

The sparse and low-rank interpretation of SVD-based denoising for vibration signals

1st Zhibin Zhao

*School of Mechanical Engineering
Xi'an Jiaotong University
Xi'an, China
zhibinzhao1993@gmail.com*

zhibinzhao1993@gmail.com

2nd Shibin Wang

*School of Mechanical Engineering
Xi'an Jiaotong University
Xi'an, China
wangshibin2008@gmail.com*

3rd David Wong

Centre for Health Informatics
University of Manchester
Manchester, UK
david.wong@manchester.ac.uk

4th Yanjie Guo

*School of Mechanical Engineering
Xi'an Jiaotong University
Xi'an, China
guoyanjie@mail.xjtu.edu.cn*

guoyanjie@mail.xjtu.edu.cn

5th Xuefeng Chen

*School of Mechanical Engineering
Xi'an Jiaotong University
Xi'an, China
chenxf@mail.xjtu.edu.cn*

Abstract—Vibration signal denoising is one of the most important steps in condition monitoring and fault diagnosis, and SVD-based methods are a vital part of advanced signal denoising due to their non-parametric and simple properties. The relationships between SVD-based denoising and other advanced signal processing methods are very significant and can help speed up the development of SVD-based denoising methods. There is limited prior work into the sparse and low-rank meaning of SVD-based denoising. In this paper, we build the relationships among SVD-based denoising, sparse l_0 -norm minimization, sparse weighted l_1 -norm minimization, and weighted low-rank models, when the dictionary is designed by left and right singular matrices in sparse minimization. Using the derived conclusion, we establish weighted soft singular value decomposition (WSSVD) for vibration signal denoising. Finally, we perform one experimental study to verify the effectiveness of WSSVD considering impulse interference and amplitude fidelity.

Index Terms—Sparse and low-rank, SVD-based denoising, vibration signal processing

1 INTRODUCTION

Condition monitoring and fault diagnosis are becoming increasingly popular in modern rotating machinery due to their complex structure and poor operating environment. Vibration signal monitoring is one of the most important monitoring methods for rotating machines since vibration signals are more sensitive to faults at the early stage than sound and temperature. Indicators in the time and frequency domains are often used to track the health condition of rotating machines. However, because fault features are often submerged by heavy background noise and discrete frequency interference, advanced signal processing methods [1], such as Wavelet Transform (WT), Spectral Kurtosis (SK), Empirical Mode Decomposition (EMD), sparsity-assisted methods, and Singular

This work was supported in part by the Natural Science Foundation of China under Grand 51605366 and 51705398, by the China Postdoctoral Science Foundation under Grant 2016M590937 and 2017T100740, and by the Fundamental Research Funds for the Central Universities. Shibin Wang is the corresponding author.

Value Decomposition based (SVD-based) methods, have been continuously proposed to realize vibration signal denoising and fault feature extraction.

Among these advanced signal processing methods, SVD-based denoising methods have received much attention because SVD is a non-parametric and simple technique to decompose the original signal into orthogonal subspaces. One of the most popular methods is called the truncated SVD (TSVD) which consists of four steps: constructing a matrix through the measured signal; decomposing the constructed matrix using SVD; choosing the first t largest singular values to realize the denoising step; reconstructing the denoised signal from the denoised matrix. The main studies in this domain can be classified into two categories: how to construct a matrix, and how to set the threshold to choose singular values. In [2], SVD and ratios of neighboring singular values were applied to the vibration signal processing and extracted features were used to train the hidden Markov model for fault recognition. In [3], SVD and a Hankel matrix was successfully used as a denoising step for ball bearing fault feature extraction. In [4], the authors proposed a reweighted singular value decomposition (RSVD) for vibration signal denoising and weak fault feature extraction. However, minimal prior research establishes the relationship between SVD-based denoising methods and other advanced signal processing methods. Pioneering work aiming to bridge this gap was done by Zhao et al [5]. They verified the similarity between Hankel-SVD and wavelet transform by theoretical analysis and experimental verification. Based on this similarity, they proposed the method called singular value decomposition packet which embedded advantages of the wavelet packet transform [6].

