IJCTA, 9(10), 2016, pp. 4433-4443

Non–Convex Group Sparsity Denoising for Bearing Fault Diagnosis Using SVM

***Archana Chandran *Neethu Mohan *K.P. Soman**

Abstract : Bearings are the pivotal components in rotating machines whose failure can result in unpredicted loss in productivity. Hence the faults on bearing need to be rectified as early as possible. In this paper four conditions of a DC motor namely good condition, defect on inner of race, defect on outer of race and defect on both inner and outer of race are obtained and subjected to classification using statistical features after a preprocessing operation for denoising. The denoising algorithm employed for preprocessing is Overlapping Group Shrinkage (OGS) and SVM is the classifier used. The accuracy in classification is found to be more when statistical features of denoised signal are fed as inputs to the classifier. Later, a vibration signal modeling system and its denoising is studied.

Keywords : machine fault diagnosis; vibration signal analysis; group sparse denoising; overlapping group shrinkage; SVM

1. INTRODUCTION

Rotating machinery plays a vital role in plants. The machines break down as a result of faults on inner race, rolling element bearing and outer race. The faults need to be diagnosed at an early stage for smooth and safe running of industry and also for preventing further economic losses and catastrophic damages. Information about the condition of sub components of machinery such as gears, bearings, shafts, couplings, engines, electric generators, pumps, fans etc can be extracted from the vibration signals measured at the external surfaces.

Vibration measurement and analysis can be extended to machine monitoring, a process intended to check the wear and tear of moving parts. This process gives idea about machine's current status [7]. Condition monitoring does analysis based on vibration and sound signals since it contain a lot of information about the internal components of the machine. The machines generate vibrations even under good condition, but the frequencies of these differ as compared to that generated under faulty conditions. This idea is useful for diagnosing faults.

Many researches have been done in modeling the vibration response and fault signals. Nowadays a lot of innovatives are coming up in the field of fault diagnosis. From eighteenth century onwards, waveforms in time domain were used for vibration signal analysis. The information obtained from frequency domain failed in analysis when the different components in the signal had a unique characteristic frequency [4]. The energy of vibration starts increasing at a particular frequency as and when the defect occurs, which couldn't be identified with the help of a waveform in frequency domain. Later in 1946, Gabor developed Short time Fourier Transform (STFT), which applies transform on the stationary segments obtained from the time domain signal. The method used a window of fixed width, a main drawback.

Good feature selection is very important in the area of pattern identification. Principal component analysis (PCA) is one predominantly used technique. Here, the data are transformed to a new coordinate system so that the

* Centre for Computational Engineering and Networking(CEN), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India. archanachandran16@gmail.com

first coordinate has the greatest variance and the consecutive coordinates will be having variance in a descending order. The reduction in dimensionality can be achieved by PCA, which keeps low-order components having most important aspects of data and the higher-order components are not taken into consideration. The drawback of PCA is it does not allow data to be used as such [11]. The original data and the transformed one vary in different aspects. Hence, other techniques that allow higher-order feature selection are preferred. A new method empirical mode decomposition (EMD), decomposes a complicated signal into intrinsic mode functions (IMFs), based on time–frequency analysis has been proposed in [12]. A SVR (Support vector regression) based model is proposed in [13] with which a bearing's life span is predicted. Artificial intelligence based classification techniques could bring about good accuracy but it suffers from defects like large time consumption for training data and huge cost as a result of computational complexities [11]. De-noising based on wavelet is a widely used technique to process non-stationary vibration signals as it could extract information apt for diagnosis and also enhance the impulsive components [14]. For identification of faults, visual inspection can prove its worth, but diagnosis is something beyond. It needs a paramount knowledge in the respective domain plus a spontaneous response. The algorithms in machine learning could solve these diagnostic problems using vibration signals [8].

Here, a bearing fault diagnosis system is proposed using Overlapping Group Shrinkage (OGS) as an initial preprocessing. OGS is an algorithm developed by Po-Yu Chen et.al., for denoising of group sparse signals [1], [2]. Denoising the dataset before feature extraction can produce an optimum accuracy. Later statistical features extracted from the denoised signals are used for classification. The statistical feature extraction technique yields an optimum computational cost and training time. The classification technique used here is Support Vector Machine (SVM).

2. PROPOSED METHOD

This session explains the proposed bearing fault diagnosis system. Figure 1 shows the block diagram of the bearing defect diagnosis system. The system involves generation of the fault signal from the simulator, which is then added with Gaussian noise ranging from 5dB to 20dB. Then the noisy signal is undergone preprocessing by means of OGS denoising algorithm. Further statistical features extracted from the denoised signal are given to SVM classifier, which produces an optimum accuracy.

