]. L Li et al. [48] proposed a generalized sparsity criterion called $G - l_p/l_q$ that can represent a series of sparse criteria, which is expressed as

$$G - l_p/l_q: \quad K_2 = \log\left(\frac{q}{p}\right) \cdot \frac{\sum_{n=1}^{N} |x(n)|^p}{(\sum_{n=1}^{N} |x(n)|^q)^{\frac{p}{q}}}. \ p > 0, q > 0. \tag{6}$$

When $p = 4$, $q = 2$, the $G - l_p/l_q$ norm is the kurtosis. In particular, the values of p and q determine the direction of optimization. When $p < q$, we need to minimize $G - l_p/l_q$ norm to find the sparsity-promoted solution. When $p > q$, we need to maximize it [37, 49].

2.2. The proposed method

2.2.1. Multi-task neural network blind deconvolution

The proposed multi-task neural network blind deconvolution (MNNBD) combines the kurtosis and $G - l_1/l_2$ norm as a hybrid optimization criterion. The kurtosis is used for time domain signal optimization, and the $G - l_1/l_2$ norm is used for the frequency domain. It complements the shortcomings of these two types of BD methods: i) The time domain-based method (MED [25]), which used the kurtosis as the criterion, may be affected by random pulses without fault characteristic frequency information; ii) The frequency domain-based method (Mini-blp-lplq [37]), which used the $G - l_p/l_q$ norm as the criterion, leads to the high sparsity of the frequency domain and loses the low energy fault-related frequency features. The MNNBD leverages the two criteria with opposite monotonicity and allows the optimizer to be mutually constrained in the time and frequency domains to avoid overfitting. Therefore, the proposed method has stable performance in several scenarios.

An overview of the MNNBD is shown in Figure 1. First, combining Eq. (5) and Eq. (6), the hybrid optimization criterion is as follows,

$$\begin{aligned} \min_{\boldsymbol{W}} K(\boldsymbol{x}) &= -K_1(\boldsymbol{x}) + \lambda \cdot K_2(\boldsymbol{es}) \\ &= -\frac{\sum_{n=1}^{N} |x(n)|^4}{(\sum_{n=1}^{N} |x(n)|^2)^2} + \lambda \cdot \text{sign}(\log(2)) \cdot \frac{\sum_{n=1}^{N} |es(n)|}{(\sum_{n=1}^{N} |es(n)|^2)^{\frac{1}{2}}} \\ s.t.\ &\boldsymbol{W} * \boldsymbol{y}, \ \|W\| = 1, \\ &\boldsymbol{es} = |\text{FFT}(|\boldsymbol{x} + i\text{Hilbert}(\boldsymbol{x})|)|, \end{aligned} \tag{7}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

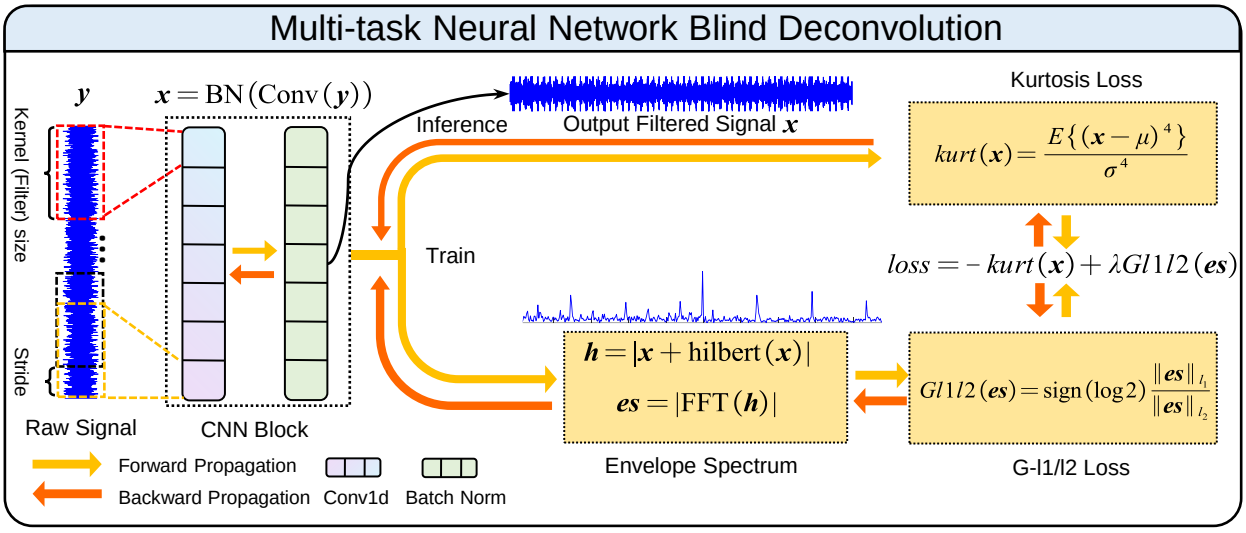


Figure 1: An overview of the multi-task blind deconvolution neural network. The key idea is to train a convolution neural network that combines the time domain and frequency domain loss function.

where, sign(·) denotes the signum function, which guarantees the direction of optimization; λ denotes a scale factor, which adjusts the weight of the time and the frequency domain criterion. λ is set to $0.1 \sim 0.001$ as an empirical constant. Specifically, we minimize the sum of the negative kurtosis and $G - l_1/l_2$ norm because deep learning optimizers are accustomed to minimizing the loss function.

Second, a one-layer, one-channel 1DCNN is used as a deconvolution filter. In particular, we pad zeros to the input so that the convolutional output has the same shape as the input. Following by a batch normalization (BN) layer that smooths the optimization landscape. It benefits the neural network to search for the global optimal solution [50]. The expression of the deconvolution neural network is

$$\boldsymbol{x} = \mathrm{BN}(\mathrm{Conv}(\boldsymbol{y})), \tag{8}$$

where,

$$\begin{aligned} \hat{\boldsymbol{y}} &= \mathrm{Conv}(\boldsymbol{y}) = \boldsymbol{W} * \boldsymbol{y}, \\ x(n) &= \mathrm{BN}(\hat{y}(n)) = \gamma \frac{\hat{y}(n) - \mu}{\sigma} + \beta. \end{aligned} \tag{9}$$

Here, $\boldsymbol{W} \in \mathbb{R}^{1\times L}$ is the weight matrix of the convolutional kernel (deconvolution filter); γ and β are the learnable parameters; μ, σ denote the mean and the standard deviation of $\boldsymbol{z}$, respectively. With the help of deep learning optimized tensor libraries, i.e., PyTorch and Tensorflow, the BD optimization problem can be solved directly using automatic differentiation without manually calculating derivatives of complex functions. Furthermore, the deep learning optimizer (SGD [51], ADAM [52] and RMSprop[53]) is helpful for an optimal solution.

2.2.2. Characteristics of the optimization criterion

In this subsection, we derive the different characteristics of the kurtosis and $G - l_1/l_2$ norm. Because the $G - l_p/l_q$ norm can present two parts of the hybrid optimization criterion, we first calculate the first-order derivatives of $G - l_p/l_q$ and then analyze the differences and effects of the two parts.