Recently, the sparse and low-rank physical meaning has been successfully applied to vibration signal denoising and fault diagnosis [7], [8]. Du et al [9] proposed a weighted low-rank sparse model for bearing fault diagnosis. Zhang et al [10] extended this model to the fault diagnosis of high-

speed aero-engine bearings by improving the construction of the matrix using k-means clustering. In [11] and [12], the sparse and low-rank model was used to denoise the vibration signal and perform bearing fault diagnosis. However, to best of our knowledge, no one has studied the sparse and low-rank interpretation of SVD-based denoising.

In this paper, we aim to bridge this gap and discuss the sparse and low-rank meaning of SVD-based denoising methods thoroughly. On the one hand, we conclude that TSVD is equivalent to the sparse l_0 -norm minimization when the dictionary is constructed by left and right singular matrices. On the other hand, the sparse weighted l_1 -norm minimization is a more general condition, and TSVD and the weighted low-rank model are the special cases of weighted l_1 -norm minimization. Meanwhile, based on the discussion of relationships, we apply the weighted soft thresholding to establish weighted soft singular value decomposition (WSSVD). Compared with RSVD, WSSVD can maintain the amplitude more efficiently due to the fact that WSSVD is based on the theory of weighted l_1 -norm minimization and has the corresponding amplitude fidelity term. Lastly, we use an open dataset to verify the performance of WSSVD in terms of impulse interference and amplitude fidelity.

This paper is organized as follows: In Section II, we briefly introduce the basic steps of SVD-based denoising methods. Section III gives the sparse and low-rank interpretation of SVD-based denoising methods. Section IV performs the experimental verification.

II. REVIEW OF SVD-BASED DENOISING METHODS

Commonly, the measured signal from the rotating machinery consists of the fault components and the other components, and the basic signal model can be described as follows:

$$y = x + e \quad (1)$$

where $y \in \mathbb{R}^{m \times 1}$ (m is the dimension of the vector) is the measured signal, $x \in \mathbb{R}^{m \times 1}$ is the fault features which we need to extract, and $e \in \mathbb{R}^{m \times 1}$ is the discrete frequency interference and background noise. SVD-based methods aim to extract the fault features from discrete frequency interference and noise. In this section, we briefly introduce the basic procedures of SVD-based denoising methods and these basic procedures are also called TSVD.

A. Construction of the matrix

Many different methods for constructing a matrix for SVD have been proposed in the past studies, including the Hankel matrix which is the most commonly used, the Toeplitz matrix, and the cycle matrix. However, in this part, we will use a more general way to describe the construction of the Hankel matrix. First, we need to define one operator $R_k(\cdot)$ which represents to extract k -th patch from the time series and the inverse operator is defined as $R_k^T(\cdot)$ to put the corresponding patch back to the time series. Meanwhile, $R(\cdot)$ and $R^T(\cdot)$ represent conversion between the time series and the matrix form, and this step can be implemented via a MATLAB function “buffer”

directly. For example, $y_k = R_k(y) \in \mathbb{R}^{l \times 1}$ and $x_k = R_k(x) \in \mathbb{R}^{l \times 1}$ represent the patches from the measured signal and the fault signal with the patch length l , respectively. Therefore, for the measured signal y , its constructed matrix can be formed as

$$\begin{aligned} Y &= [y_1, y_2, \dots, y_n] \\ &= [R_1(y), R_2(y), \dots, R_n(y)] = R(y) \end{aligned} \quad (2)$$

where $Y \in \mathbb{R}^{l \times n}$ is the matrix version of y , n is the number of total patches, and l is the length of each patch. The value n depends on l , m which is the length of the measured signal, and the overlap, s which is the overlap between every two adjacent blocks. If the overlap, s , is equal to $l - 1$, the constructed matrix is equivalent to the Hankel matrix.

