

Article

Laplace Prior-Based Bayesian Compressive Sensing Using K-SVD for Vibration Signal Transmission and Fault Detection

Yunfei Ma ¹, Xisheng Jia ^{1,*}, Qiwei Hu ¹, Daoming Xu ², Chiming Guo ¹, Qiang Wang ¹ and Shuangchuan Wang ¹

¹ Shijiazhuang Campus, Army Engineering University, Shijiazhuang 050003, China; fcz1992@sina.com (Y.M.); h_q_w@sina.com (Q.H.); guochiming@nudt.edu.cn (C.G.); wangqiang1987023@sina.com (Q.W.); wscws0427@163.com (S.W.)

² Radar Sergeant School, Air Force Early Warning Academy, Wuhan 430019, China; alexalex521@163.com

* Correspondence: xs_jia@hotmail.com; Tel.: +86-0311-879-94538

Received: 3 April 2019; Accepted: 1 May 2019; Published: 9 May 2019



Abstract: Vibration signal transmission plays a fundamental role in equipment prognostics and health management. However, long-term condition monitoring requires signal compression before transmission because of the high sampling frequency. In this paper, an efficient Bayesian compressive sensing algorithm is proposed. The contribution is explicitly decomposed into two components: a multitask scenario and a Laplace prior-based hierarchical model. This combination makes full use of the sparse promotion under Laplace priors and the correlation between sparse blocks to improve the efficiency. Moreover, a K-singular value decomposition (K-SVD) dictionary learning method is used to find the best sparse representation of the signal. Simulation results show that the Laplace prior-based reconstruction performs better than typical algorithms. The comparison between a fixed dictionary and learning dictionary also illustrates the advantage of the K-SVD method. Finally, a fault detection case of a reconstructed signal is analyzed. The effectiveness of the proposed method is validated by simulation and experimental tests.

Keywords: vibration; signal; Bayesian compressive sensing; K-SVD; gearbox; bearing

1. Introduction

High-speed vibration signal transmission plays an important role in capturing equipment failure, which is of interest in many applications. Compared with wired monitoring processes, wireless systems greatly increase flexibility, maintainability, and scalability [1]. However, due to the limitations of wireless transmission bandwidth, real-time monitoring must compress the signal for transmission. In previous studies, vibration signal compression was applied for structural health monitoring (SHM) [2,3]. However, the sampling frequency in SHM is much lower than that in mechanical monitoring. For example, only a 240-Hz sampling frequency [4] is needed in bridge structure monitoring, while a sampling frequency of at least 5–20 kHz [5] is required to realize mechanical monitoring, which causes great difficulties in vibration compression. At present, methods based on the transform domain dominate vibration signal compression. Among them, wavelet transform [6], arithmetic coding [7], and Huffman coding [8] are widely used. However, some deficiencies remain in the research on vibration signal compression.

Compressive sensing (CS) theory [9–11], which is based on the observation that a small collection of a sparse signal projections may contain sufficient information, emerged in recent years. CS can be regarded as a breakthrough of the Nyquist sampling theorem that requires fewer measurements. A number of algorithms for recovering originally sparse signals were proposed. These algorithms can

be classified into three categories: convex optimization algorithms [12], greedy-based algorithms [13], and Bayesian algorithms [14].

Bayesian compressive sensing (BCS) combines Bayesian estimation with CS to obtain the maximum posterior probability of the original signal rather than performing point estimation. One advantage of BCS is that the noise generated in signal transmission is taken into consideration, and the original signal is not strictly required to be sparse. Furthermore, BCS may be extended to multitask CS (MCS) [15], which performs multiple sets of CS measurements jointly. MCS theory is based on the observation that measurements of different tasks are statistically related when multitask reconstructions are performed under the same scenario. Moreover, Zhang [16] et al. exploited the situation in which elements in the nonzero row of the matrix are temporally correlated, and proposed two sparse Bayesian learning algorithms. Babacan [17] et al. first used the Laplace prior-based BCS algorithm and obtained better sparseness for reconstruction. Another advantage of the Laplace prior is the log-concavity, which eliminates local minima. BCS is widely used in the fields of image processing [18], electrocardiograph (ECG) signal reconstruction [19], and radar signal estimation [20], but BCS is rarely used for mechanical vibration signal reconstruction.

The premise of CS is that a signal must be sparse or sparse in a certain domain. An over-complete dictionary that decomposes the original signal efficiently is required. Dictionaries can be divided into two categories: fixed dictionaries and learning dictionaries. Fixed dictionaries, such as discrete cosine transform, Fourier transform, and wavelet transform, have strong dependence on prior knowledge of the original signal. Learning dictionaries are updated adaptively according to the original signal to achieve increasing attention. Typical learning dictionaries, such as the method of optimal directions (MOD) [21] and K-singular value decomposition (K-SVD) [22], produce good results in application. The K-SVD algorithm was successfully applied in the fields of image denoising and CS. Zhou et al. [23] replaced the orthogonal basis function with a K-SVD-trained over-complete dictionary. Yang et al. [24] improved the K-SVD method by means of the correlation coefficient matching criterion and dictionary cutting. Shi et al. [25] combined the K-SVD algorithm with the idea that high- and low-resolution dictionaries can be cogenerated. In addition to K-SVD, Jafari et al. [26] found a link between sparsity in the dictionary and sparsity in decomposition. They proposed a greedy adaptive learning algorithm for finding sparse atoms. Ophir [27] et al. presented a multiscale dictionary learning method for different applications. The algorithm can not only reduce the training time but also improve the reconstruction quality.

However, vibration signal transmission is not the main purpose of machinery condition monitoring. In the past few years, considerable attention was paid to the combination of CS and fault detection. For example, Wang et al. [28] proposed a proximal decomposition algorithm for reconstruction of sparse time–frequency (TF) representation. The experiments on bearings and gears show that the proposed method can retain TF features through small measurements. Tang et al. [29] described a CS framework of characteristic harmonics for detecting bearing faults. In their framework, the processes of sampling and fault detection are performed simultaneously. Sun et al. [30] introduced the block sparse Bayesian learning method for CS reconstruction. The Bayesian algorithm works well by exploiting the block property and inner structures of the original signal. Experiments illustrate that the Bayesian method is suitable for signal reconstruction and fault detection.

In this paper, we introduce a Laplace prior into the hierarchical MCS model in combination with K-SVD dictionary learning. The contribution of this work is twofold. Firstly, we develop a new technique, named Laplace prior-based correlated-sparse-block Bayesian CS (Lap-CBCS), which imposes sparseness over the original signal and extends the Laplace prior-based BCS algorithm to the multitask scenario. Secondly, we use the K-SVD dictionary for sparse decomposition. For a given complex signal, such a dictionary can be trained for sparse promotion. Compared to the fixed transform, K-SVD offers improved decomposition and good signal reconstruction performance. The proposed method is referred to as Lap-CBCS-KSVD. Finally, an application of Lap-CBCS-KSVD for fault classification is presented using planetary gearbox data. The classification results of the support vector machine

and random forest (RF) methods demonstrate the effectiveness of Lap-CBCS-KSVD. The new method provides technical support for wireless monitoring of mechanical equipment.

The structure of this paper can be summarized as follows: in Section 2, we review CS theory and K-SVD dictionary optimization. In Section 3, the Laplace prior-based correlated-sparse-block BCS algorithm is presented. In Section 4, the framework of the Lap-CBCS-KSVD method is proposed. Section 5 compares the simulation results of the proposed method with those of typical algorithms. Section 6 presents the fault classification results using the reconstructed signal. Section 7 summarizes all the content in this paper.

2. Related Work: Compressive Sensing and KSVD

2.1. Compressive Sensing

CS [11] uses a low-dimensional signal to approximate the original signal. Denoting Ψ as the sparse transform, the original signal $x(x \in R^N)$ can be represented as

$$x = \Psi\theta \tag{1}$$

where θ is the coefficient vector in the Ψ -domain. If θ_0 represents the $N-M$ smallest coefficients of θ set 0, then we have that $\|\theta_0 - \theta\|_2 / \|\theta\|_2$ is negligibly small when $M \ll N$. Based on this observation, the CS measurements may be represented as

$$y = \Phi x = \Phi\Psi\theta = \Theta\theta \tag{2}$$

Solving a sparse vector θ with respect to Θ is a commonly discussed problem [11,12]. However, BCS [14] solves the maximum posteriori estimate of the original signal from a probabilistic perspective.