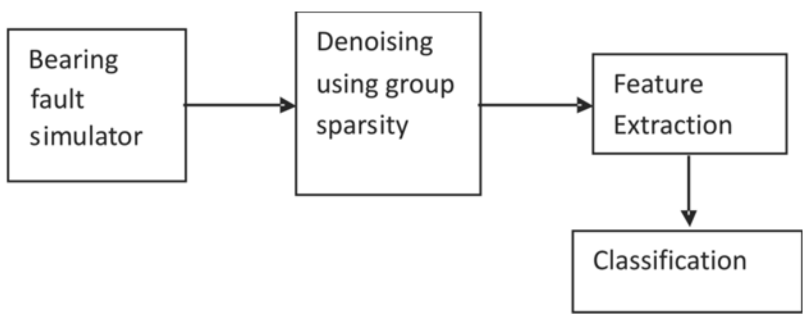

Fig. 1. Block diagram of the proposed bearing fault diagnosis system.

A. Bearing Fault Simulation

A machine with a DC motor rotating at a speed ranging from 0 to 3000 rpm, a 20 mm diameter motor shaft coupled to bearings together constitute a simulator setup [5]. The DC motor is controlled by an ON/OFF switch provided in the control panel and the speed is adjusted by means of a potentiometer. A control panel is provided in the setup for controlling and displaying various parameters. The selector is used to vary the speed of motor and gear as desired. The panel also consists of an ammeter which gives information about the current drawn by the motor, which depends on the load. Here four SKF30206 bearings are used. In this, one is a new bearing free from defects. The other three bearings had defects made from electrical discharge machining (EDM) [6].

Data generated belongs to four different types, type 1-bearing good condition (GOOD), type 2 - bearing having fault in inner race (IRF), type 3-bearing having fault in outer race (ORF), type 4-bearing having fault at both inner and outer race (IORF). Type 2, 3 and 4 are faulty data.

B. Preprocessing Operation- Denoising

The sparse signals in noise can be estimated by using optimization of convex type along with a convex regularizer which promotes sparsity [1]. A non-convex regularization term can introduce more sparsity without altering the convexity of overall cost function, hence maintaining the pros of convex optimization such as unique minimum and robustness. This idea is exploited in OGS, used for denoising of signals which are sparse. OGS is a type of FOCUSS algorithm used to obtain solutions of sparse type for linear equations containing design matrix [1]. The method requires that the convex function gets minimized. The algorithm uses less memory, minimum data indexing, good convergence behaviour.

Denoising is the estimation of clean signal $c(i)$, $i \in \mathrm{I}$ from observations of noisy type $n(i)$,

$$n(i) = c(i) + a(i),\ i \in \mathrm{I} \tag{1}$$

the signal $c(i)$ has property of group sparse and $a(i)$ is Gaussian noise. The c domain is given by I = {0, ... , N – 1}. The optimization problem is formulated as,

$$\mathrm{G}(c) = \frac{1}{2}\| n - c \|_2^2 + \lambda \mathrm{P}(c) \tag{2}$$

In the problem, c can either be wavelet coefficients or STFT coefficients of a signal or the signal of sparse itself. The STFT coefficients exhibit a strange behavior of grouping which is exploited in OGS algorithm so as to improve the denoising. The penalty function or regularizer, P(c), increases the sparse behavior of signal and is of form

$$\mathrm{P}(c) = \sum_{i \in l} \left[\sum_{\substack{i \in l \\ J}} | c(i + j) |^2 \right]^{\frac{1}{2}} \tag{3}$$

where the group is defined by set .

This penalty function in (3) can be rewritten as,

$$\mathrm{P}(c) = \sum_i | c_i , \mathrm{K} \|_2 \tag{4}$$

Where K is the group size and i represents the number of groups. Here, 2D denoising is used thus the group sizes are K1 and K2, where K1 is number of spectral samples and K2 is number of temporal samples.