B. Decomposition

When we have the constructed matrix, we can perform the signal decomposition (SVD). For the measured signal y , its constructed matrix Y can be decomposed into the sum of subspaces by SVD

$$Y = U\Sigma V^T = \sum_{i=1}^{\min(l,n)=N} \sigma_i u_i v_i^T \quad (3)$$

where $U \in \mathbb{R}^{l \times l}$ is the left singular matrix whose columns are set as u_i , $\Sigma \in \mathbb{R}^{l \times n}$ is the diagonal matrix whose diagonal elements are set as the singular value σ_i , and $V \in \mathbb{R}^{n \times n}$ is the right singular matrix whose columns are set as v_i . It should be mentioned that the singular value series σ_i is decreasing ($\sigma_1 \geq \sigma_2 \geq \dots$).

C. Denoising

In most SVD-based denoising methods, we often assume that the fault features are lying in the corresponding subspace with larger singular values. Thus, how to select a threshold to shrinkage the singular values is important, and many researchers have proposed subtle methods embedding the physical meanings. If the threshold τ is used to shrink singular values, only the first t largest singular values remain, i.e.

$$\hat{X} = \sum_{i=1}^t \sigma_i u_i v_i^T \quad (\forall \sigma_i > \tau) \quad (4)$$

where \hat{X} is an approximation of the desired matrix X .

D. Reconstruction

When we get the denoised matrix \hat{X} (if the Hankel matrix is used as the constructed matrix, the direct or anti-diagonals averaging methods mentioned in [4] can be used to reconstruct the time series), we can recover the fault signal using a simple least square estimation

$$\min_x \|\hat{X} - R(x)\|_F^2. \quad (5)$$

This optimization problem can be solved by letting the derivative of (5) equal to zero, and the close-form solution is

$$\hat{x} = \sum_{k=1}^n R_k^T(\hat{X}) ./ \sum_{k=1}^n R_k^T(\mathbf{1}) \quad (6)$$

where $\mathbf{1} \in \mathbb{R}^{l \times n}$ is a matrix whose elements are all equal to one, \hat{x} is the optimal solution of the optimization problem (5), and $./$ represents element-wise division.

III. SPARSE AND LOW-RANK INTREPRETATION

To make the interpretation clearer, we first define a dictionary D and its representation coefficients α as follows:

$$D = [u_1 v_1^T, \dots, u_i v_i^T, \dots, u_N v_N^T] \in \mathbb{R}^{(l \times n) \times N}, \quad (7)$$

$$\alpha = [\alpha_1, \dots, \alpha_i, \dots, \alpha_N]^T \in \mathbb{R}^{N \times 1}.$$

where α_i is a scalar. Then if we further define a representation $\sum_{i=1}^N \alpha_i u_i v_i^T = D \cdot \alpha$, we can define the following optimization problem to find an approximation of the matrix X .

$$\min_{\alpha} \frac{1}{2} \|Y - D \cdot \alpha\|_F^2 + \lambda P(\alpha) \quad (8)$$

where $\|\cdot\|_F$ is the Frobenius norm, λ is the trade-off parameter, and $P(\alpha)$ is the regularization of α . If the optimization problem (8) does not have the regularization, the optimal solution is directly the singular values of Y due to the fact that D is a combination of singular matrices of Y .

In addition, We give *Lemma 1* [13], which is the key of the derivation

Lemma 1:

$$\|D \cdot \alpha - D \cdot \hat{\alpha}\|_F^2 = \|\alpha - \hat{\alpha}\|_2^2 \quad (9)$$

Proof 1:

$$\begin{aligned} \|D \cdot \alpha - D \cdot \hat{\alpha}\|_F^2 &= \left\| \sum_{i=1}^N \alpha_i u_i v_i^T - \sum_{i=1}^N \hat{\alpha}_i u_i v_i^T \right\|_F^2 \\ &= \|U \text{diag}(\alpha) V^T - U \text{diag}(\hat{\alpha}) V^T\|_F^2 \quad (10) \\ &= \|\text{diag}(\alpha) - \text{diag}(\hat{\alpha})\|_F^2 \\ &= \|\alpha - \hat{\alpha}\|_2^2 \end{aligned}$$

where $\text{diag}(\cdot)$ means to generate the singular value matrix. The second equation to the third equation is due to the fact that U and V are all orthonormal matrices.