In the field of BCS, a Gaussian prior-based model is widely used. However, the Laplace distribution, conjugated by Gaussian and exponential distributions, recently emerged. The results of Reference [17] show that the Laplace prior-based model has better sparseness promotion than the Gaussian prior-based model, while also being log-concave. Because the Laplace prior is not conjugate to the Gaussian prior, a hierarchical model is adopted here using the relevance vector machine (RVM) [31]. In this paper, we introduce this hierarchical model to the multitask scenario in Section 3.

2.2. K-SVD

To obtain a stronger sparse representation, this paper replaces the traditional transform bases with an over-complete dictionary. The basic idea is to use the K-SVD algorithm proposed by Aharon et al. [22] to train various signal blocks and adaptively update the dictionary atoms until an over-complete dictionary is obtained. Extended by K-means, the K-SVD algorithm effectively reduces the number of atoms in the dictionary and confirms that the remaining atoms still represent all the information. Compared to the fixed sparse dictionary, K-SVD efficiently avoids the strong dependence on prior knowledge and poor adaptability. Given training matrix Y and sparse matrix G , the process of dictionary D learning is described as

$$\min_D \|Y - DG\|_F^2 \text{ s.t. } \|g_i\|_0 \leq T \tag{3}$$

Let E_i represent the error after removing the i -th atom, d_j represent the j -th column in dictionary D , and g_i indicate the i -th row in sparse matrix G . We have the following equation:

$$\|Y - DG\|_F^2 = \left\| Y - \sum_{j=1}^K d_j g^j \right\|_F^2 = \left\| (Y - \sum_{j \neq i} d_j g^j) - d_i g^i \right\|_F^2 = \|E_i - d_i g^i\|_F^2 \tag{4}$$

Since the elements in \mathbf{g}^i may be 0, the atoms must shrink during training. We define ω_i as

$$\omega_i = \{k | \mathbf{g}^i(k) \neq 0\} \tag{5}$$

which represents the nonzero index set of \mathbf{g}^i . Another matrix $\mathbf{\Omega}_i$, which places 1 at the position $(\omega_i(i), i)$, is also introduced; thus, we have $E_R^i = E_i \mathbf{\Omega}_i$ and $\mathbf{g}_R^i = \mathbf{g}_i \mathbf{\Omega}_i$ for zero shrinking. Then, E_R^i can be decomposed as

$$E_R^i = \mathbf{U} \mathbf{\Delta} \mathbf{V}^T \tag{6}$$

Atom \mathbf{d}_i in the dictionary is replaced by the first column of matrix \mathbf{U} . The coefficient \mathbf{g}_R^i is updated by the product of the first column of matrix \mathbf{V} and $\mathbf{\Delta}(1,1)$. At this point, the first column is updated, and the remaining columns can be updated in the same manner to generate a new dictionary.

3. Laplace Prior-Based Correlated-Sparse-Block BCS Method

In this section, we propose a new algorithm called Lap-CBCS. A hierarchical model that combines Laplace priors and multitask BCS is established; then, the signal is reconstructed by estimating the hyperparameters shared by all signal blocks in the model. The improved algorithm makes full use of the sparseness promotion with Laplace priors and the correlation between signal blocks to improve the accuracy of the reconstruction.

3.1. Distribution of the Lap-CBCS Hierarchical Model

Consider an N -dimensional signal \mathbf{x} . Let $\mathbf{\Phi} \in R^{M \times N}$ represent the measurement matrix, and let $\mathbf{\Psi}$ be an $N \times N$ transform basis such that $\mathbf{x} = \mathbf{\Psi} \boldsymbol{\theta}$. Assume that L tasks of CS measurements are performed, with L observations statistically related as defined below.

$$\mathbf{y}_i = \mathbf{\Phi}_i \mathbf{x}_i + \mathbf{n}_i = \mathbf{\Phi}_i \mathbf{\Psi}_i \boldsymbol{\theta}_i + \mathbf{n}_i = \mathbf{\Theta}_i \boldsymbol{\theta}_i + \mathbf{n}_i \quad (i = 1, \dots, L). \tag{7}$$

In Equation (7), \mathbf{n}_i represents a noisy vector, described as a zero-mean Gaussian random variable with unknown precision α_0 (variance $1/\alpha_0$). The Gaussian likelihood function of Equation (7) can be expressed as

$$P(\mathbf{y}_i | \boldsymbol{\theta}_i, \alpha_0) = (2\pi/\alpha_0)^{-M/2} \exp(-\frac{\alpha_0}{2} \|\mathbf{y}_i - \mathbf{\Theta}_i \boldsymbol{\theta}_i\|_2^2) \tag{8}$$

The likelihood function of the original signal $\boldsymbol{\theta}_i$ is assigned Laplace priors, as shown in Equation (9). Because Laplace priors are not conjugate to the Gaussian distribution, we introduce a hierarchical prior to solve this problem.

$$p(\boldsymbol{\theta}_i | \lambda) = \frac{\lambda}{2} \exp(-\frac{\lambda}{2} \|\boldsymbol{\theta}_i\|_1) \tag{9}$$

In Equation (10), γ is a set of hyperparameters determining the prior distribution of $\boldsymbol{\theta}_i$. To apply Laplace priors to a Bayesian model, hyperparameter λ, v is introduced for γ_i , as shown in Equations (11) and (12). By combining Equations (10) and (11), we obtain extended Equation (13).

$$p(\boldsymbol{\theta}_i | \gamma) = \prod_{j=1}^N N(\boldsymbol{\theta}_{i,j} | 0, \gamma_j^{-1}) \tag{10}$$

$$p(\gamma_j | \lambda) = Ga(\gamma_j | 1, \lambda/2) = \frac{\lambda}{2} \exp(-\frac{\lambda \gamma_j}{2}) \tag{11}$$

$$p(\lambda | v) = Ga(\lambda | v/2, v/2) \tag{12}$$

$$p(\boldsymbol{\theta}_i | \lambda) = \prod_i \int p(\boldsymbol{\theta}_i | \gamma_i) p(\gamma_i | \lambda) d\gamma_i = \frac{\lambda^{N/2}}{2^N} \exp(-\lambda^{1/2} \sum_i |\boldsymbol{\theta}_i|) \tag{13}$$

In this manner, the original signal θ_i can be estimated with hyperparameters γ , λ , and v . Moreover, the noisy precise α_0 can be represented by hyperparameters a and b . Figure 1 shows the hierarchical a priori model of MCS using Laplace priors.

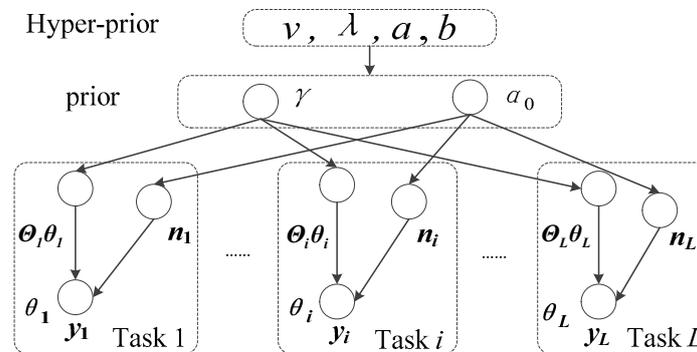


Figure 1. A hierarchical model of the Laplace prior-based correlated-sparse-block Bayesian compressive sensing (Lap-CBCS) approach.

3.2. Bayesian Estimation for Hyperparameters

The greatest difference between Lap-CBCS and multitask BCS is the embedded layer of hyperparameter λ . This section shows how to estimate the hyperparameters and θ_i . Given γ , λ , and v , the posterior of θ_i with known measurement y_i can be expressed as

$$\prod_{i=1}^L p(\theta_i, \gamma, \lambda, v | y_i) = \prod_{i=1}^L p(\theta_i | y_i, \gamma, \lambda, v) p(\gamma, \lambda, v | y_i) \tag{14}$$

Since $p(\theta_i | y_i, \gamma, \lambda, v) \propto p(\theta_i, y_i, \gamma, \lambda, v)$, $p(\theta_i | y_i, \gamma, \lambda, v)$ is a multivariate Gaussian distribution with mean μ_i and covariance Σ_i .