The cost function in (2) can be rewritten as,

$$\mathrm{G}(c) = \frac{1}{2}\| n - c \|_2^2 + \lambda \sum_i \| c_{i,\mathrm{K}} \|_2 \tag{5}$$

Now to induce more sparsity, the problem defined in (5) has been reframed by introducing a non-convex penalty function as,

$$\mathrm{G}(c) = \frac{1}{2}\| n - c \|_2^2 + \lambda \sum_{i \in \mathrm{I}} \varphi\left(\|c_{i,\mathrm{K}}\|_2 ; a\right) \tag{6}$$

The value of a in (6) needs to be restricted for making the objective function convex. The solution for these objective functions are obtained by using the method of majorization-minimization [1], [2]. In this paper we have used two penalty functions namely (i) absolute value function (abs) or the l_1 norm, (ii) arctangent penalty function (atan). In OGS-abs, $\varphi(t) = |t|,\ t \in \mathrm{R}$, then the objective function is reduced as defined in (3) and which ensures the convex regularization. In OGS-atan,ensures the non-convex regularization. The signal's amplitude is fairly preserved in OGS-atan than in OGS-abs [2].

$$\varphi(t) = \frac{2}{a\sqrt{3}}\left(\tan^{-2}\left(\frac{1+2at}{\sqrt{3}}\right)-\frac{\pi}{6}\right), \forall t > 0 \quad (7)$$

C. Statistical Feature Extraction

Standard deviation (d) gives a measure of how much each data item in a group differ from its mean. Thus less d implies the points are very close to mean and high d implies data are distributed over wider range.

$$d = \sqrt{\sum_{n=1}^{m}\frac{(a_n-\bar{a})^2}{m-1}} \quad (8)$$

Variance (v) gives a numerical measure of data scatter and is square of standard deviation. It can be termed as the average calculated for the squared differences from the mean.

$$v = \sum_{n=1}^{m}\frac{(a_n-\bar{a})^2}{m-1} \quad (9)$$

Mean ( ) also called expected value is defined as the central value of a number set.

$$\bar{a} = \frac{1}{m}\sum_{n=1}^{m} a_n \quad (10)$$

Kurtosis (k) gives an idea about the peakness of a distribution. Good bearings have a low value for kurtosis, whereas the value becomes high for bearings affected with faults. High peak distributions are called leptokurtic and low peak distributions are called platykurtic and normal distribution is called mesokurtic.

$$k = \left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^4\right\}-\frac{3(n-1)^2}{(n-1)(n-3)} \quad (11)$$

Sum means sum of all signal values. Range is obtained by subtracting signal's minimum value from its maximum value. Maximum value means highest value in the signal. Minimum Value means the lowest value present in the signal.

D. Classification

Support Vector Machine (SVM) is used extensively in the field of fault diagnosis problems since it yielded good and desirable success rates[10]. SVM is a supervised learning method which makes use of maximum margin concept in which a linear separator (hyperplane) separates the data in different classes. The algorithm does this by finding maximum separation between data in different classes. The features extracted from vibration signal are in general not linearly seperable and hence the feature space need to be mapped nonlinearly to the output space , which is possible in SVM by means of a non linear kernel [6].

The equation of linear classifier is given by

$$w^{T}x - \gamma = 0 \quad (12)$$

The 2 class linear SVM formulation is given as

$$\min_{w,\gamma}\frac{1}{2}(w^{T}w) \quad \text{subject to} \quad D(Aw - e\gamma) > e \quad (13)$$

where $W \in R^n$, A is $m \times n$ data matrix belonging to R^n and D is $m \times n$ diagonal matrix and e is an $m \times 1$ vector of ones.

In case of nonlinear data, projection onto high dimension space is done with mapping function, ϕ.

$$\phi : R^n \rightarrow R^k \tag{14}$$

$\phi(x) \in R^k$ and $k >> n$.The kernel function $k(x_i, x_j)$is used which obeys the condition,

$$k(x_i,\ x_j) = <\phi(x_i).\ \phi(x_j)> \tag{15}$$

[17], [18], [9], [15]. The 2 class SVM is to be extended to multiclass classification. The package LIBSVM [16] does it through M, one vs rest classifiers, where M is the number of classes. After training the algorithm, testing can be done with a new dataset, the prediction is done by SVM. SVM could produce high accuracies on vibration data. An optimal SVM architecture requires that the dimensionality of input is reduced drastically. Here, we extract the statistical features of the denoised signal like mean, variance, standard deviation, kurtosis, sum, range, minimum and maximum value and achieve the reduction in dimensionality.

3. RESULTS AND DISCUSSION

The vibration data set contains 400 signals of length 8190 of which the first hundred belong to type1, *i.e,* good signal, which is free from all kind of faults, next hundred belong to type2, which is resulted by fault on inner race, next hundred belong to type3, resulted by fault on outer race and the last hundred of type 4 arises as a result of fault on both inner and outer race.

All the types of signals namely type1, type 2, type 3 and type 4 make use of group sizes K1 = 1 and K2 = 2, and λ of 10 for OGS-abs and λ of 15 for OGS-atan. Figure 2, 3, 4 and 5 shows denoising on type 1, 2, 3 and 4 signals.