A. Case 1: $P(\alpha) = \|\alpha\|_0$

If the regularization $P(\alpha) = \|\alpha\|_0$, the optimization problem (8) can be reformulated as the sparse approximation problem

$$\min_{\alpha} \frac{1}{2} \|Y - D \cdot \alpha\|_F^2 + \lambda \|\alpha\|_0 \quad (11)$$

where $\|\cdot\|_0$ is the l_0 -norm representing the number of non-zero values. According to the SVD of Y and *Lemma 1*, the optimization problem (11) can be simplified into

$$\begin{aligned} \min_{\alpha} \frac{1}{2} \|D \cdot \sigma - D \cdot \alpha\|_F^2 + \lambda \|\alpha\|_0 \\ \Rightarrow \min_{\alpha} \frac{1}{2} \|\sigma - \alpha\|_2^2 + \lambda \|\alpha\|_0 \end{aligned} \quad (12)$$

where $\sigma = [\sigma_1, \dots, \sigma_i, \dots, \sigma_N]^T$ is the vector of singular values.

Lemma 2: The optimal solution of (12) is the hard thresholding $\hat{\alpha}_i = \text{hard}(\sigma_i, \sqrt{2\lambda}) \quad \forall i \in [1, N]$.

Proof 2: The problem can be separated into N sub-problems which are defined as follows:

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \frac{1}{2} (\sigma_i - \alpha_i)^2 + \lambda |\alpha_i|^0 \quad (13)$$

where $|\alpha_i|^0 = 1$ when $\alpha_i \neq 0$. Otherwise, $|\alpha_i|^0 = 0$.

From the optimization problem (13), there are only two options for the value α_i . In the first condition, $\alpha_i = \sigma_i$ sets the cost value as λ . In the second condition, $\alpha_i = 0$ sets the cost value as $\frac{1}{2} \sigma_i^2$. Therefore, we just need to compare these two cost values, and if $|\sigma_i| \leq \sqrt{2\lambda}$, $\alpha_i = 0$ is the optimal solution. Otherwise, $\alpha_i = \sigma_i$ is the optimal solution for $|\sigma_i| > \sqrt{2\lambda}$. The final solver can be concluded as:

$$\hat{\alpha}_i = \text{hard}(\sigma_i, \sqrt{2\lambda}) = \text{sgn}(\sigma_i) \max \{ |\sigma_i|, \sqrt{2\lambda} \} \quad (14)$$

where $\text{sgn}(\cdot)$ extracts the sign of a real number, $\max \{ \cdot, \cdot \}$ extracts the larger value in the curly braces.

According to *Lemma 2*, we can directly find an approximation of X , e.g. $\hat{X} = \sum_{i=1}^{t'} \sigma_i u_i v_i^T$, $\forall \sigma_i > \sqrt{2\lambda}$. Based on *Lemma 2*, we can have the following theorem.

Theorem 1: TSVD is equivalent to the sparse approximation problem (l_0 -norm minimization) in (11), when the threshold τ is equivalent to $\sqrt{2\lambda}$.

B. Case 2: $P(\alpha) = \|\alpha\|_1$

If the regularization $P(\alpha) = \|\alpha\|_1$, the optimization problem (8) can be reformulated as the sparse approximation problem (l_1 -norm minimization)

$$\min_{\alpha} \frac{1}{2} \|Y - D \cdot \alpha\|_F^2 + \lambda \|\alpha\|_1 \quad (15)$$

where $\|\cdot\|_1$ is the l_1 -norm representing the sum of absolute values. Similar to the derivation in (12), the optimiation problem (15) can be simplified into

$$\min_{\alpha} \frac{1}{2} \|\sigma - \alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (16)$$

Lemma 3: The optimal solution of (16) is called the soft thresholding [14] $\hat{\alpha}_i = \text{soft}(\sigma_i, \lambda) \quad \forall i \in [1, N]$.