$$\mu_i = \sum_i \Theta_i^T y_i, \Sigma_i = (\Theta_i^T \Theta_i + \Lambda_{n-\lambda})^{-1} \tag{15}$$

where $\Lambda_{n-\lambda} = \text{diag}(1/\gamma_1, 1/\gamma_2, 1/\gamma_3, \dots, 1/\gamma_N) = \text{diag}\{1/\gamma_j\}$. We attempt to utilize $p(\gamma, \lambda, \alpha_0 | y_i)$ to estimate the Bayesian hyperparameters $\{\gamma, \lambda, \alpha_0\}$. Since $p(\gamma, \lambda, \alpha_0 | y_i) = p(\gamma, \lambda, \alpha_0, y_i) / p(y_i) \propto p(\gamma, \lambda, \alpha_0, y_i)$, we need only maximize the term $p(\gamma, \lambda, \alpha_0, y_i)$, which is computable and given by

$$p(\gamma, \lambda, \alpha_0, y_i) = \int p(y_i | \theta_i, \alpha_0) p(\theta_i | \gamma) p(\gamma | \lambda) p(\lambda | v) p(\alpha_0) d\theta_i \tag{16}$$

Furthermore, applying the logarithm yields the posterior distribution function and omits the constant term in operation. We decompose constant C as in Reference [23] and apply the matrix inversion lemma. Finally, the complete presentation of $L_{Laplace}^{multi-task}$ is obtained.

$$L_{Laplace}^{multi-task} = \sum_{i=1}^L \text{Iog} p(\gamma, \lambda, \alpha_0, y_i) = \sum_{i=1}^L \text{Iog} \int p(y_i | \theta_i, \alpha_0) p(\theta_i | \gamma) p(\gamma | \lambda) p(\lambda | v) p(\alpha_0) d\theta_i \tag{17}$$

Then, we differentiate $L_{Laplace}^{multi-task}$ with respect to γ and λ and set the result to 0. This process can be implemented readily via simplification to yield

$$\gamma_j^{new} = \frac{-L + \sqrt{L^2 + 4\lambda L \sum_{i=1}^L (\mu_{i,j}^2 + \Sigma_{i,jj})}}{2\lambda L}, \lambda^{new} = \frac{2(N-1)}{\sum_j r_j} \tag{18}$$

where $\mu_{i,j}$ represents the estimated mean value of the j -th hyperparameter γ_j for signal i and $\Sigma_{i,jj}$ is the j -th diagonal component of the variance matrix of signal i . Note that γ_j^{new} is a function of $\mu_{i,j}$ and $\Sigma_{i,jj}$, while $\mu_{i,j}$ and $\Sigma_{i,jj}$ are functions of γ_j^{new} : this situation suggests an iterative solution using Equations (15) and (18). Compared to the Gaussian hyperparameter α_i^{new} in Reference [14], we have $\gamma_j^{new} > \alpha_i^{new}$, which indicates that Laplace priors can better promote sparseness.

4. Framework of the Lap-CBCS-KSVD Method

A flowchart of the Lap-CBCS-KSVD method is shown in Figure 2. The model includes three parts: dictionary off-line training, data compression, and signal reconstruction. The measurement matrix and block length are set before the acquisition node is activated. When a signal is collected, the signal must be split because the collected data are huge and the entire data cannot be used as training samples. Each signal block is decomposed using dictionary Ψ after training, and the original signal is compressed with a measurement matrix Φ . The compressed signal is transmitted through the wireless sensor network, and the reconstruction stage is completed at the manage node. In this part, the Laplace prior-based hierarchical model is introduced for Bayesian estimation. Based on the observation of received signal y_i ($i = 1, \dots, L$), the sparse coefficients θ_i ($i = 1, \dots, L$) are estimated with respect to matrix $\Theta = \Phi\Psi$. Finally, the original signal blocks x_i ($i = 1, \dots, L$) are reconstructed through inverse transformation. The whole algorithm works as follows:

1. Original signal x is split into L blocks according to the block length set. During the off-line stage, these blocks are used as samples for K-SVD dictionary training.
2. During signal acquisition, the signal blocks are decomposed and compressed using K-SVD dictionary Ψ and measurement matrix Φ .
3. The compressed signal blocks are transmitted to the upper node via a wireless sensor network.
4. The Laplace prior-based hierarchical model is established, and Bayesian estimation is conducted for sparse coefficients θ_i .
5. The inverse transformation is applied to obtain signal blocks and reconstructed signal x' .

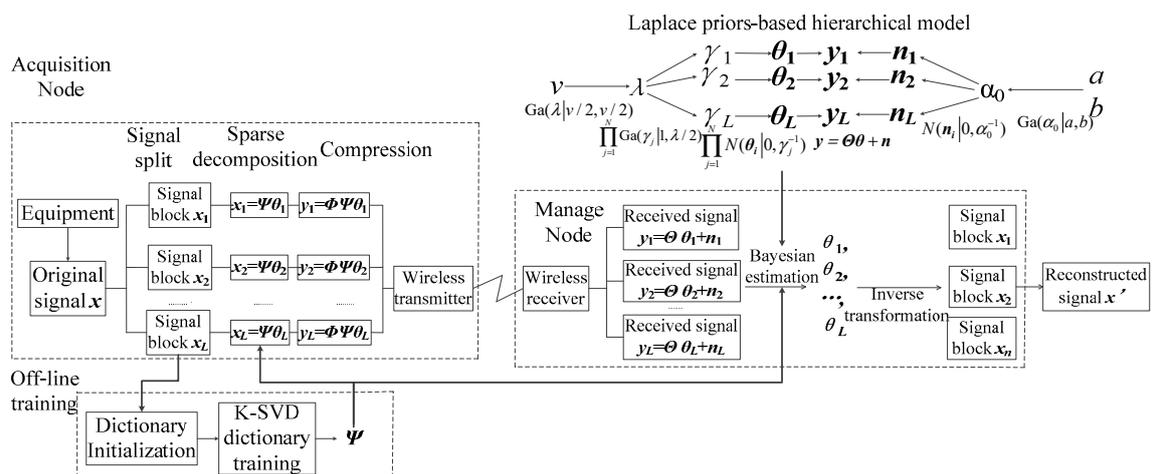


Figure 2. Flowchart of the method proposed in this paper.

5. Simulation

In this section, we examine the performance of our proposed Lap-CBCS-KSVD algorithm for real accelerometer data from a bearing and a gearbox. Case 1 comes from the ball bearing fault test bed [32] belonging to the electric engineering laboratory of Case Western Reserve University. The test bench consists of a 2-hp motor, a torque decoder, and a power tester. The tested bearing supports the rotor of a motor. The drive-end bearing is an SKF6205, and the fan-end bearing is an SKF6203.

The bearing faults can be divided into three types according to the fault location: inner raceway fault, outer raceway fault, and ball fault. We select two datasets: IR007_0_105 and B028_0_3005.

Case 2 involves raw data collected from a gearbox test bed. The failure of a gearbox will substantially affect the stable operation of mechanical equipment. Thus, gearbox monitoring is of great significance in real scenarios. The experimental gearbox is a JZQ175. The electromagnetic speed-regulating motor provides 4 kW of power, and the air-cooled magnetic powder brake provides load for the gearbox. The data acquisition system consists of four 3056B4 piezoelectric sensors produced by Dytran Instruments, Inc. During the experiment, the sampling frequency was set to 20 kHz.

We conducted two types of preset fault experiments for the gearbox: crack failure experiments and broken-tooth failure experiments. Four failure states were considered: 5-mm crack fault, 5-mm broken-tooth fault, 8-mm crack fault, and 10-mm broken-tooth fault. The selected datasets are shown in Table 1, and the signal after sampling is shown in Figure 3. In this paper, the compression ratio is denoted as

$$CR = \frac{N - M}{N} \tag{19}$$

Table 1. Experimental datasets.