The $G-l_p/l_q$ norm can be expanded as [37]

$$\begin{aligned} K &= \left(\frac{\|\boldsymbol{x}\|_{l_p}}{\|\boldsymbol{x}\|_{l_q}}\right)^p \\ &= \frac{\sum_{n=1}^{N}|x(n)|^p}{\left(\sum_{n=1}^{N}|x(n)|^q\right)^{p/q}} \\ &= \frac{\sum_{n=1,n\neq n'}^{N}|x(n)|^p+|x(n')|^p}{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{p/q}}, \end{aligned} \tag{10}$$

where, n' is the index of the maximum value of $|\boldsymbol{x}|$. Consequently, the derivative of the maximum absolute value for $|\boldsymbol{x}|$ is as follows,

$$\begin{aligned} &\frac{\partial K(\boldsymbol{x})}{\partial(|x(n')|)} = \\ &\frac{p|x(n')|^{p-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{p/q}}{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{2p/q}} \\ &-\frac{\left(\frac{p}{q}\right)\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{p/q-1}q|x(n')|^{q-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^p+|x(n')|^p\right)}{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{2p/q}} \\ &=\frac{p|x(n')|^{p-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{p/q}}{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{2p/q}} \\ &-\frac{p|x(n')|^{q-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{p/q-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^p+|x(n')|^p\right)}{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)^{2p/q}}. \end{aligned} \tag{11}$$

The numerator of Eq. (11) can be simplified to

$$\begin{aligned} &|x(n')|^{p-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)-|x(n')|^{q-1}\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^p+|x(n')|^p\right) \\ =&\frac{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q\right)}{|x(n')|^q}-\frac{\left(\sum_{n=1,n\neq n'}^{N}|x(n)|^p+|x(n')|^p\right)}{|x(n')|^p}, \end{aligned} \tag{12}$$

when $p>q>0$, the following inequality holds

$$\frac{\sum_{n=1,n\neq n'}^{N}|x(n)|^q+|x(n')|^q}{|x(n')|^q}>\frac{\sum_{n=1,n\neq n'}^{N}|x(n)|^p+|x(n')|^p}{|x(n')|^p}, \tag{13}$$

substituting Eq.(13) into Eq.(11), we have

$$\frac{\partial K(\boldsymbol{x})}{\partial|x(n')|}>0, \tag{14}$$

similarly, when $0<p<q$, we have

$$\frac{\partial K(\boldsymbol{x})}{\partial|x(n')|}<0. \tag{15}$$

Inequalities (14) and (15) show that the $G - lplq$ is a monotonic function respective of the relationship between p and q. That's why we maximize the kurtosis but minimize the $G - l_1/l_2$ norm.

Furthermore, we generate Bernoulli distributed random variables ($P(X = 1) = p_0$, $P(X = 0) = 1 - p_0$, $0 < p_0 < 1$) with 100 points and calculate the $G - l_4/l_2$ (kurtosis) and $G - l_1/l_2$ norm values. p_0 denotes the probability of random variables being 1. The smaller p_0 means variables have a higher level of sparsity. The characteristics of two $G - l_p/l_q$ norms are shown in Figure 2. Here, we obtain the following observations: i) When $p > q$, the $G - l_p/l_q$ has a monotonically increasing relationship with sparsity; when $p < q$, it shows the opposite property. ii) The l_4/l_2 norm is exponentially associated with the change of sparsity, while the l_1/l_2 norm is almost linearly related to the change of sparsity. The phenomenon is determined by $p > q$ or $p < q$ [49]. iii) The value of the l_1/l_2 norm is 100 times larger than that of the l_4/l_2 norm, so we set a scale factor λ to keep their equilibrium.

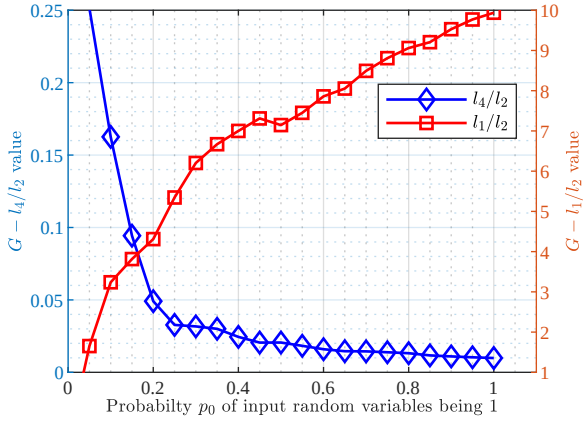


Figure 2: Comparison of two $G-l_p/l_q$ norm values versus sparsity of input random variables. The smaller p_0 is, the variables are higher level of sparsity.

2.2.3. Discussion of the optimization function

Overfitting will lose some fault characteristics components, so the time and frequency optimization criteria should be mutually constrained and kept within a suitable range. Two $G-l_p/l_q$ norm criteria optimized in opposite directions are well-suited to achieve this work. Here, an experiment confirms this opinion. The result is shown in Figure 3. First, comparing the two graphs in the first and second rows, the same criteria used in time and frequency domains can not recover the correct signal. Because the criteria have the same monotonicity, optimizing the time and frequency domains is unrestricted. Second, comparing the envelope spectrum curves in the third and fourth rows, the kurtosis tends to extract a single principal frequency component, but the $G - l_1/l_2$ norm extracts more principal frequency components. Because we expect to obtain sufficient fault frequency components, the $G - l_1/l_2$ norm is more suitable for the frequency domain. In summary, using the kurtosis for the time domain optimization and the $G - l_1/l_2$ norm for the frequency domain shows the most desirable results.

3. Experiment

In this section, we conduct three case studies to investigate our proposed method: One simulated fault signal with heavy noise, two real-world vibration signals with a manually injected bearing faulty and an accelerated degradation bearing faulty.

3.1. Experimental setups

Settings. We choose the same initial parameters for all experiments to ensure fairness. We select five representative methods as the baseline, MED [25], MCKD [28], Mini-blp-lplq [37], SF-SLSN [54], and BADBD using the kurtosis as the optimization criterion [41]. All baseline methods have the official implementation on MATLAB, and we use their recommended initial parameters. In particular, we implement our method in Python 3.8 with PyTorch, an open-source deep learning framework. The kernel size of MNNBD is 1×80. We utilize Kaiming initialization for kernel weight

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

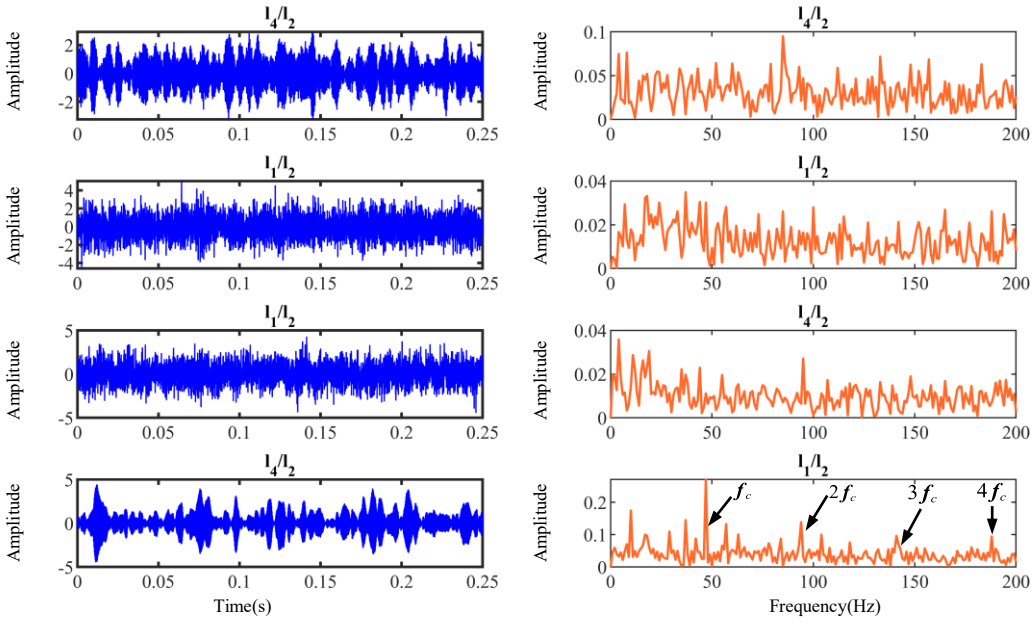


Figure 3: Comparing different combinations of optimization criterion. f_c denotes the fault characteristic. Only using the kurtosis for time domain optimization and the $G - l_1/l_2$ norm for frequency domain can exact the correct fault feature.

Table 1
The cyclic characteristic frequencies of different bearing defects signals. IR denotes inner race, OR denotes outer race.

IR Defect	OR Defect	Ball Defect
$n\text{BPFI} \pm k f_r \ (n, k \in Z)$	$n\text{BPFO} \ (n \in Z)$	$n\text{BSF} \pm k\text{FTF} \ (n, k \in Z)$

parameters and use Adam as the optimizer with 200 epochs of training. The learning rate and scale factor λ are set to 0.01. All experiments are conducted in Windows 10 with an Intel i9 10900k CPU at 3.70 GHz and one NVIDIA RTX 3080Ti 12GB GPU. At last, we use the kurtosis and the envelop spectrum of the filtered signal to evaluate the performance of the compared methods.