Proof 3: Additionally, it is very easy to deduce the optimal value based on sub-gradient method [15]. The sub-gradient of $|\alpha_i|$ is defined as follows:

$$\partial(|\alpha_i|) = \begin{cases} 1 & \alpha_i > 0 \\ [-1, 1] & \alpha_i = 0 \\ -1 & \alpha_i < 0 \end{cases} \quad (17)$$

Thus, we can let the first derivative to be zero, and the equality (18) can be used to find the optimal value of the optimization problem (16).

$$\alpha_i - \sigma_i + \lambda \partial(|\alpha_i|) = 0. \quad (18)$$

Then, the final solver can be concluded as:

$$\hat{\alpha}_i = \text{soft}(\sigma_i, \lambda) = \text{sgn}(\sigma_i) \max\{|\sigma_i| - \lambda, 0\} \quad (19)$$

According to *Lemma 3*, the approximation of X can be set as $\hat{X} = \sum_{i=1}^{t'} \hat{\alpha}_i u_i v_i^T$. Based on results in [16], the above approximation is equivalent to the following optimization problem (also called low-rank approximation based on nuclear norm)

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_* \quad (20)$$

where $\|X\|_*$ is the nuclear norm of X which represents the sum of singular values.

Although TSVD is not strictly equivalent to the l_1 -norm minimization and the low-rank approximation (20), there are significant similarities. The only difference exists in hard thresholding and soft thresholding, and hard thresholding is cruder than soft thresholding.

C. Case 3: $P(\alpha) = \|\alpha\|_{1,w}$

In this case, we will discuss a more general version of l_1 -norm minimization. If the regularization $P(\alpha) = \|\alpha\|_{1,w}$, the optimization problem (8) can be reformulated as the sparse approximation problem (weighted l_1 -norm minimization)

$$\min_{\alpha} \frac{1}{2} \|Y - D \cdot \alpha\|_F^2 + \lambda \|\alpha\|_{1,w} \quad (21)$$

where $\|\cdot\|_{1,w}$ is the weight l_1 -norm representing the weighted sum of absolute values ($\|\alpha\|_{1,w} = \sum_{i=1}^N w_i |\alpha_i|$). Similar to the derivation in (12) and (16), the optimization problem (21) can be simplified into

$$\min_{\alpha} \frac{1}{2} \|\sigma - \alpha\|_2^2 + \lambda \|\alpha\|_{1,w} \quad (22)$$

Lemma 4: The optimal solution of (21) is called the weighted soft thresholding $\hat{\alpha}_i = \text{soft}(\sigma_i, \lambda w_i) \quad \forall i \in [1, N]$.

In addition, the proof of *Lemma 4* is similar to *Lemma 3*, so we will not repeat it here.

Meanwhile, we introduce the weighted low-rank model directly to discuss the relationships between these three methods.

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_{*,w} \quad (23)$$

where $\|X\|_{*,w} = \sum_{i=1}^N w_i \alpha_i$ and α_i represents i -th singular value of the matrix X .

Theorem 2: TSVD is equivalent to the weighted l_1 -norm minimization in (11) and weighted low-rank model in (23), when the weight $w = [\underbrace{0, \dots, 0}_t, 1, \dots, 1]^T$.

Proof 4: If the weight w satisfies $w_1 \leq w_2 \leq \dots \leq w_N$, the optimal solution of weighted low-rank model is the weighted soft thresholding $\hat{\alpha}_i = \text{soft}(\sigma_i, \lambda w_i) \quad \forall i \in [1, N]$ [17]. Therefore, these three methods are equivalent to each other when $w = [\underbrace{0, \dots, 0}_t, 1, \dots, 1]^T$.

In fact, the method called reweighted SVD (RSVD) proposed by Zhao [4] is also similar to the weighted l_1 -norm

minimization. The core idea of reweighted SVD is to design an information indicator vector called periodic modulation intensity (PMI) p to rescale singular values. The weight w is defined as follows:

$$w_i = \begin{cases} 0, & p_i < 1 \\ \frac{p_i}{\sum_i p_i}, & p_i \geq 1 \end{cases} \quad (24)$$

And the result $\hat{\alpha}_i$ is set as $w_i \sigma_i$. However, this step will destroy amplitude information of fault features, which is not suitable for condition monitoring. Based on the above discussion about *Theorem 1* and *Theorem 2*, we use the weighted soft thresholding to calculate the desired singular values. Meanwhile, the weight w is defined as follows

$$w_i = \frac{1}{c_i}, \text{ where } c_i = \frac{p_i}{\max(p)} \quad (25)$$

So, we construct the model called weighted soft singular value decomposition (WSSVD) depicted in Algorithm 1 and its desired singular values is calculated using $\alpha_i = \text{soft}(\sigma_i, \lambda w_i)$ (*Lemma 4*). Meanwhile, the approximation of X can be set as $\hat{X} = \sum_{i=1}^N \hat{\alpha}_i u_i v_i^T$.