Dataset	Fault location	Speed	Load	Equipment	Source
IR007_0_105	Inner raceway	1797 r/s	0 Nm	Bearing	Public
B028_0_3005	Ball	1797 r/s	0 Nm	Bearing	Public
Crack5mm	Tooth foot	800 r/s	10 Nm	Gearbox	Measured
Break5mm	Gear	800 r/s	20 Nm	Gearbox	Measured
Crack8mm	Tooth foot	800 r/s	20 Nm	Gearbox	Measured
Break10mm	Gear	800 r/s	15 Nm	Gearbox	Measured

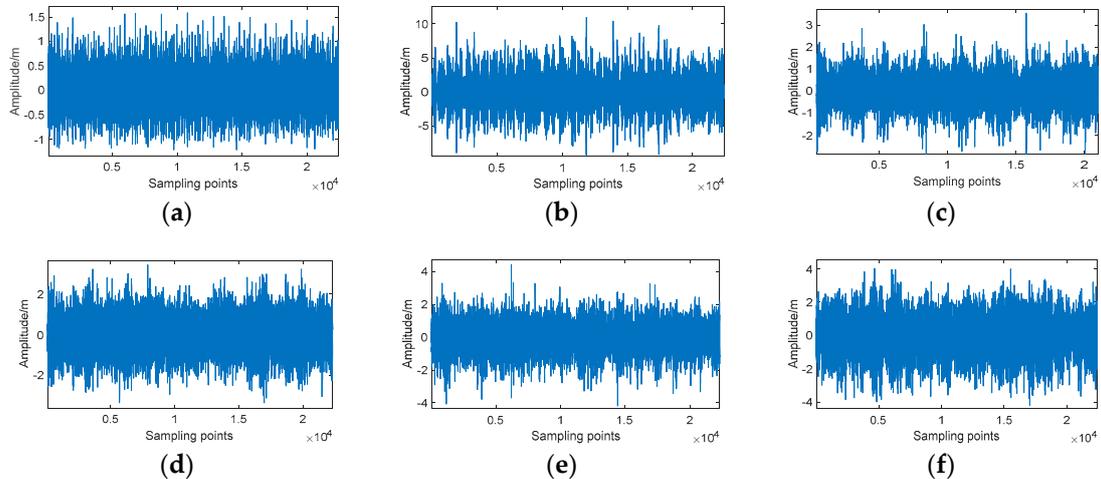


Figure 3. Experimental data collected from bearings, (a) IR007_0_105 and (b) B028_0_3005, and gearbox (c) Crack5mm; (d) Break5mm; (e) Crack8mm; (f) Break10mm.

The mean square error (*MSE*), peak signal-to-noise ratio (*PSNR*), and Pearson correlation coefficient (*r*) were used as evaluation metrics.

$$MSE = \frac{\|u' - u\|_2}{\|u\|_2} \tag{20}$$

$$PSNR = 10 \lg(u_{max}^2 / (\frac{1}{N} \sum_{i=1}^N (u_i - u'_i)^2)) \tag{21}$$

$$r_{u,u'} = \frac{N \sum uu' - \sum u \sum u'}{\sqrt{N \sum u^2 - (\sum u)^2} \sqrt{N \sum (u')^2 - (\sum u')^2}} \tag{22}$$

where \mathbf{u} represents the original signal, \mathbf{u}' represents the reconstructed signal, and \mathbf{u}_{max} represents the largest component of vector \mathbf{u} .

5.1. Comparison with Other Reconstruction Algorithms

In this part, several widely used CS algorithms are compared with Lap-CBCS. Figure 4 presents the reconstruction results of basic pursuit (BP) [12], orthogonal matching pursuit (OMP) [13], BCS [14], and regular orthogonal matching pursuit (ROMP) [33].

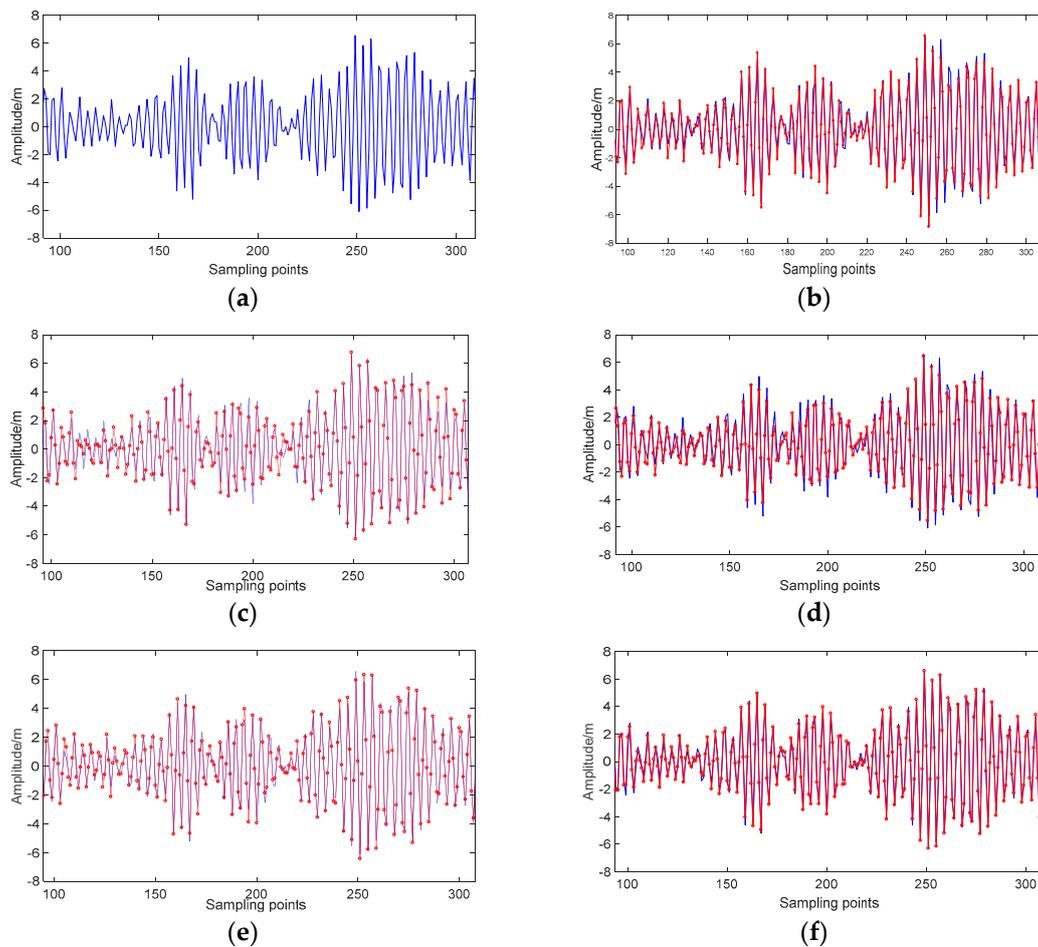


Figure 4. (a) Original bearing signal and reconstructed signal of B028_0_3005 using (b) BP ($MSE = 0.2839$); (c) OMP ($MSE = 0.3175$); (d) ROMP ($MSE = 0.3315$); (e) BCS ($MSE = 0.2741$); and (f) Lap-CBCS ($MSE = 0.2358$) algorithms.

The B028_0_3005 bearing data were chosen as an example to illustrate the reconstruction performance. The blue lines represent raw signals, and the red lines with spikes represent reconstructed signals. To guarantee the principle of a single variable, the discrete cosine transform (DCT) transform was selected as the sparse basis. In contrast to Lap-CBCS, the BCS algorithm uses the traditional Gaussian prior model for reconstruction. The BP and OMP algorithms are typical representatives of convex optimization and greedy iteration algorithms. The results in Figure 4 show that Lap-CBCS recovered the original signal better than the other four algorithms; the MSE results confirmed this conclusion.

Then, we investigated the effectiveness of different compression ratios on the reconstruction performance of different algorithms. As shown in Figure 5, we varied the compression ratio (CR) from 0 to 1. A smaller MSE indicates more accurate reconstruction. For each data point, 100 groups of experiments were conducted to calculate the average ρ and variance τ . Figure 5 gives the range of $[\rho - 2\tau, \rho + 2\tau]$ for each point, which can be regarded as the 95% confidence interval. On the basis of all

the results with different CRs, it can be seen that the MSE of Lap-CBCS was smaller than that of BP, BCS, and OMP, which confirms the effectiveness of the proposed algorithm. Moreover, the variance of Lap-CBCS was the smallest. Figure 5 shows that the variance of each point increased with increasing CR, which indicates the instability of reconstruction algorithms under high CRs.