The classification accuracies are summerised in tables I to IV for the 4 types of signals with different amount of noise added ranging from 5dB to 20 dB. Training and testing for different values of the cost and gamma is done in SVM. Classification is tried for C-SVM and nu-SVM using linear and polynomial kernel types. Here the classification is done in 2 parts. First, the statistical features of the noisy signal is taken as such and classified, further statistical features of the denoised signal is taken and given for classification. From table 1, for an amount of 5dB noise, the SVM accuracy is improved noticeably for nu-SVM with polynomial kernel type. From table 2, we can note that 10dB denoising, SVM accuracy is more for denoising by OGS-abs for C-SVM and linear kernel type. C-SVM and linear kernel type also gives high accuracy for 15 dB denoising as can be seen from table 3.nu-SVM and C-SVM with polynomial kernel type gives high accuracy for 20dB denoising.

4. DENOISING ON A VIBRATION SIGNAL MODEL OF ROLLING BEARINGS

We are considering mainly three types of fault signals here namely IRF, ORF and IORF signals.

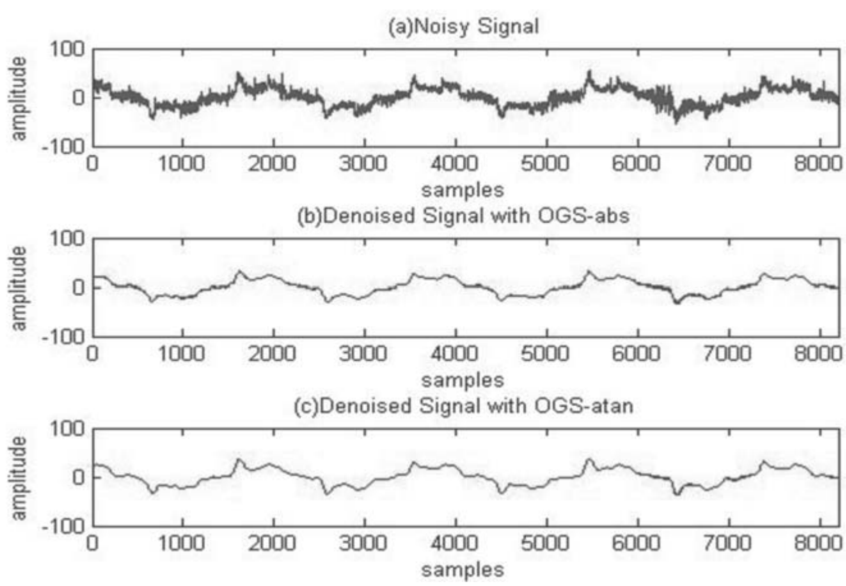

Fig. 2. Type1 signal (GOOD) (*a*) Noisy Signal (*b*) Denoised Signal with OGS-abs (*c*) Denoised Signal with OGS-atan