Bearing fault characteristic frequency. When the balls periodically pass through defective points, it causes an impulse response to the vibration signal, defined as the bearing defect's characteristic frequency. Let n, d, and D be the number of balls, the ball diameter, and the bearing pitch diameter, respectively. f_r denotes the shaft rotation speed, and ϕ denotes the angle of the load from the radial plane. The characteristic frequencies are represented as follows,

$$\begin{cases} \text{BPFO} = \frac{nf_r}{2}\left(1 - \frac{d}{D}\cos\phi\right) \\ \text{BPFI} = \frac{nf_r}{2}\left(1 + \frac{d}{D}\cos\phi\right) \\ \text{FTF} = \frac{f_r}{2}\left(1 - \frac{d}{D}\cos\phi\right) \\ \text{BSF} = \frac{Df_r}{2d}\left(1 - \left(\frac{d}{D}\cos\phi\right)^2\right), \end{cases} \tag{16}$$

where BPFO, BPFI is the ball pass frequency of the outer and inner races, FTF is the fundamental train frequency (cage speed), and BSF is the ball spin frequency [55, 56]. Moreover, the cyclic characteristic frequencies of different bearing defects signals are shown in Table 1. Hereafter, we use the above frequencies to identify the characteristics of the fault signals.

Table 2
The summary of the simulated signal parameters

Periodic impulses		Random shocks	
f_i	4000Hz	f_j	6000Hz
ζ_r^i	0.1Hz	ζ_r^j	0.1Hz
T_α	1/47s	φ_i	0
φ_i	0		
Periodic harmonic		Noise	
C_1/C_2	0.175/0.145	SNR	-15dB
β_1/β_2	$\pi/2/\pi/2$	Sampling rate	20kHz
f_1/f_2	10Hz/20Hz	Time	1s

3.2. Case 1: simulated fault signal with heavy noise

In this case, we use simulated signals to verify the performance of the proposed method. The mathematical model of a single bearing fault vibration signal refers to [32]. It consists of four components: i) The rolling ball passes through the fault defect and generates the periodic fault impulse responses. ii) The incidental knocking or electromagnetic interference cause the random impulses. iii) The gear meshing, shaft rotation, or other circular motions cause harmonic interference. iv) The Gaussian white noise is also superimposed on the signal model. The equation is as follows,

$$x(t) = \underbrace{\sum_{i=1}^{M_0} A_i s_i\left(t - iT_\alpha - \tau_i\right)}_{\text{fault impulse}} + \underbrace{\sum_{j=1}^{M_1} B_j s_j\left(t - T_d\right)}_{\text{random impulse}} + \underbrace{\sum_{k=1}^{M_2} C_k \cos\left(2\pi f_k t + \beta_k\right)}_{\text{harmonic interference}} + \underbrace{n(t)}_{\text{noise}}, \tag{17}$$

where, A_i denotes the magnitude of the i-th shock, T_α is the time interval of periodic shocks, τ_i is a random variable indicating rollers random slippage and usually set to $\tau = 1\% \sim 2\% T_\alpha$. B_j, T_d are random variables. C_k, f_k, and β_k are parameters of periodic harmonic signal. $s_i(t)$ denotes the impulse response of the single-degree-of-freedom (SDOF) system. The expression is,

$$s_i(t) = e^{-\eta t}\cos\left(2\pi f_i t + \varphi_i\right), \tag{18}$$

where, $\eta = f_r\zeta/\sqrt{1-\zeta^2}$ with ζ, f, and φ are the damping factor, resonant frequency of the SDOF, and the initial phase of the i-th impulse response, respectively.

We generate an inner race fault simulation signal referred to [27]. The fault characteristic frequency is 47Hz, with the Gaussian noise's signal-to-noise ratio (SNR) set to -15dB to simulate heavy noise. Table 2 shows the parameters of the simulated signal. All BD methods filter 20,000 point signals directly to get the results.

The results are shown in Figure 4 and Table 3. First, Table 3 shows that all BD methods extract sharp peaks from the strong noise signal, making the kurtosis value of the filtered signal higher than the raw signal. In particular, the MNNBD has the highest kurtosis, which implies it has the most significant impulse responses. The time domain waveforms in Figure 4 show that the MNNBD recovers a discernible impulse response than others, and the filtered signal has less burr. Moreover, it has the fastest computation time, which suggests that the MNNBD has real-time processing capabilities with Pytorch-based GPU acceleration. Second, from the envelope spectrum, we find that the Mini-blp-lplq, SF-SLSN, Kurt-BADBD and MNNBD can extract the fault frequency and its harmonics from the raw signal. But the MED and MCKD fail to extract the fault frequency under strong noise, and the MCKD extracts the single peak impulse in the time domain. At last, the experimental results illustrate that the MNNBD has a stable fault feature extraction capability under strong noise interference.

3.3. Case 2: manually injected fault signals under noisy condition by HIT

In this case, we conduct experiments using our collected dataset [22]. We carry out a bearing fault test in MIIT Key Laboratory of Aerospace Bearing Technology and Equipment, Harbin Institute of Technology. Figure 5 shows our bearing test rig. We utilize angular contact ball bearings HC7003 for the test. This bearing is used for high-speed rotating machines compared to a common deep groove ball bearing. The accelerometer is directly attached to a bearing to collect vibration signals. Consistent with the traditional dataset, we inject faults at the outer race (OR), inner race

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

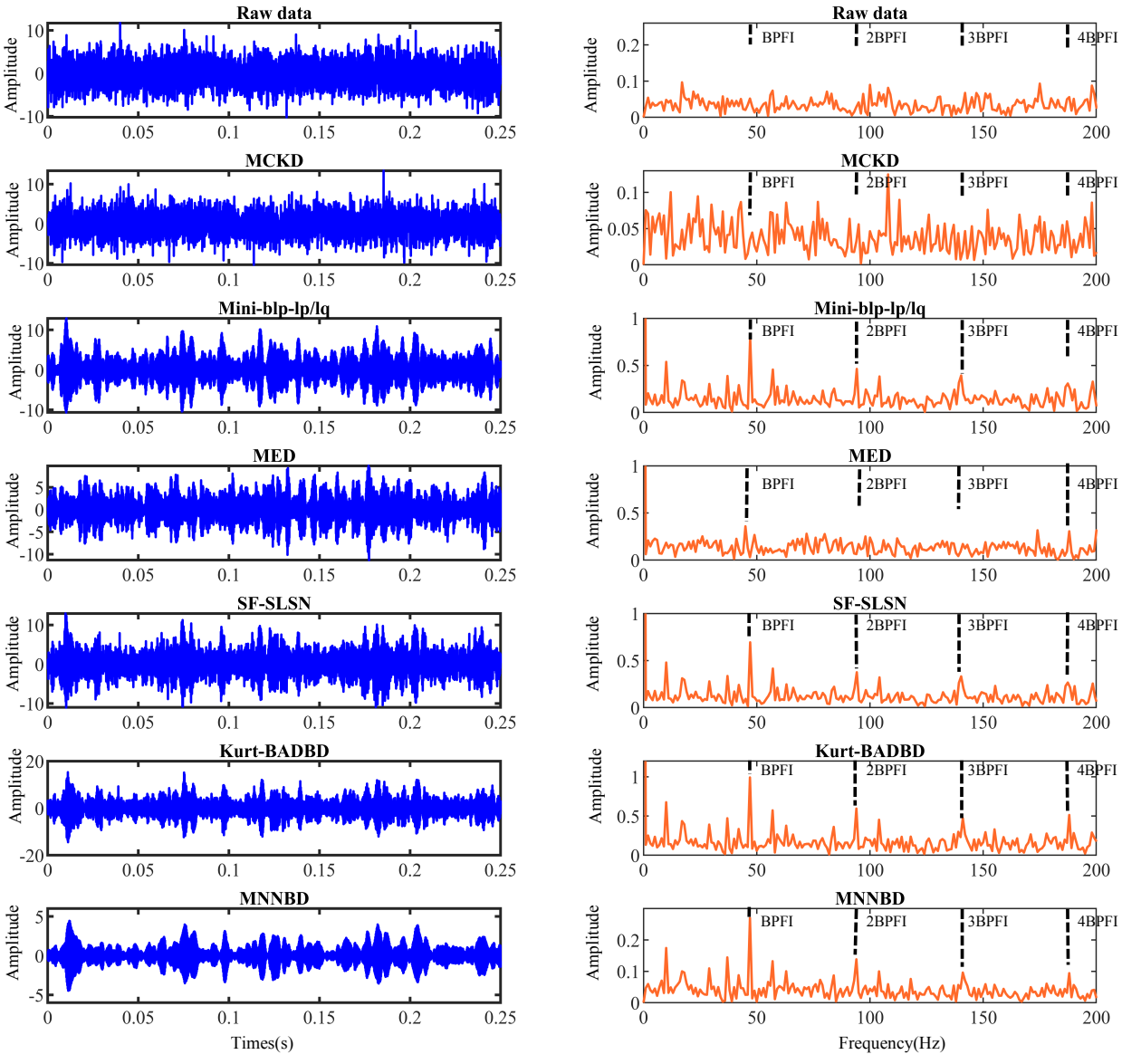


Figure 4: The results of deconvolution methods in time domain waveform and envelope spectrum on the simulated signal. BPFI is the ball pass frequency of the inner race.