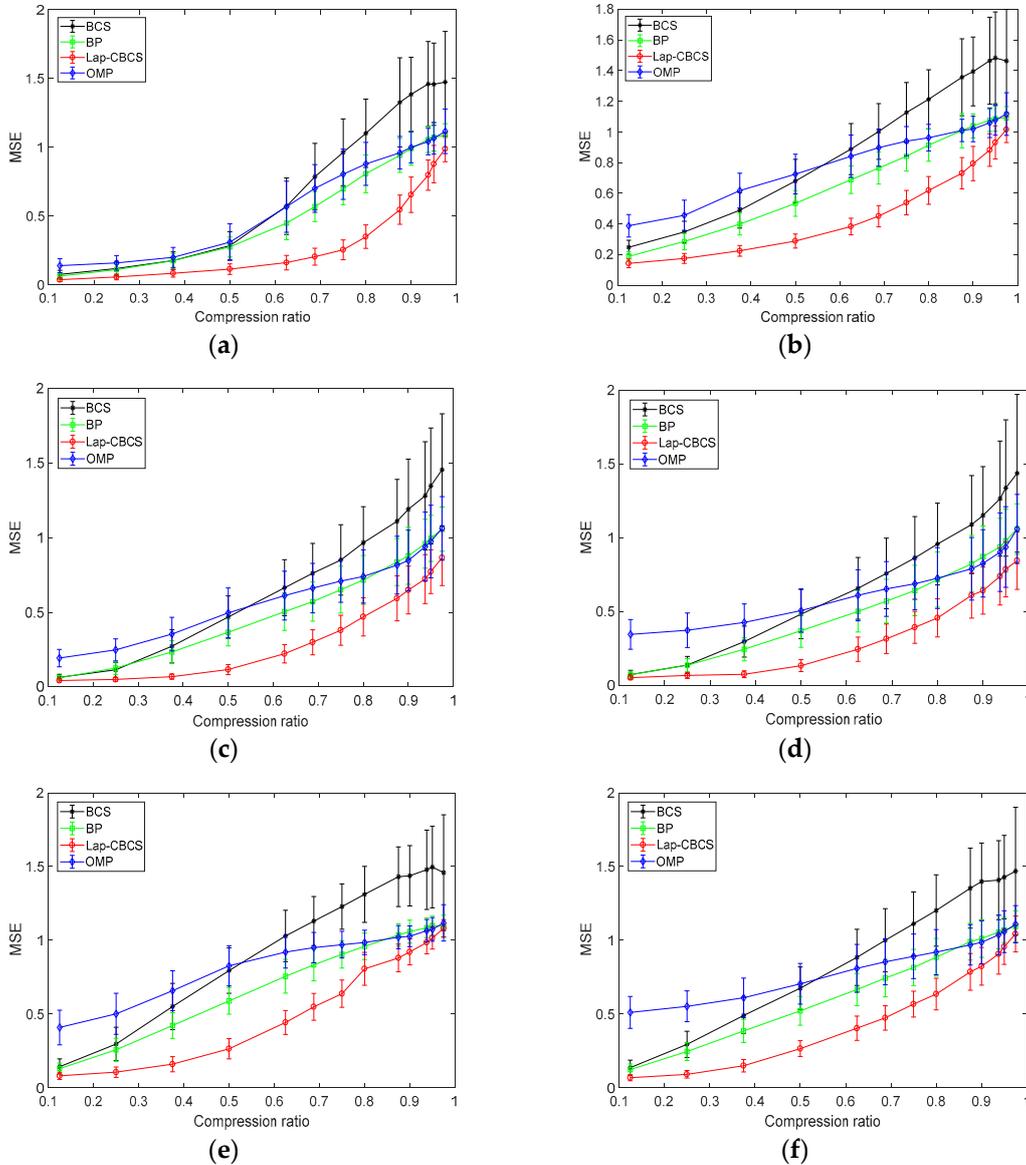


Figure 5. Comparison of the MSE 95% confidence intervals of Lap-CBCS and other reconstruction algorithms on six datasets: (a) B028_0_3005; (b) IR007_0_105; (c) Break5mm; (d) Break10mm; (e) Crack5mm; (f) Crack8mm.

5.2. Robustness and Cost Analysis of the Reconstruction Algorithms

More experiments are needed to compare the robustness of OMP, BP, BCS, and Lap-CBCS. The datasets IR007_0_105 and break10mm_800r_15nm are used in this section. One hundred experiments were conducted to calculate the average MSE for each signal-to-noise ratio (SNR). The SNR is defined in Equation (23). Figure 6 shows that the four algorithms were not substantially different when the SNR was small; however, the PSNR of Lap-CBCS increased significantly when the SNR was large, indicating that Lap-CBCS is superior to the other three algorithms.

$$SNR(dB) = 20 \lg(\|\Phi x\|_2 / \|y\|_2) \tag{23}$$

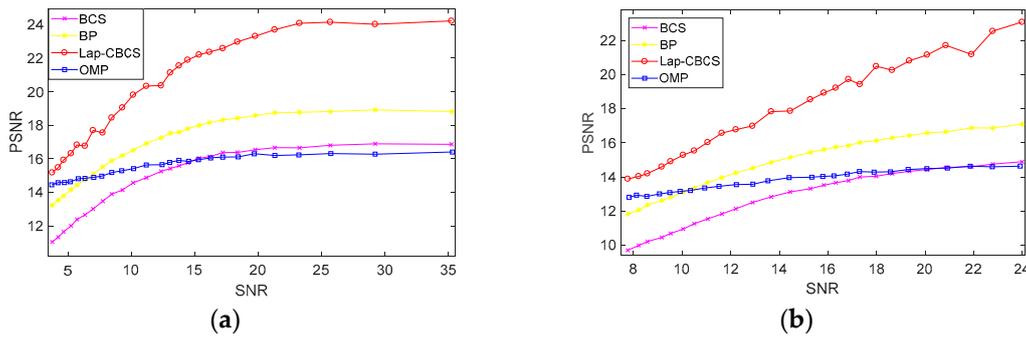


Figure 6. Comparison of the PSNR for different SNR values for datasets (a) IR007_0_105; (b) break10mm.

Furthermore, Figure 7 compares the average running time of Lap-CBCS, OMP, BP, and BCS. Lap-CBCS and BCS consumed more time when the CR was small, which indicates the complexity of Bayesian algorithms. However, the time costs of Lap-CBCS and BCS decreased as the CR approached 1. At this time, BP had a greater time cost than BCS and Lap-CBCS.

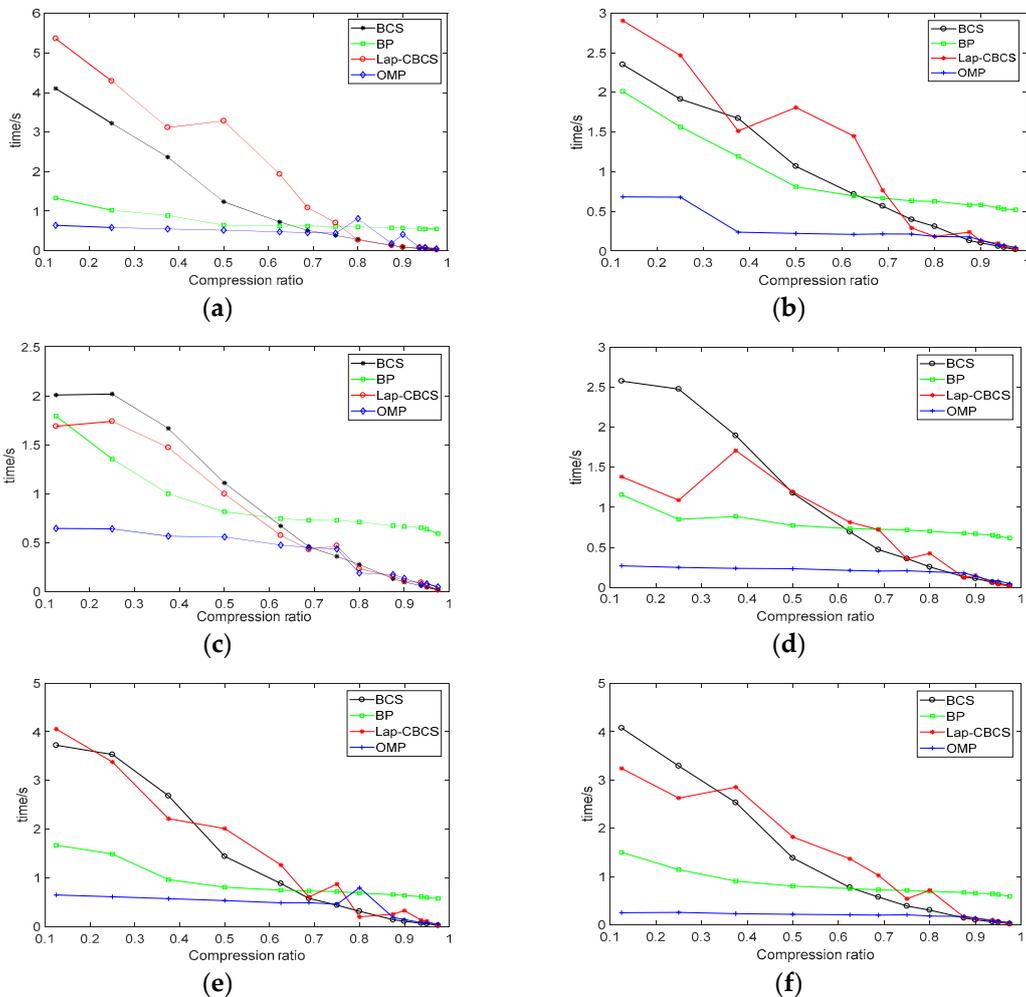


Figure 7. Average time consumption of the four algorithms. The states of the six subpictures are (a) B028_0_3005; (b) IR007_0_105; (c) Break5mm; (d) Break10mm; (e) Crack5mm; (f) Crack8mm.