Algorithm 1 WSSVD

Input: The measured signal $y \in \mathbb{R}^{m \times 1}$, $\lambda > 0$, the length of each patch l , and the length of overlap s .

Constructing the matrix:

$$Y = R(y)$$

SVD decomposition:

$$Y = U \Sigma V^T = \sum_{i=1}^{\min(l,n)=N} \sigma_i u_i v_i^T$$

Calculating the weights:

$$w_i = \frac{1}{c_i}, \text{ where } c_i = \frac{p_i}{\max(p)}$$

Denoising:

$$\alpha_i = \text{soft}(\sigma_i, \lambda w_i)$$

Reconstruction:

$$\hat{X} = \sum_{i=1}^N \hat{\alpha}_i u_i v_i^T$$

$$\hat{x} = \sum_{k=1}^n R_k^T(\hat{X}) ./ \sum_{k=1}^n R_k^T(1)$$

Output:

IV. EXPERIMENTAL STUDY

In this section, we use an open dataset to verify the performance of WSSVD considering the impulse interference and the amplitude fidelity.

A. Experimental Description

In this experimental study, we use the measured data from NSF I/UCR Center for Intelligent Maintenance Systems (www.imscenter.net) [18]. This experiment is a lifetime experiment collecting acceleration data every ten minutes, and the sampling frequency is 20 kHz. The bearing in the test rig had an outer race fault, and the fault characteristic frequency is 236.4 Hz according to geometric parameters of this bearing.

Root Mean Square (RMS) of the whole life cycle is shown in Fig. 1. The early stage in file No.558 and the medium stage in file No.703 are labelled in the curve. For verification of the impulse interference resistance ability of WSSVD, we add

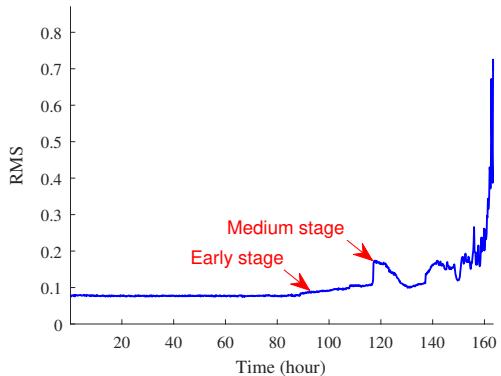


Fig. 1. RMS of vibration signals.

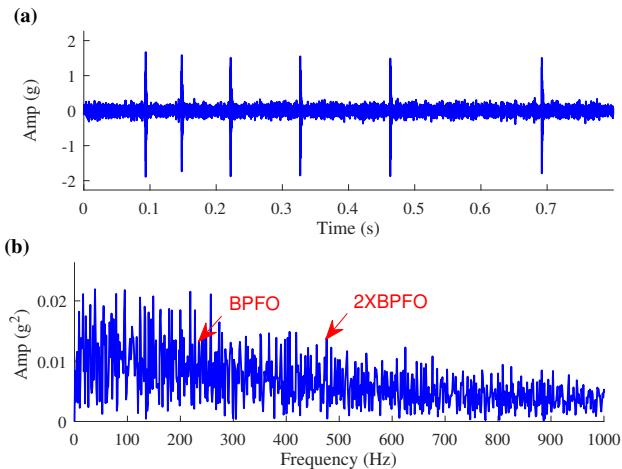


Fig. 2. (a) The measured data in time domain, and (b) its corresponding square envelope spectrum.

some random impulses in the measured data as shown in Fig. 2 (a). It can be observed from Fig. 2 (b) that the first order fault characteristic frequency is submerged in the background noise and random impulses.