(IR), and ball with three levels (minor, moderate, severe). Table 4 summarizes ten classes of bearings in our dataset. In the test, we choose the constant motor speed (1800 r/min) and use NI USB-6002 to collect vibration signals with the 12kHz sampling rate. We record 47s of bearing vibration (561,152 points per category). Different from the traditional dataset, bearing faults here are cracks of the same size yet different depths. Therefore, the vibration signals between different faults are more similar, making the noise affects the raw signal easier. Figure 6 shows the raw signals in the time and frequency domain. Note that the fault characteristic frequencies of the Ball defect 1 and 2 are unremarkable compared to other faults. It is due to the defect of the ball being slight, but there are variations compared to the healthy bearing signal. To further distinguish the feature extraction capability of the compared methods, we add 0dB noise to the raw signal.

For each category of fault signals, we cut 24000 points for blind deconvolution. The results are shown in Table 5 Figure 7. First, as shown in Figure 7, the noise heavily affects the time domain of the signal and leads the periodic

Table 3
The kurtosis value and run time (per 20000 points) of compared methods on the simulated signal.

Methods	Kurtosis ↑	Time cost(s) ↓
Raw Signal	3.0840	-
MCKD	3.2379	4.8775
Mini-blp-lplq	3.5989	5.2037
MED	3.8465	2.4657
SF-SLSN	3.6283	1.7546
Kurt-BADBD	3.7471	5.7824
MNNBD	**4.0901**	**1.1778**

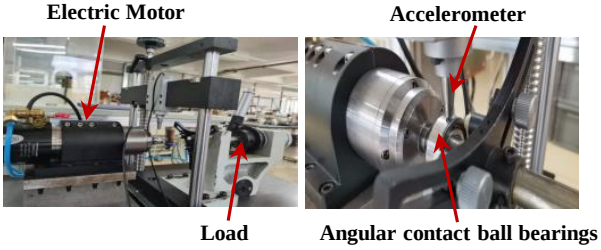


Figure 5: The test rig for angular contact ball bearing.

Table 4
Ten classes in our dataset. OR and IR denote that the faults appear at outer race and inner race, respectively.

Label	Fault Mode	Label	Fault Mode
C1	Health	C6	OR cracking (Moderate)
C2	Ball cracking (Minor)	C7	OR cracking (Severe)
C3	Ball cracking (Moderate)	C8	IR cracking (Minor)
C4	Ball cracking (Severe)	C9	IR cracking (Moderate)
C5	OR cracking (Minor)	C10	IR cracking (Severe)

impulses to be completely drowned out. Similarly, the fault characteristic frequencies are attenuated, and the noise swamp higher harmonics. Second, comparing BD methods, we find that the MCKD and Kurt-BADBD fail to recover the correct fault signals on the HIT dataset and exhibit lower kurtosis values. At last, the Mini-blp-lplq, MED, SF-SLSN, and MNNBD show favorable feature extraction capabilities, which maintain fault feature frequencies and recover sparse impulse responses in the time domain. In particular, the MNNBD displays the most analogous characteristics to the raw signal in the time domain and extracts the fourth and fifth harmonic components, which are drowned in the noisy signal as shown in Figure 7 (b). More evidence is shown in Table 5, the signal recovered by the MNNBD has superior performance in terms of kurtosis values. The results suggest that the MNNBD outperforms others in handling noisy conditions.

3.4. Case 3: manually injected fault signals by CWRU

In this case, we conduct experiments using manually injected fault signals. Case Western Reserve University Bearing Data Center (CWRU) collects this bearing vibration dataset. A diameter of 7, 14 and 21 mils single point defects are injected into the outer race, inner race and ball of bearings using an electro-discharge machine. Hence, the dataset consists of ten categories, nine types of faulty bearings and one healthy bearing. Two deep groove ball bearings, 6205-2RS JEM SKF and 6203-2RS JEM SKF are installed in the fan-end (FE) and drive-end (DE) of an electric motor. In particular, four levels (0HP, 1HP, 2HP, 3HP) of load are attached to the motor, which slightly affects the motor speed (1797r/min, 1772/min, 1750r/min, 1730r/min). Vibration signals are measured at two sampling rates: 12kHz and 48kHz. This study uses the vibration signal measured at the DE side with the 12kHz sampling rate.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

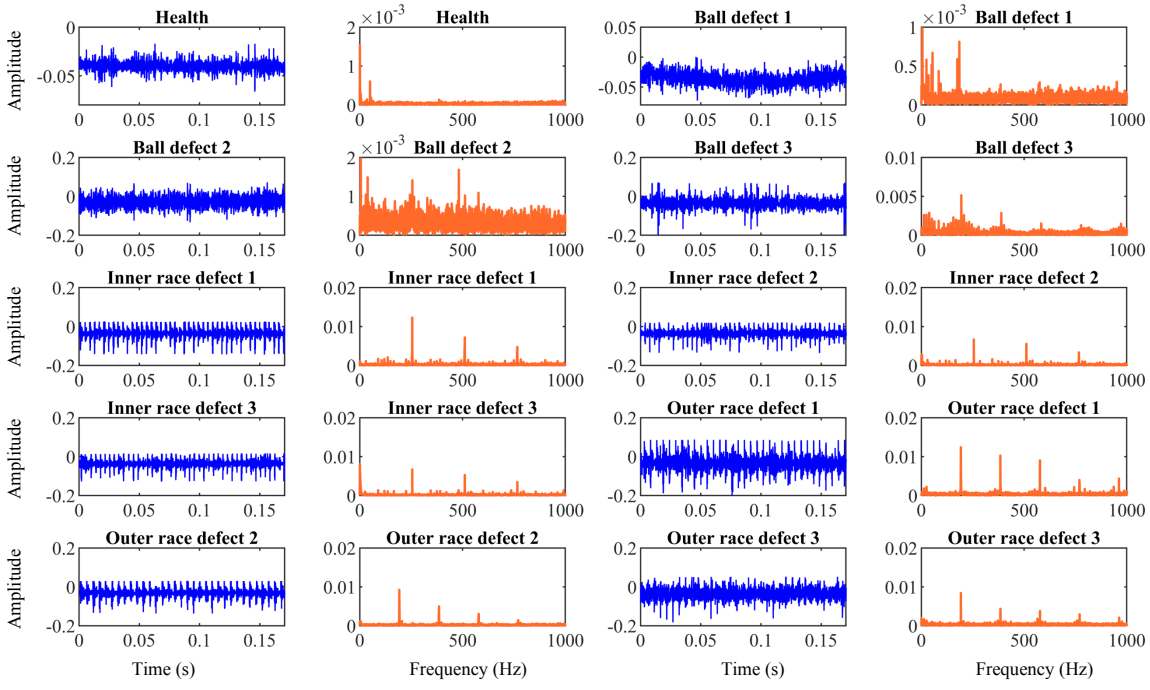


Figure 6: Raw Signal with respect to ten categories of the HIT dataset.