In summary, when $CR > 0.7$, there was little difference in the time consumption of Lap-CBCS, BCS, and OMP; thus, Lap-CBCS achieved a good balance between cost and efficiency under high CR. Therefore, Lap-CBCS is suitable for reconstructing highly compressed signals.

5.3. Comparison with K-SVD Dictionary Learning and Traditional Sparse Representation

Traditional fixed dictionaries are generally obtained via an orthogonal transform, such as fast Fourier transform (FFT), DCT, and wavelet packet transform (WPT). When the signal characteristics are consistent with the atomic features in the dictionary, an efficient representation can be obtained. However, for real signals, the sparseness is unknown. These fixed orthogonal bases are not sufficiently flexible to represent such signals. An adaptive over-complete dictionary is needed to ensure that the atomic scale in the dictionary is close to that of the original signal.

To compare the sparseness of dictionaries, a threshold ε was set to 2% of the peak-to-peak value of signal x . Thus, we have $\varepsilon = |\max(x) - \min(x)| \times 2\%$. The data points in the range $[-\varepsilon, \varepsilon]$ were set to 0, and the number of nonzero elements in signal x was counted as N_0 . With signal length N , the sparseness of signal x can be represented as

$$\eta = N_0/N \quad (24)$$

In this section, we compare the K-SVD adaptive dictionary with the fixed dictionaries of FFT, DCT, and WPT. The bearing data 12k_Drive_End_OR007@3_0_144 and 12k_Drive_End_B028_0_3005 were randomly selected for the simulation. Firstly, we compared the sparseness of the original signal and that of the transformed signals. A signal segment consisting of 150 points was chosen for comparison. For K-SVD training, the number of atoms was set to 50, and the number of iterations was set to 10. Figure 8 displays the original signal and the results of the K-SVD, DCT, and WPT transforms. The signal after K-SVD dictionary optimization had the smallest sparseness.

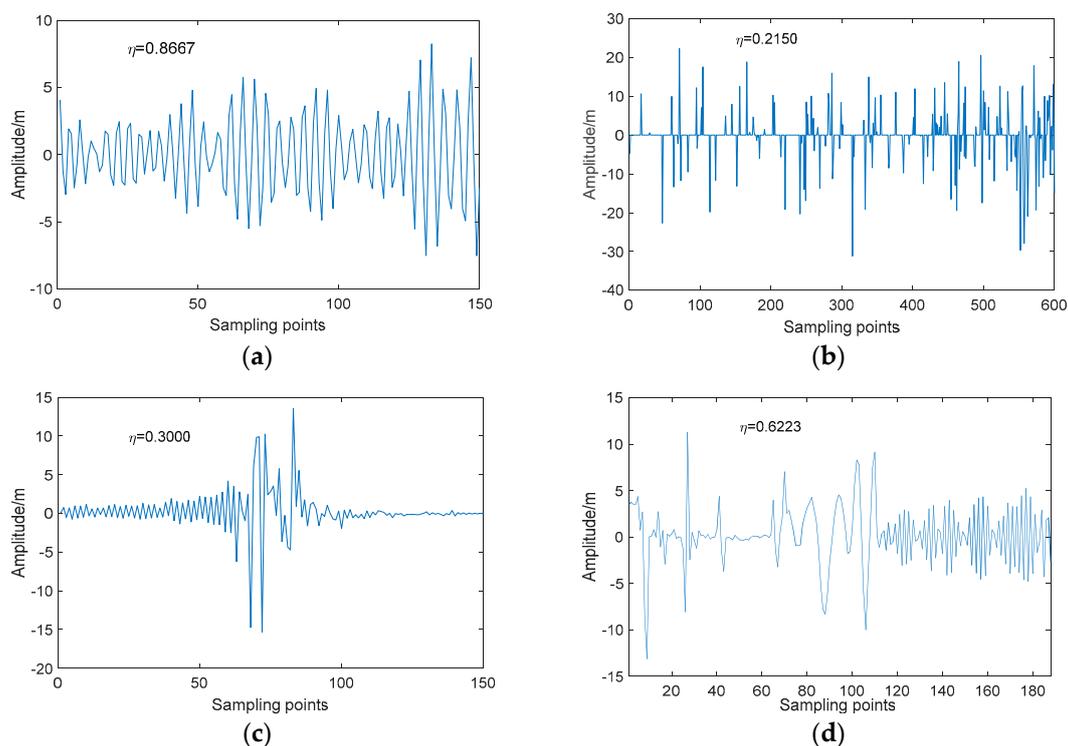


Figure 8. Sparse decomposition of (a) original vibration signal under three types of dictionaries: (b) KSVD; (c) DCT; (d) WPT.

Next, the reconstruction effects were compared. The signal block length was set to 32, and we prepared 200 blocks for training. The remaining 500 blocks were chosen for assessing the dictionary validity. To guarantee a single variable, Lap-CBCS was adopted as the common reconstruction algorithm. Figure 9 shows the MSE, PSNR, and r of the signals reconstructed using different sparse dictionaries.

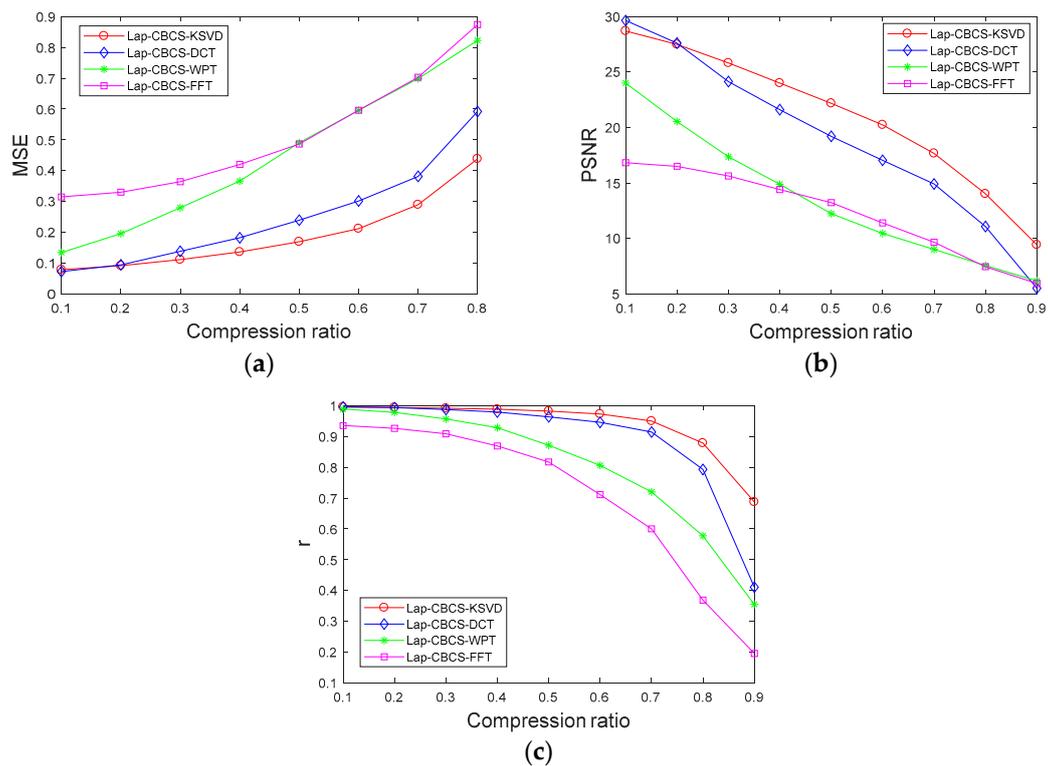


Figure 9. Comparison of different sparse dictionaries in terms of three metrics: (a) MSE; (b) PSNR; (c) r .

Figure 9 shows that the signal reconstructed using the K-SVD dictionary had a small MSE and a large PSNR, which indicates good performance. Comparison of the correlation coefficients confirmed this conclusion. In addition, many factors affected the performance of the K-SVD algorithm. Therefore, all these factors, which are discussed below, must be fully considered.

Figure 10a shows the effects of different initial dictionaries for K-SVD training. As can be seen, r does not vary substantially for the three initial dictionaries. Figure 10b exploits the relationship between the number of atoms and r . The results show that r varied in a small range of [0.92, 0.95], indicating that K-SVD training is not sensitive to the number of atoms. Therefore, no special requirements for the initial dictionary and atoms exist in this paper.