B. Results

The length of each patch l and the overlap s are set as 40 and $l-1$, respectively. The trade-off parameter λ in WSSVD is set as 2 by trial and error. As shown in Fig. 3 (a), the random impulses are denoised by the proposed method completely. In addition, the first three order fault characteristic frequencies are enhanced in the square envelope spectrum shown in Fig. 3 (b) and these amplitudes are close to 10^{-3} . Compared to RSVD, we first tune the weight in it finely, and the modified weight is defined as

$$w_i = \begin{cases} 0, & p_i < 0.1 \\ \frac{p_i}{\sum_i p_i}, & p_i \geq 0.1 \end{cases}. \quad (26)$$

As shown in Fig. 4, RSVD can also extract the fault information due to the fact that RSVD also has the function of choosing the useful subspaces. However, the amplitude (10^{-5}) extracted by RSVD is much smaller than that (10^{-3}) extracted by WSSVD.

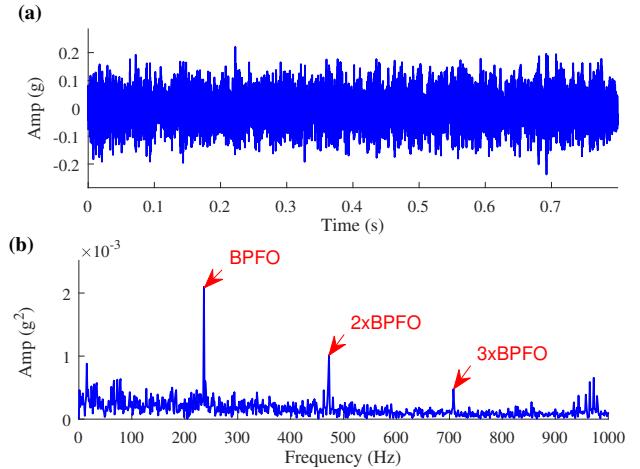


Fig. 3. (a) The extracted results of vibration signals at the early stage via WSSVD, and (b) its corresponding square envelope spectrum.

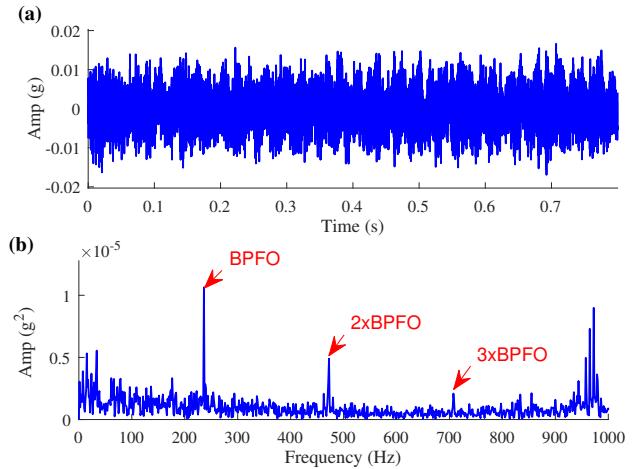


Fig. 4. (a) The extracted results of vibration signals at the early stage via RSVD, and (b) its corresponding square envelope spectrum.

Besides, we also use TSVD to copy with this dataset and the threshold in TSVD is calculated by the difference spectrum which is used to capture the abrupt change. The extracted results are shown in Fig. 5, and it can be observed that TSVD can only extract the random impulses and ignore the fault information.