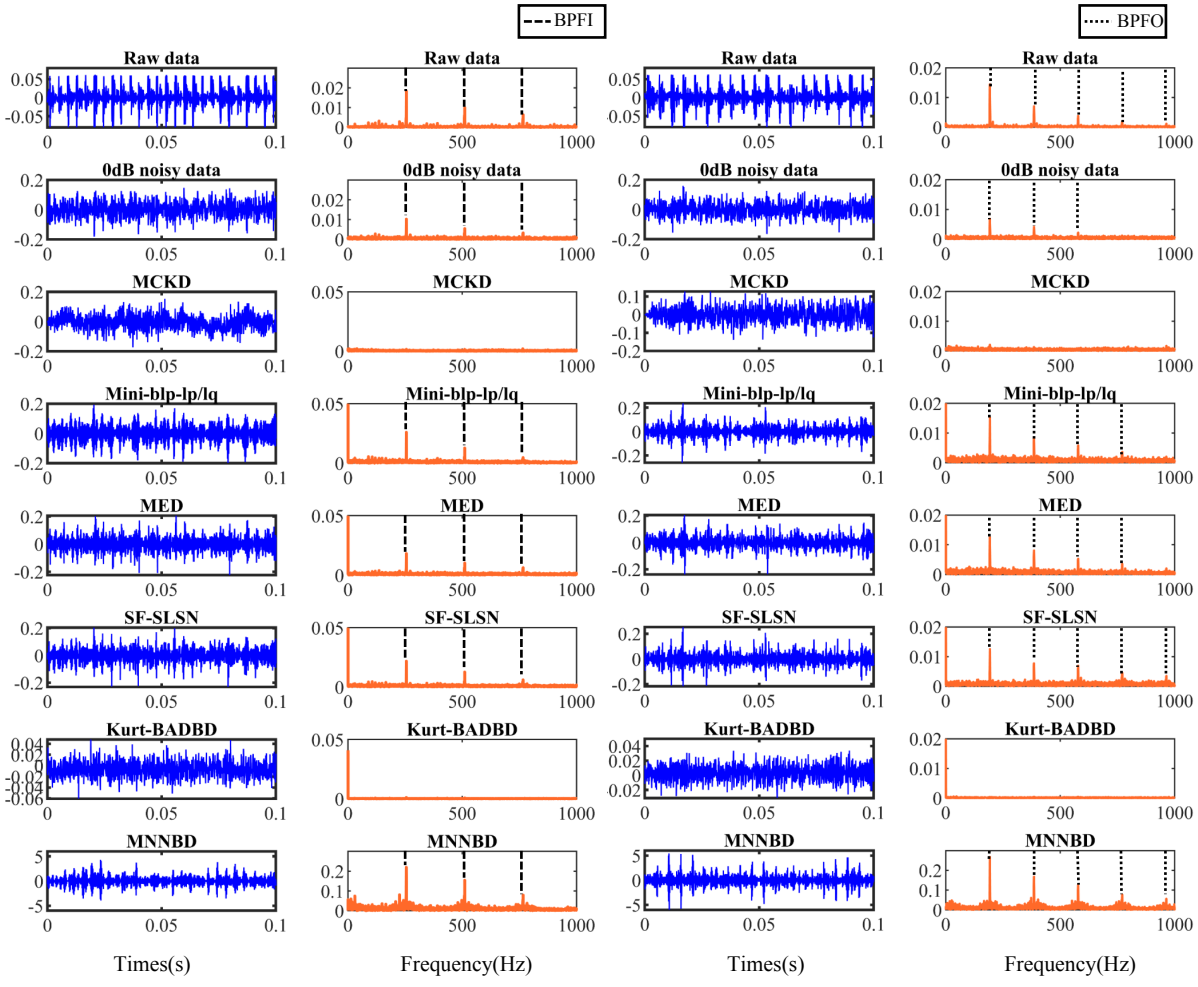
Table 5
The kurtosis of compared mathods on HIT datasets. Ball denotes ball fault, IR denotes inner race fault and OR represents outer race fault.

	Ball1	Ball2	Ball3	IR1	IR2	IR3	OR1	OR2	OR3
Raw Signal	3.0367	3.0452	3.2462	3.2690	3.0985	3.1777	3.2781	3.1066	3.0938
MCKD	3.0884	3.0631	3.0948	3.0326	3.0546	3.0569	3.0872	3.0407	3.0703
Mini-blp-lplq	3.0641	3.0564	4.7936	3.7698	3.3771	7.2924	3.9526	3.7728	3.2118
MED	3.4039	3.3388	5.2767	3.8305	3.3842	7.4536	4.2026	3.9123	3.4517
SF-SLSN	3.2051	3.0277	4.8651	3.8806	3.2974	7.4380	4.2973	4.0891	3.1763
Kurt-BADBD	3.0154	2.9578	3.1323	3.0319	2.9797	3.0275	3.1436	2.9686	2.9841
MNNBD	**3.4258**	**3.4711**	**6.1792**	**4.6555**	**3.4502**	**8.4603**	**5.6116**	**5.1808**	**3.8410**

We use the 2HP load dataset to test the feature exaction performance. For each class of fault signals, 24000 points are picked for blind deconvolution. The results are shown in Table 6. The MNNBD has the highest kurtosis values for both outer race faulty and ball faulty and shows a significant improvement compared to the raw signal. Two other kurtosis-based optimization methods (Kurt-BADBD and MED) and SF-SLSN also yield favorable results. However, MCKD and Mini-blp-lplq have unstable performance and, in some cases, have even fewer kurtosis values than the raw signal. The MCKD requires priori knowledge, but deviations of the fault characteristic frequencies between real scenarios and theoretical values may exist. The Mini-blp-lplq focuses only on the envelope spectrum components, so the time domain-based kurtosis value may be low.

Specifically, in IR1 and IR3, the MED and Kurt-BADBD have the highest kurtosis value. Here, we plot these methods' time domain waveform and envelope spectrum curves. As shown in Figure 8, in IR 1, the MED has the highest kurtosis value, but it extracts the wrong single-peaked impulse. It confirms our view that using kurtosis without constraining may extract a single peak. The same result appears in the Kurt-BADBD on IR3. Although the kurtosis value of the Kurt-BADBD is high, the fault frequency is barely extracted. In contrast, the MNNBD

 doi:10.20944/preprints202211.0



(a) Inner Race Defect (b) Outer Race Defect

Figure 7: Time domain and envelope spectrum of recovered signals by all methods on the HIT dataset. BPFI is the ball pass frequency of the inner race and BPFO is the ball pass frequency of the outer race.

performs consistently well in both cases. The fault frequency and its harmonics are visible and recover distinct periodic impulses in the time domain. As a result, the MNNBD has stable performance in multiple scenarios through the mutual constraints of time and frequency domain optimization.

3.5. Case 4: accelerated degradation fault signals by XJTU-SY

In this case, we conduct experiments using accelerated degradation fault signals. The dataset is provided by the Institute of Design Science and Basic Component at XiâĂŹan Jiaotong University (XJTU) and the Changxing Sumyoung Technology Co., Ltd. (SY) [57]. Three different operating conditions (2100rpm and 12KN, 2250rpm and 11 kN, 2400 rpm and 10 kN) are set in the accelerated degradation experiments; five LDK UER204 bearings are tested under each condition; two accelerometers are installed horizontally and vertically on the housing of the bearings; the sampling frequency is set to 25.6 kHz with 1.28s (32768 data points) each time. The sensors collect data every minute until a major bearing faulty occurs. Here, we choose the Bearing2_1 dataset and employ the BD methods per minute (32768 points) to recover the fault signals. This bearing runs for 8h11min and eventually suffers an inner race fault.

The failure of the bearings will lead to a significant rising in the kurtosis value. We recover the fault signal using BD methods and calculate their kurtosis values. Then, we calculate the absolute value of the gradient of the kurtosis

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

Table 6
The kurtosis of compared mathods on CWRU-2HP datasets. Ball denotes ball fault, IR denotes inner race fault and OR represents outer race fault.