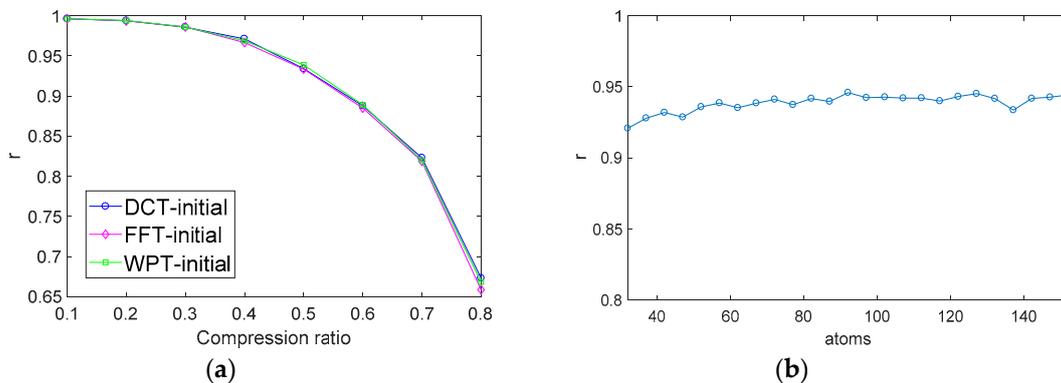


Figure 10. Effects of different parameter configurations in K-SVD training: (a) different initial dictionaries; (b) different atoms.

6. Application of Fault Detection with the Reconstructed Signal

The proposed reconstruction method was validated using signals collected from the planetary gearbox. The structure of the mechanical test rig is shown in Figure 11. Wear failure was seeded on one

tooth of the sun gear, planet gear, and ring gear. The experiments were conducted under the speeds, 400 rpm and 800 rpm, and loads, 0.4 Nm and 1.2 Nm. The sampling frequency was set to 20 kHz. The specific failures are shown in Figure 12.

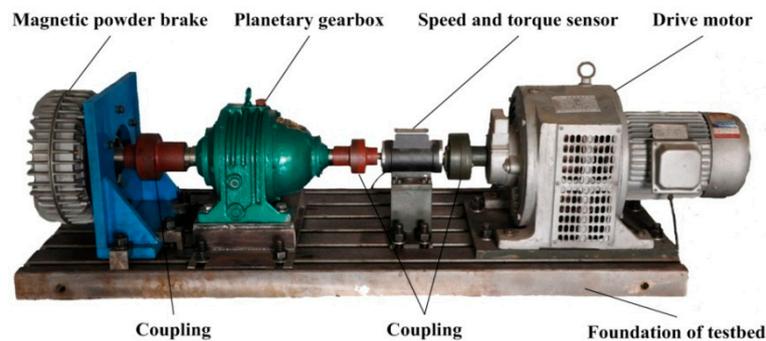


Figure 11. Planetary gearbox test rig.

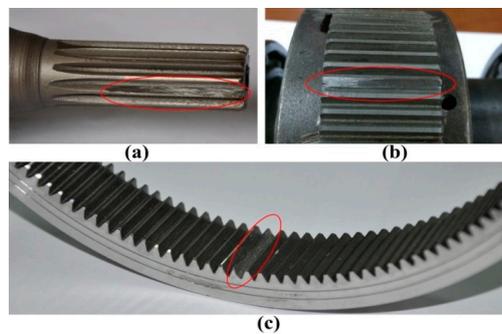


Figure 12. Seeded wear failure: (a) sun gear; (b) planet gear; (c) ring gear.

The ultimate purpose of condition monitoring is to identify fault types. In this section, we use the reconstructed signals, rather than the original signal, for failure classification. The reconstruction effectiveness was validated by comparing the classification accuracies of different methods. The data used in our test can be divided into four categories: (1) normal state, (2) planet gear failure, (3) ring gear failure, and (4) sun gear failure. The training samples and testing samples are shown in Tables 2 and 3. We used the raw signals from the test rig as the training samples and used the reconstructed signals as the testing samples. The signal block length was set to 200 sampling points, and the number of blocks is shown in Tables 2 and 3.

Table 2. Training samples.

State	Motor Load	Motor Speed	Number of Blocks	Data Resource
Normal	0.4 Nm	400 rpm	600	Raw signal
	1.2 Nm	800 rpm	600	
Planet gear failure	0.4 Nm	400 rpm	600	Raw signal
	1.2 Nm	800 rpm	600	
Ring gear failure	0.4 Nm	400 rpm	600	Raw signal
	1.2 Nm	800 rpm	600	
Sun gear failure	0.4 Nm	400 rpm	600	Raw signal
	1.2 Nm	800 rpm	600	

Table 3. Testing samples.

State	Motor Load	Motor Speed	Number of Blocks	Data Resource
Normal	0.4 Nm	400 rpm	300	Reconstructed signal
	1.2 Nm	800 rpm	300	
Planet gear failure	0.4 Nm	400 rpm	300	Reconstructed signal
	1.2 Nm	800 rpm	300	
Ring gear failure	0.4 Nm	400 rpm	300	Reconstructed signal
	1.2 Nm	800 rpm	300	
Sun gear failure	0.4 Nm	400 rpm	300	Reconstructed signal
	1.2 Nm	800 rpm	300	

In the application of fault diagnosis for a planetary gearbox, various features may be sensitive to different equipment states. Therefore, feature extraction is needed before fault classification. We decomposed the planetary gearbox signal using the WPT by three levels and took the energy spectra of eight decomposition nodes as fault features. Support vector machines (SVMs) are state-of-the-art large-margin classifiers that are widely used in pattern recognition and many other applications. RF, which constructs a large set of independent decision trees, was also introduced as a classification algorithm. The results of these trees were combined for the classification task. In this paper, SVM and RF were selected to compare the effectiveness of different reconstruction algorithms. Higher classification accuracy indicates more accurate reconstruction.

The classification accuracy represents the ratio of the number of testing samples identified accurately to the total number of all samples. Figure 13 shows that, as the CR increased, the classification accuracy decreased gradually. However, Lap-CBCS-KSVD achieved the best classification accuracy when the CR was approximately 0.4–0.8. Thus, the same conclusion as that in Section 5.2 was obtained here; Lap-CBCS-KSVD is recommended for reconstructing highly compressed signals. Moreover, Lap-CBCS-KSVD was better than Lap-CBCS-DCT when the CR varied from 0.5–0.8, a common range in practice. The results of all DCT-based methods showed that Lap-CBCS was better than other commonly used reconstruction algorithms, such as BP, OMP, BCS, and ROMP. In accordance with Figure 8, the sparseness η of WPT was the largest, suggesting that the sparse promotion of WPT was the worst among FFT, DCT, and WPT. This constitutes the key reason why Lap-CBCS-WPT reconstructed signals delivered such a bad performance in Figure 13. Moreover, we studied the confusion matrices for different compression ratios, taking CR = 0.1 and 0.9 as examples. With four states represented by N, P, R, and S, we determined the confusion matrices of SVM using the Lap-CBCS-KSVD reconstructed signal. From Table 4, we can find that the planet gear failure, ring gear failure, and sun gear failure were likely to be confused with the normal state when CR was large. However, when CR was decreasing, as in Table 5, the reconstructed signal was more and more easily classified.

Finally, we considered another approach for machinery monitoring. With regard to “sending the whole signal uncompressed and fault detect at the center”, we carried out further experiments, and attained the results described below. The classification accuracy using SVM for original signal was 98.75% and the one making use of random forest was 98.33%. The fault classification of the original signal was much better as compared with that of the reconstructed signal. However, we are interested in stating that, if the classification accuracy of 90% is acceptable, a compressed signal with CR = 0.5 is expected to significantly lower the cost of sensor and transmission. Accordingly, there exists a balance between the classification accuracy and transmission cost. The present paper makes it possible to attain efficient transmission if the classification accuracy is acceptable.