V. CONCLUSION

This paper aims to discuss the sparse and low-rank meaning of SVD-based denoising methods. The core conclusions are classified into two categories when the dictionary is designed using left and right singular matrices. The first one is that TSVD is equivalent to sparse l_0 -norm minimization, and the second is that TSVD and the weighted low-rank model are the special cases of sparse weighted l_1 -norm minimization. Based on the description of weighted soft thresholding, we form a model called WSSVD which can maintain amplitude more efficiently than RSVD. Also, further studies can adopt other advanced theories and methods existing in sparse and

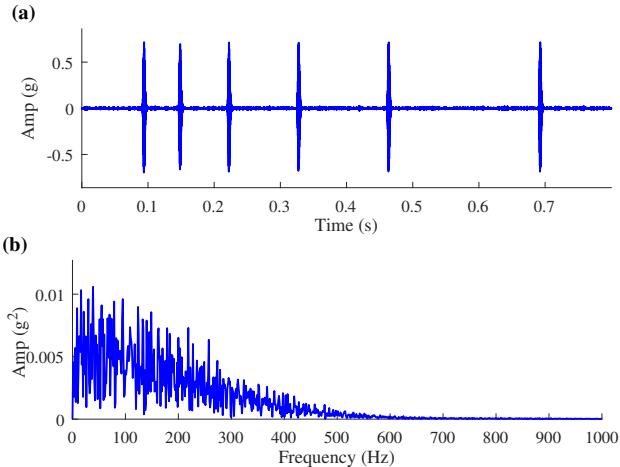


Fig. 5. (a) The extracted results of vibration signals at the early stage via TSVD, and (b) its corresponding square envelope spectrum.

low-rank fields to make SVD-based denoising methods more efficient.

REFERENCES

- [1] Z. Zhao, B. Qiao, S. Wang, Z. Shen, and X. Chen, "A weighted multi-scale dictionary learning model and its applications on bearing fault diagnosis," *Journal of Sound and Vibration*, vol. 446, pp. 429–452, 2019.
- [2] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen, "Study on hankel matrix-based svd and its application in rolling element bearing fault diagnosis," *Mechanical systems and signal processing*, vol. 52, pp. 338–359, 2015.
- [3] R. Golafshan and K. Y. Sanliturk, "Svd and hankel matrix based denoising approach for ball bearing fault detection and its assessment using artificial faults," *Mechanical Systems and Signal Processing*, vol. 70, pp. 36–50, 2016.
- [4] M. Zhao and X. Jia, "A novel strategy for signal denoising using reweighted svd and its applications to weak fault feature enhancement of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 94, pp. 129–147, 2017.
- [5] X. Zhao and B. Ye, "Similarity of signal processing effect between hankel matrix-based svd and wavelet transform and its mechanism analysis," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1062–1075, 2009.
- [6] X. Zhao and B. Ye, "Singular value decomposition packet and its application to extraction of weak fault feature," *Mechanical Systems and Signal Processing*, vol. 70, pp. 73–86, 2016.
- [7] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, and X. Chen, "Non-convex sparse regularization and convex optimization for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7332–7342, 2018.
- [8] Z. Zhao, S. Wu, B. Qiao, S. Wang, and X. Chen, "Enhanced sparse period-group lasso for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 3, pp. 2143–2153, 2018.
- [9] Z. Du, X. Chen, H. Zhang, B. Yang, Z. Zhai, and R. Yan, "Weighted low-rank sparse model via nuclear norm minimization for bearing fault detection," *Journal of Sound and Vibration*, vol. 400, pp. 270–287, 2017.
- [10] H. Zhang, X. Chen, and X. Zhang, "A clustering low-rank approach for aero-enging bearing fault detection," in *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2019, pp. 1–6.
- [11] G. Xin, Y. Qin, L.-M. Jia, S.-J. Zhang, and J. Antoni, "Low-rank and sparse model: A new perspective for rolling element bearing diagnosis," in *2018 International Conference on Intelligent Rail Transportation (ICIRT)*. IEEE, 2018, pp. 1–5.
- [12] L. Yu, W. Dai, S. Huang, and W. Jiang, "Sparse time-frequency representation for the transient signal based on low-rank and sparse decomposition," *Journal of Acoustics*, vol. 1, no. 1, 2019.
- [13] Z. Zha, X. Liu, X. Huang, H. Shi, Y. Xu, Q. Wang, L. Tang, and X. Zhang, "Analyzing the group sparsity based on the rank minimization methods," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 883–888.
- [14] D. L. Donoho, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [15] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University, Autumn Quarter*, vol. 2004, pp. 2004–2005, 2003.
- [16] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [17] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2862–2869.
- [18] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of sound and vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.