	Ball1	Ball2	Ball3	IR1	IR2	IR3	OR1	OR2	OR3
Raw Signal	2.7877	11.7201	3.2610	5.5783	22.5785	7.8830	7.8161	2.9979	22.3884
MCKD	3.1558	7.8357	2.9963	4.4113	5.8016	5.3381	4.6540	2.9476	9.3193
Mini-blp-lplq	2.8701	7.8224	3.6616	3.9114	14.0924	5.4151	5.1231	2.9692	11.9543
MED	3.8150	50.3977	5.0319	**340.8989**	47.7752	25.0033	**23.4282**	3.6353	59.2419
SF-SLSN	3.4848	29.3699	3.9084	9.7609	33.3053	9.0638	9.5724	3.5410	23.1831
Kurt-BADBD	3.7738	51.5993	4.9035	18.9389	45.6358	**47.4366**	18.3497	3.6444	56.3053
MNNBD	**4.0792**	**54.0683**	**6.0863**	21.9892	**53.2156**	28.2633	20.3630	**3.9908**	**64.3170**

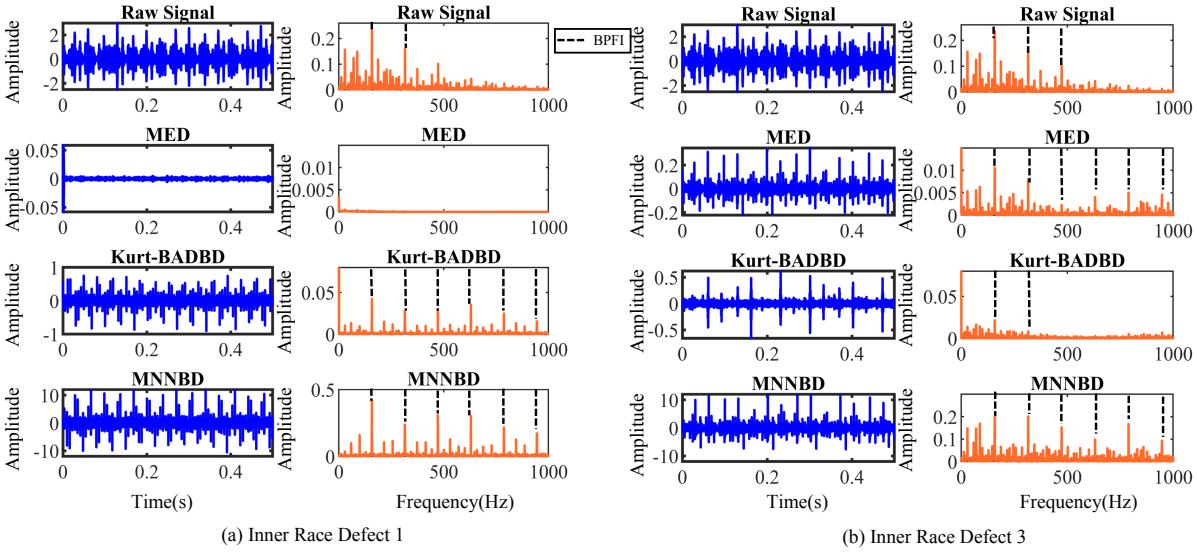


Figure 8: Time domain and envelope spectrum of recovered inner race fault signals by top 3 kurtosis methods on the CWRU dataset. BPFI is the ball pass frequency of the inner race.

Grad(·) to find where the first intense change occurs. The equation is as follows,

$$\Delta K(\boldsymbol{x}) = |\mathrm{Grad}(Kurt(\boldsymbol{x}))|. \tag{19}$$

The occurrence times of the first high ΔK values for the different BD methods are shown in Table 7. The MNNBD and MED detect the faults at 395 minutes, but the other methods detect the faults delay. Comparing the kurtosis of each BD method at 395 minutes, MNNBD has the highest value. In particular, as shown in Figure 9, SF-SLSN shows a large kurtosis at 122 minutes, but it is not a fault feature. It suggests that MNNBD is more sensitive to fault signals and extracts peak impulses when the early fault occurs.

Furthermore, we choose several methods which perform well for analysis. As shown in Figure 10, first, at the early stage, the glitch appears in the kurtosis of the MED and Mini-blp-lplq, and it will interfere with the correct fault identification. Second, we plot the time domain waveforms and the envelope spectrum curves at 395 minutes. The MNNBD and MED exhibit a clear periodic impulse response in the time domain, but others still mix with noise. Then, the MED and MNNBD also display wealthy fault-related frequency components in the frequency domain. However, the Mini-blp-lplq keeps the shaft rotation frequency as a primary component and loses the higher BPFI harmonic components. This phenomenon suggests that optimizing only the envelope spectrum shows more sparse frequency components but may lose the characteristic frequency component in some cases. At last, the results show that the

 doi:10.20944/preprints202211.0

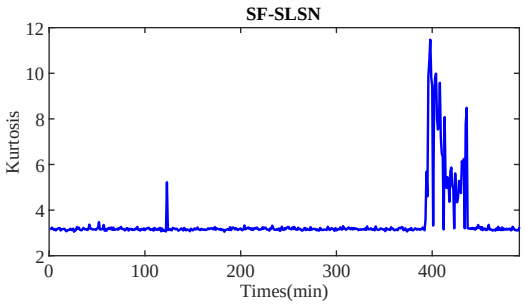


Figure 9: The kurtosis value of SF-SLSN over the full lifetime of the bearing.

Table 7
The earliest moment of the abrupt variation in kurtosis and the corresponding kurtosis value.

	Time(min)	Kurtosis of 395 minutes
Raw Signal	395	5.9092
MCKD	435	3.2660
Mini-blp-lplq	402	4.7729
MED	395	16.1468
SF-SLSN	122	9.8697
Kurt-BADBD	402	4.2718
MNNBD	395	**21.2140**

Table 8
The sensitivity of learning rate in terms of the kurtosis value.

Learning rate	0.1	0.01	0.001	0.0001
Kurtosis	**4.0942**	4.0903	3.8668	3.0514

MNNBD exhibits excellent performance in the time and frequency domains and does not appear the kurtosis fluctuation when dealing with healthy vibration signals. So the MNNBD shows potential for early fault detection.

3.6. Hyperparameters sensitivity

The MNNBD needs to set some initial hyperparameters (learning rate, λ and kernel size (filter length)) before optimizing. In this case, we conduct experiments to test the hyperparameters sensitivity of the MNNBD. We use the simulated signal in case 1 to test how the change of hyperparameters will affects the BD results. First, as shown in Table 8, the higher learning rate shows better performance in kurtosis. However, when the learning rate exceeds 0.01, the learning rate has little effect on the results. Second, from Figure 11, we can find that the kernel size between 60 and 200 shows the appropriate deconvolution results. Because the larger kernel size will decrease the computation speed, we recommend 1×80 kernel size for optimizing the MNNBD. At last, the kurtosis-based methods (MED, Kurt-BADBD) can also recover the desired signal in this simulated signal, the MNNBD shows stable performance without frequency constraint ($\lambda \leq 0.01$). Despite that, we recommend setting λ to 0.01 so that the two criteria are balanced and suitable for a wide range of scenarios.

4. Conclusion

In this article, we have proposed a multi-task convolutional neural network blind deconvolution (MNNBD) for solving the BD problem. The proposed method has integrated the kurtosis and the $G - l_1/l_2$ norm for the time and frequency domains and has utilized a convolutional neural network for optimization. Furthermore, we have analyzed the characteristics of the hybrid optimization criterion. Two sparsity criteria with opposite monotonicity have constrained the optimization procedure in both time and frequency domains. The kurtosis for the time domain and the $G - l_1/l_2$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