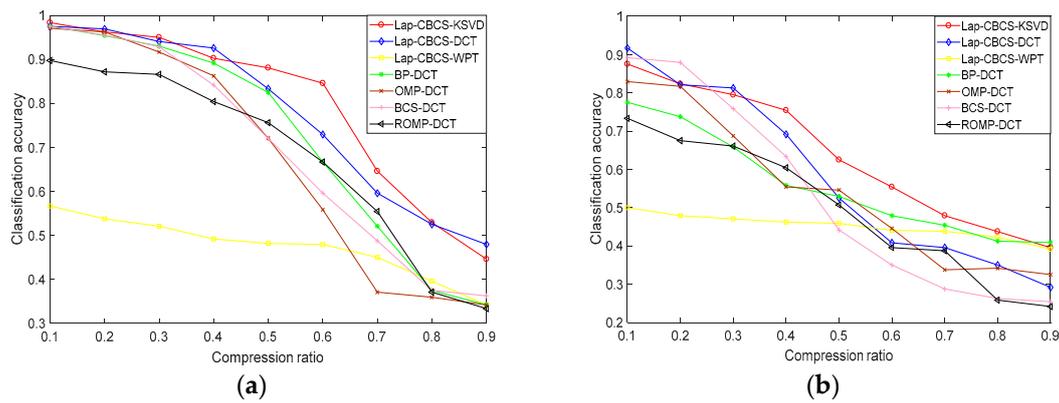


Figure 13. Pattern classification accuracy for different reconstruction methods using (a) SVM classifier; (b) random forest classifier.

Table 4. Confusion matrix (accuracy = 44.58%) of SVM when CR = 0.9.

	<i>N</i>	<i>P</i>	<i>R</i>	<i>S</i>
<i>N</i>	1.0000	0	0	0
<i>P</i>	0.9833	0.0167	0	0
<i>R</i>	0.6000	0	0.2333	0.1667
<i>S</i>	0.4667	0	0	0.5333

Table 5. Confusion matrix (accuracy = 98.33%) of SVM when CR = 0.1.

	<i>N</i>	<i>P</i>	<i>R</i>	<i>S</i>
<i>N</i>	1	0	0	0
<i>P</i>	0	1	0	0
<i>R</i>	0	0	0.9833	0.0167
<i>S</i>	0.05	0	0	0.95

7. Conclusions

Wireless transmission for vibration signals has great potential. However, the high sampling frequency requires an efficient signal compression approach. This paper developed a new technique that aims to impose sparseness over the original signal by means of Laplace priors and extending the algorithm to a multitask scenario. In addition, a K-SVD training method was used for signal sparse decomposition. The reconstruction performance of Lap-CBCS was compared with that of typical algorithms, such as OMP, BP, and BCS. This paper also discussed the benefits of K-SVD dictionary learning. Finally, we presented a fault detection case using the reconstructed signal and several algorithms for verification. The experimental results showed that Lap-CBCS-KSVD achieved good classification accuracy under the SVM and RF models.

Author Contributions: Y.M., X.J. and Q.H. conceived and designed the experiments; Y.M., D.X. and C.G. performed the experiments; Y.M., Q.W. and S.W. analyzed the data; Y.M. and X.J. wrote the paper.

Funding: This work was supported in part by the National Natural Science Foundation of China under the grant 71871220.

Acknowledgments: I would like to express my gratitude to all those who helped me during the writing of this paper. Special acknowledgment is given to my supervisor Professor Xisheng Jia who provided greatly beneficial instructions. Lastly, I would like to thank Justin Romberg, Shihao Ji, and Derin Babacan for sharing their codes.

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

References

1. Tang, B.-P.; Huang, Q.-Q.; Deng, L.; Liu, Z.-R. Research progress and challenges of wireless sensor networks for machinery equipment condition monitoring. *J. Vib. Meas. Diagn.* **2014**, *34*, 1–7.
2. Qi, L.-Q. Research on Wireless Sensor Technology Applied in Bridge Structure Health Monitoring. Ph.D. Thesis, Harbin Institute of Technology, Harbin, China, 2017.
3. Huang, Y.; Beck, J.L.; Wu, S. Bayesian compressive sensing for approximately sparse signals and application to structural health monitoring signals for data loss recovery. *Probab. Eng. Mech.* **2016**, *46*, 62–79. [[CrossRef](#)]
4. Sazonov, E.; Krishnamurthy, V.; Schilling, R. Wireless intelligent sensor and actuator network-A scalable platform for time-synchronous applications of structural health monitoring. *Struct. Health Monit.* **2010**, *9*, 465–476. [[CrossRef](#)]
5. Vogl, A.; Wang, D.T.; Storås, P.; Bakke, T.; Taklo, M.M.V.; Thomson, A.; Balgård, L. Design, process and characterization of a high-performance vibration sensor for wireless condition monitoring. *Sens. Actuators A Phys.* **2009**, *153*, 155–161. [[CrossRef](#)]
6. Shi, M.J.; Luo, R.Z.; Fu, Y.H. Fault diagnosis of rotating machinery based on wavelet and energy feature extraction. *J. Electron. Meas. Instrum.* **2015**, *29*, 1114–1120.
7. Rissanen, J.; Langdon, G.G. Arithmetic coding. *IBM J. Res. Dev.* **1979**, *23*, 149–162. [[CrossRef](#)]
8. Gallager, R. Variations on a theme by Huffman. *Trans. Inf. Theory* **1978**, *24*, 668–674. [[CrossRef](#)]
9. Candès, E.J. Compressive sampling. In Proceedings of the International Congress of Mathematics, Madrid, Spain, 22–30 August 2006; Volume 3, pp. 1433–1452.
10. Candès, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215. [[CrossRef](#)]
11. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
12. Tropp, J.A. Computational methods for sparse solution of linear inverse problems. *Proc. IEEE* **2009**, *98*, 948–958. [[CrossRef](#)]
13. Tropp, J.A.; Gilbert, A.C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [[CrossRef](#)]
14. Ji, S.; Xue, Y.; Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Process* **2008**, *56*, 2346–2356. [[CrossRef](#)]
15. Ji, S.; Dunson, D.; Carin, L. Multi-task compressive sensing. *IEEE Trans. Signal Process* **2009**, *57*, 92–106. [[CrossRef](#)]
16. Zhang, Z.; Rao, B.D. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 912–926. [[CrossRef](#)]
17. Babacan, S.; Molina, R.; Katsaggelos, A. Bayesian compressive sensing using Laplace priors. *IEEE Trans. Image Process* **2010**, *19*, 53–63. [[CrossRef](#)]
18. Torkamani, R.; Sadeghzadeh, R.A. Bayesian compressive sensing using wavelet based Markov random fields. *Signal Process. Image Commun.* **2017**, *58*, 65–72. [[CrossRef](#)]
19. Luo, K.; Li, J.-Q.; Wang, Z.-G.; Cai, Z.-P. Prior-block sparse Bayesian learning algorithm for compressed sensing based ECG recovery. *Chin. J. Sci. Instrum.* **2014**, *35*, 1883–1889.
20. Ma, W.-J. DOA estimation through Bayesian compressive sensing algorithm. Ph.D. Thesis, Harbin Institute of Technology, Harbin, China, 2014.
21. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Inf. Theory* **2006**, *15*, 3736–3745. [[CrossRef](#)]
22. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [[CrossRef](#)]
23. Zhou, Y.-T.; Wang, L.-L.; Tang, H.-M. Sparsity adaptive algorithm for image inpainting based on compressive sensing. *J. China Railw. Soc.* **2014**, *36*, 52–59.
24. Yang, A.-P.; Tian, Y.-Z.; He, Q.-Y. Image denosing based on improved K-SVD and non-local regularization. *Comput. Eng.* **2015**, *41*, 249–253.
25. Shi, J.; Wang, X.-H. Image super resolution reconstruction based on improved K-SVD dictionary learning. *Acta Electron. Sin.* **2013**, *41*, 997–1000.
26. Jafari, M.G.; Plumbley, M.D. Fast dictionary learning for sparse representations of speech signals. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 1025–1031. [[CrossRef](#)]

27. Ophir, B.; Lustig, M.; Elad, M. Multi-scale dictionary learning using wavelets. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 1014–1024. [[CrossRef](#)]
28. Wang, Y.X.; Xiang, J.W.; Mo, Q.Y.; He, S.L. Compressed sparse time–frequency feature representation via compressive sensing and its applications in fault diagnosis. *Measurement* **2015**, *68*, 70–81. [[CrossRef](#)]
29. Tang, G.; Hou, W.; Wang, H.Q.; Luo, G.G.; Ma, J.W. Compressive Sensing of Roller Bearing Faults via Harmonic Detection from Under-Sampled Vibration Signals. *Sensors* **2015**, *15*, 25648–25662. [[CrossRef](#)]
30. Sun, J.D.; Yu, Y.; Wen, J.T. Compressed-Sensing Reconstruction Based on Block Sparse Bayesian Learning in Bearing-Condition Monitoring. *Sensors* **2017**, *17*, 1454. [[CrossRef](#)]
31. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
32. Available online: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 27 March 2019).
33. Needell, D.; Vershynin, R. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Sel. Top. Signal Process.* **2010**, *4*, 310–316. [[CrossRef](#)]



© 2019 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/>).