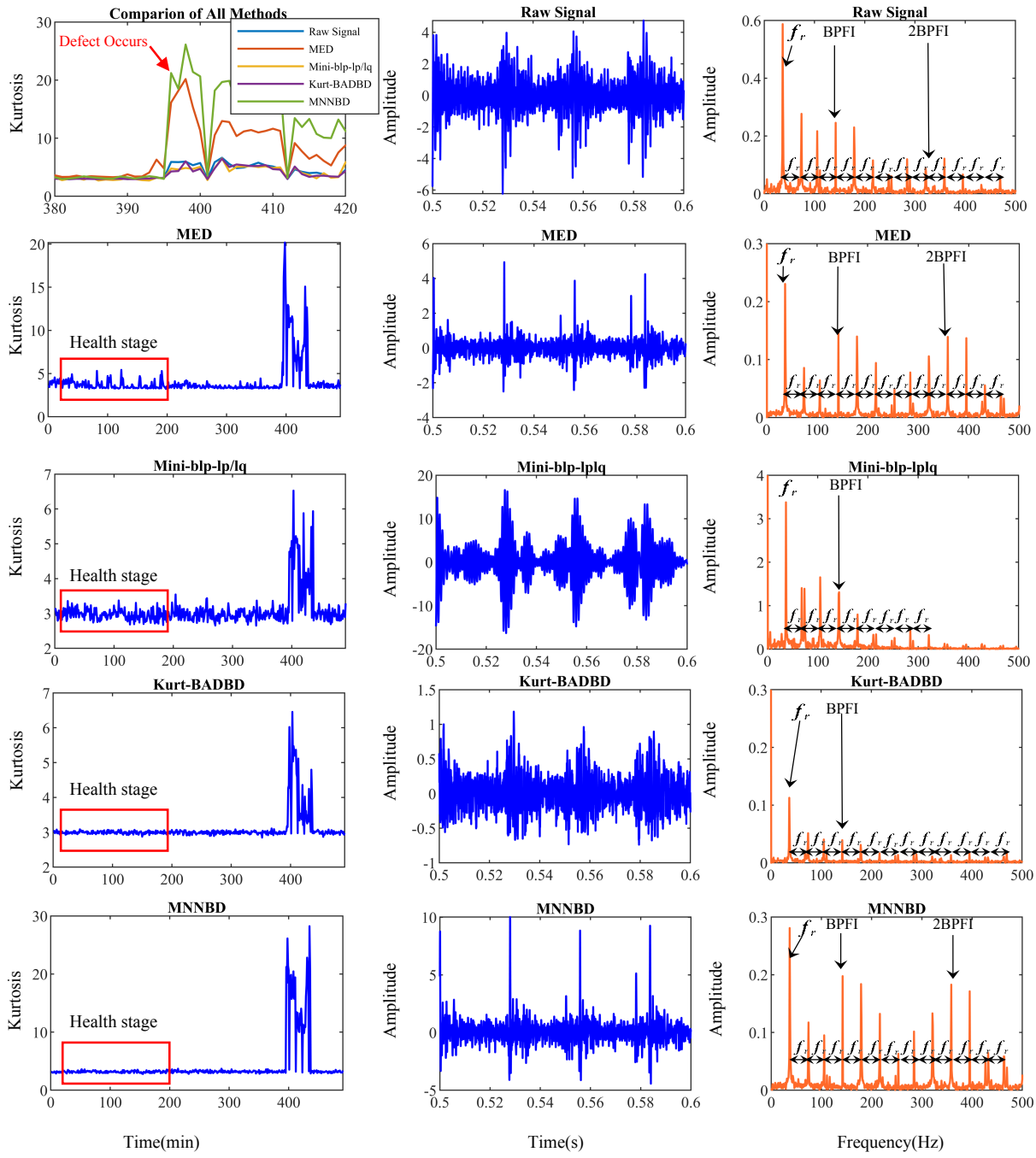


Figure 10: Comparison of the kurtosis variation of several BD methods over the full lifetime of the bearing, together with the time domain waveforms and the envelope spectrum curves at 395 minutes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 November 2022 doi:10.20944/preprints202211.0

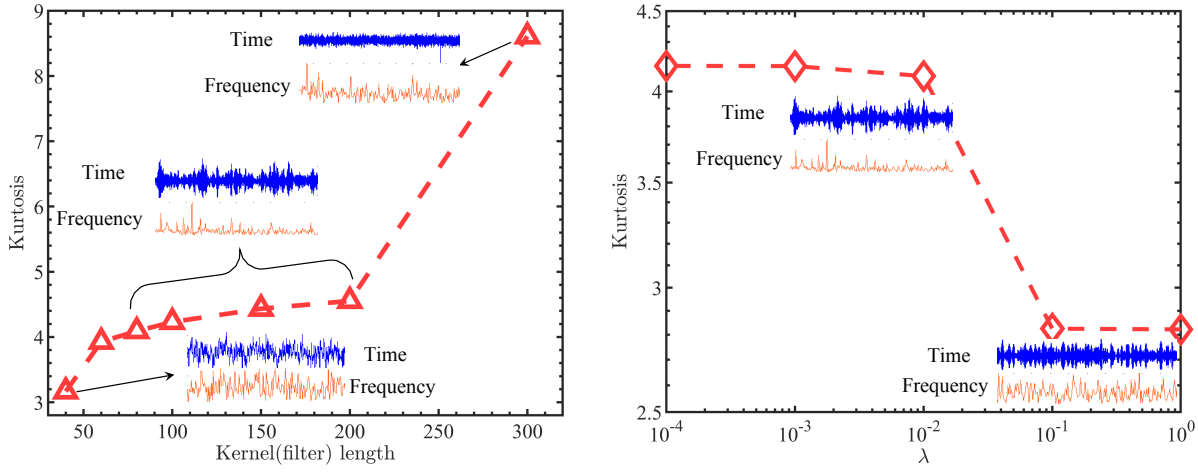


Figure 11: The sensitivity of kernel size (filter length) and scale factor λ in terms of the kurtosis value.

norm for the frequency domain optimization can obtain the ideal results. Finally, the superiority of the MNNBD has been verified in multiple scenarios. Our method has shown a more stable performance compared to others. Since the proposed blind deconvolution method has been embedded into the deep learning framework, it should be widely developed in downstream intelligent fault diagnosis tasks in the future.

CRediT authorship contribution statement

Jing-Xiao Liao: Writing - Original Draft, Methodology, Software, Data curation. **Hang-Cheng Dong:** Writing - Review & Editing, Investigation, Formal analysis, Validation. **Lei Luo:** Writing - Review & Editing, Validation. **Jinwei Sun:** Project administration, Conceptualization, Writing - Review & Editing. **Shiping Zhang:** Funding acquisition, Project administration, Supervision.

Acknowledgement

The bearing faults test of this work is supported by Prof. Xiaoli Zhao, Mr. Hongyuan Zhang in MIIT Key Laboratory of Aerospace Bearing Technology and Equipment, Harbin Institute of Technology.

Declaration of competing interests

The authors declare that they have no known competing ïňĄnan- cial interests or personal relationships that could have appeared to inïňĆuence the work reported in this paper.

External source codes

MED & MCKD: https://www.mathworks.com/matlabcentral/fileexchange/53484-minimum-entropy-deconvolution-multipack-med-meda-omeda-momeda-mckd;
SF-SLSN: https://github.com/aresmiki/SF-SLSN;
Mini-blp-lplq: https://github.com/aresmiki/Mini-blp-lplq;
BADBD: https://github.com/FangBo-0219/BADBD;
We will share our code after this manuscript is published.

References

[1] R. B. Randall, V.-b. C. Monitoring, Industrial, aerospace and automotive applications, VIBRATION-BASED CONDITON MONITORING. West Sussex (2011) 13–20.

[2] Y. Miao, B. Zhang, J. Lin, M. Zhao, H. Liu, Z. Liu, H. Li, A review on the application of blind deconvolution in machinery fault diagnosis, Mech Syst Signal Process 163 (2022) 108202.

[3] A. H. Bonnett, C. Yung, Increased efficiency versus increased reliability, IEEE Industry Applications Magazine 14 (2008) 29–36.

[4] A. Rai, S. H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, Tribology International 96 (2016) 289–306.

[5] X. Lou, K. A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mech Syst Signal Process 18 (2004) 1077–1095.

[6] Z. Peng, W. T. Peter, F. Chu, A comparison study of improved hilbert–huang transform and wavelet transform: Application to fault diagnosis for rolling bearing, Mech Syst Signal Process 19 (2005) 974–988.

[7] P. K. Kankar, S. C. Sharma, S. P. Harsha, Rolling element bearing fault diagnosis using wavelet transform, Neurocomputing 74 (2011) 1638–1645.

[8] Y. Xu, Y. Deng, J. Zhao, W. Tian, C. Ma, A novel rolling bearing fault diagnosis method based on empirical wavelet transform and spectral trend, IEEE Transactions on Instrumentation and Measurement 69 (2019) 2891–2904.

[9] J. Zheng, J. Cheng, Y. Yang, Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis, Mech Syst Signal Process 40 (2013) 136–153.

[10] M. S. Hoseinzadeh, S. E. Khadem, M. S. Sadooghi, Quantitative diagnosis for bearing faults by improving ensemble empirical mode decomposition, ISA transactions 83 (2018) 261–275.

[11] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE transactions on signal processing 62 (2013) 531–544.

[12] Y. Wang, R. Markert, J. Xiang, W. Zheng, Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system, Mech Syst Signal Process 60 (2015) 243–251.

[13] G. W. Stewart, On the early history of the singular value decomposition, SIAM review 35 (1993) 551–566.

[14] H. Li, T. Liu, X. Wu, Q. Chen, A bearing fault diagnosis method based on enhanced singular value decomposition, IEEE Transactions on Industrial Informatics 17 (2020) 3220–3230.

[15] V. Vrabie, P. Granjon, C. Serviere, Spectral kurtosis: from definition to application, in: 6th IEEE international workshop on Nonlinear Signal and Image Processing (NSIP 2003), 2003, p. xx.

[16] J. Antoni, The spectral kurtosis: a useful tool for characterising non-stationary signals, Mech Syst Signal Process 20 (2006) 282–307.

[17] Y. Wang, J. Xiang, R. Markert, M. Liang, Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications, Mech Syst Signal Process 66 (2016) 679–698.

[18] C. A. Cabrelli, Minimum entropy deconvolution and simplicity: A noniterative algorithm, Geophysics 50 (1985) 394–413.

[19] M. Buzzoni, J. Antoni, G. d'Elia, Blind deconvolution based on cyclostationarity maximization and its application to fault identification, Journal of Sound and Vibration 432 (2018) 569–601.

[20] Y. Zhang, X. Li, S. Wang, Y. Sun, Investigation assembly state of spindle bearing based on improved maximum correlated kurtosis deconvolution and support vector machine, Journal of Vibroengineering 20 (2018) 963–978.

[21] S. Wang, J. Xiang, A minimum entropy deconvolution-enhanced convolutional neural networks for fault diagnosis of axial piston pumps, Soft Computing 24 (2020) 2983–2997.

[22] J. X. Liao, H. C. Dong, Z. Q. Sun, J. Sun, S. Zhang, F. Fan, Attention-embedded quadratic network (qttention) for effective and interpretable bearing fault diagnosis, arXiv preprint arXiv:2206.00390 (2022).

[23] F. Fan, W. Cong, G. Wang, A new type of neurons for machine learning, International journal for numerical methods in biomedical engineering 34 (2018) e2920.

[24] J. X. Liao, B. J. Hou, H. C. Dong, H. Zhang, J. Ma, J. Sun, S. Zhang, F. L. Fan, Heterogeneous autoencoder empowered by quadratic neurons, arXiv preprint arXiv:2204.01707 (2022).

[25] R. A. Wiggins, Minimum entropy deconvolution, Geoexploration 16 (1978) 21–35.

[26] K. Pearson, âĂIJdas fehlergesetz und seine verallgemeiner-ungen durch fechner und pearson.âĂİ a rejoinder, Biometrika 4 (1905) 169–212.

[27] B. Zhang, Y. Miao, J. Lin, Y. Yi, Adaptive maximum second-order cyclostationarity blind deconvolution and its application for locomotive bearing fault diagnosis, Mech Syst Signal Process 158 (2021) 107736.

[28] G. L. McDonald, Q. Zhao, M. J. Zuo, Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection, Mech Syst Signal Process 33 (2012) 237–255.

[29] Y. Miao, M. Zhao, J. Lin, X. Xu, Sparse maximum harmonics-to-noise-ratio deconvolution for weak fault signature detection in bearings, Measurement Science and Technology 27 (2016) 105004.

[30] G. L. McDonald, Q. Zhao, Multipoint optimal minimum entropy deconvolution and convolution fix: application to vibration fault detection, Mech Syst Signal Process 82 (2017) 461–477.

[31] B. Chen, W. Zhang, D. Song, Y. Cheng, Blind deconvolution assisted with periodicity detection techniques and its application to bearing fault feature enhancement, Measurement 159 (2020) 107804.

[32] Y. Miao, M. Zhao, J. Lin, Y. Lei, Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings, Mech Syst Signal Process 92 (2017) 173–195.

[33] Y. Cheng, B. Chen, W. Zhang, Adaptive multipoint optimal minimum entropy deconvolution adjusted and application to fault diagnosis of rolling element bearings, IEEE Sensors Journal 19 (2019) 12153–12164.

[34] M. Zhao, J. Lin, Y. Miao, X. Xu, Detection and recovery of fault impulses via improved harmonic product spectrum and its application in defect size estimation of train bearings, Measurement 91 (2016) 421–439.

[35] W. T. Peter, D. Wang, The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as âĂIJtwo automatic vibration-based fault diagnostic methods using the novel sparsity measurement–parts 1 and 2âĂİ, Mech Syst Signal Process 40 (2013) 499–519.

[36] H. Zhang, X. Chen, Z. Du, R. Yan, Kurtosis based weighted sparse model with convex optimization technique for bearing fault diagnosis, Mech Syst Signal Process 80 (2016) 349–376.

[37] L. He, D. Wang, C. Yi, Q. Zhou, J. Lin, Extracting cyclo-stationarity of repetitive transients from envelope spectrum based on prior-unknown blind deconvolution technique, Signal Processing 183 (2021) 107997.
[38] S. Ruder, An overview of multi-task learning in deep neural networks, arXiv preprint arXiv:1706.05098 (2017).
[39] G. Villarrubia, J. F. De Paz, P. Chamoso, F. D. la Prieta, Artificial neural networks used in optimization problems, Neurocomputing 272 (2018) 10–16.
[40] W. Bian, L. Ma, S. Qin, X. Xue, Neural network for nonsmooth pseudoconvex optimization with general convex constraints, Neural Networks 101 (2018) 1–14.
[41] B. Fang, J. Hu, C. Yang, Y. Cao, M. Jia, A blind deconvolution algorithm based on backward automatic differentiation and its application to rolling bearing fault diagnosis, Measurement Science and Technology 33 (2021) 025009.
[42] Y. Cheng, N. Zhou, W. Zhang, Z. Wang, Application of an improved minimum entropy deconvolution method for railway rolling element bearing fault diagnosis, Journal of sound and vibration 425 (2018) 53–69.
[43] Y. Cheng, B. Chen, G. Mei, Z. Wang, W. Zhang, A novel blind deconvolution method and its application to fault identification, Journal of Sound and Vibration 460 (2019) 114900.
[44] B. Fang, J. Hu, C. Yang, X. Chen, Minimum noise amplitude deconvolution and its application in repetitive impact detection, Structural Health Monitoring (2022) 14759217221114527.
[45] W. C. Gray, Variable norm deconvolution, 19, Stanford University Ph. D. thesis, 1979.
[46] Y. Lei, Z. He, Y. Zi, X. Chen, New clustering algorithm-based fault diagnosis using compensation distance evaluation technique, Mech Syst Signal Process 22 (2008) 419–435.
[47] D. Wang, Y. Zhao, C. Yi, K.-L. Tsui, J. Lin, Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings, Mech Syst Signal Process 101 (2018) 292–308.
[48] L. Li, Sparsity-promoted blind deconvolution of ground-penetrating radar (gpr) data, IEEE Geoscience and Remote Sensing Letters 11 (2014) 1330–1334.
[49] X. Jia, M. Zhao, Y. Di, P. Li, J. Lee, Sparse filtering with the generalized lp/lq norm and its applications to the condition monitoring of rotating machinery, Mech Syst Signal Process 102 (2018) 198–213.
[50] S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How does batch normalization help optimization?, Advances in neural information processing systems 31 (2018).
[51] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010, Springer, 2010, pp. 177–186.
[52] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[53] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747 (2016).
[54] L. He, C. Yi, D. Wang, F. Wang, J.-h. Lin, Optimized minimum generalized lp/lq deconvolution for recovering repetitive impacts from a vibration mixture, Measurement 168 (2021) 108329.
[55] R. B. Randall, J. Antoni, Rolling element bearing diagnosticsâĂŤa tutorial, Mech Syst Signal Process 25 (2011) 485–520.
[56] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, Mech Syst Signal Process 64 (2015) 100–131.
[57] B. Wang, Y. Lei, N. Li, N. Li, A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, IEEE Transactions on Reliability 69 (2018) 401–412.