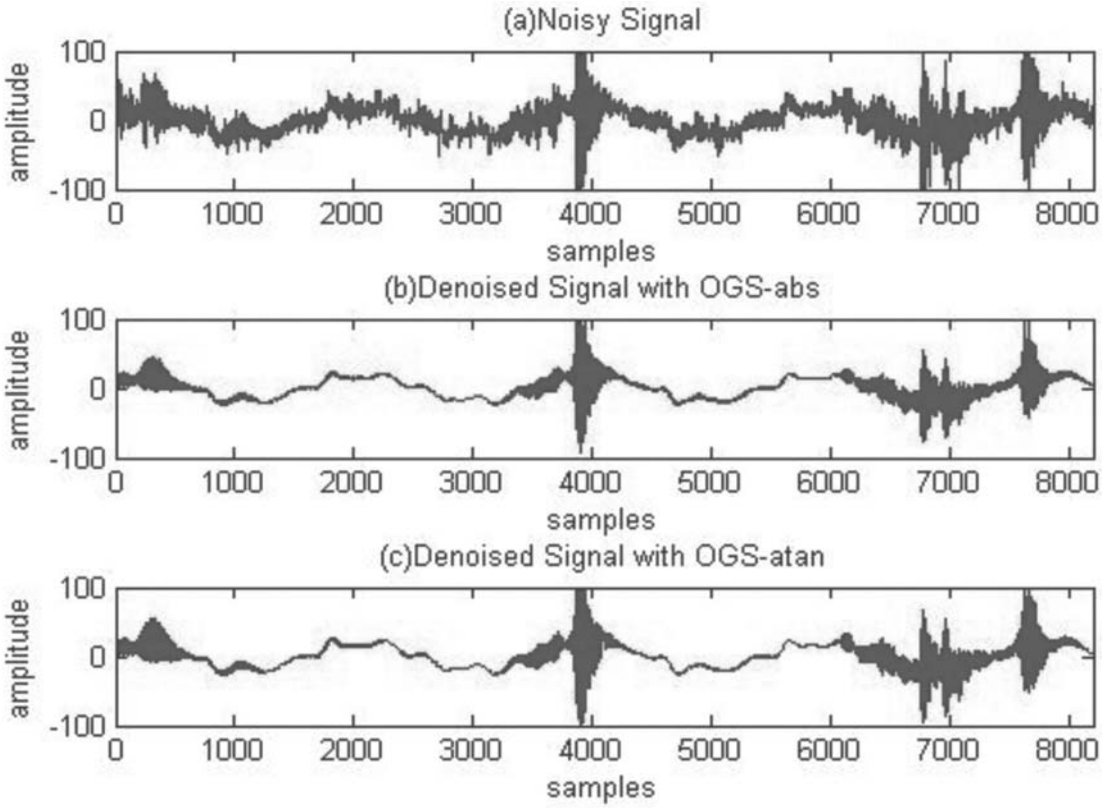

Fig. 3. Type2 signal (IRF) (*a*) Noisy signal (*b*) Denoised signal with OGS-abs (*c*) Denoised signal with OGS-atan.

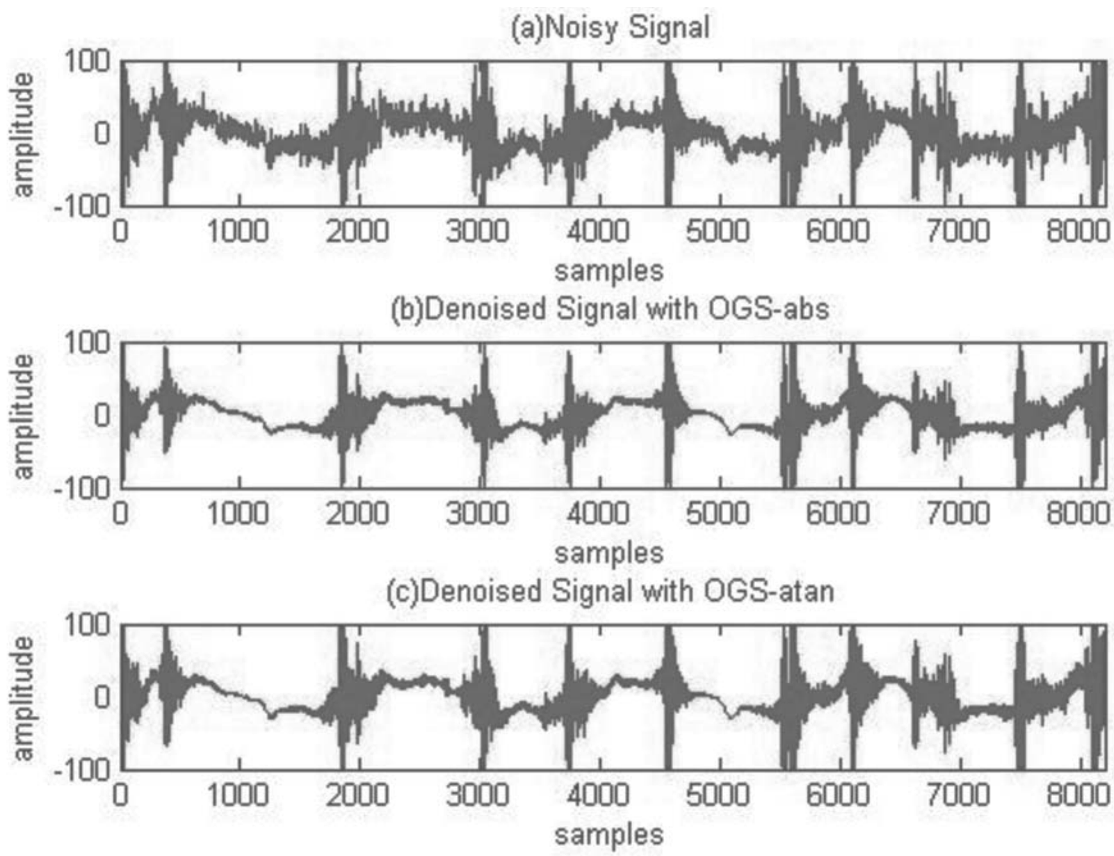

Fig . 4. Type 3 signal (ORF) (*a*) Noisy signal (*b*) Denoised signal with OGS-abs (*c*) Denoised signal with OGS-atan.

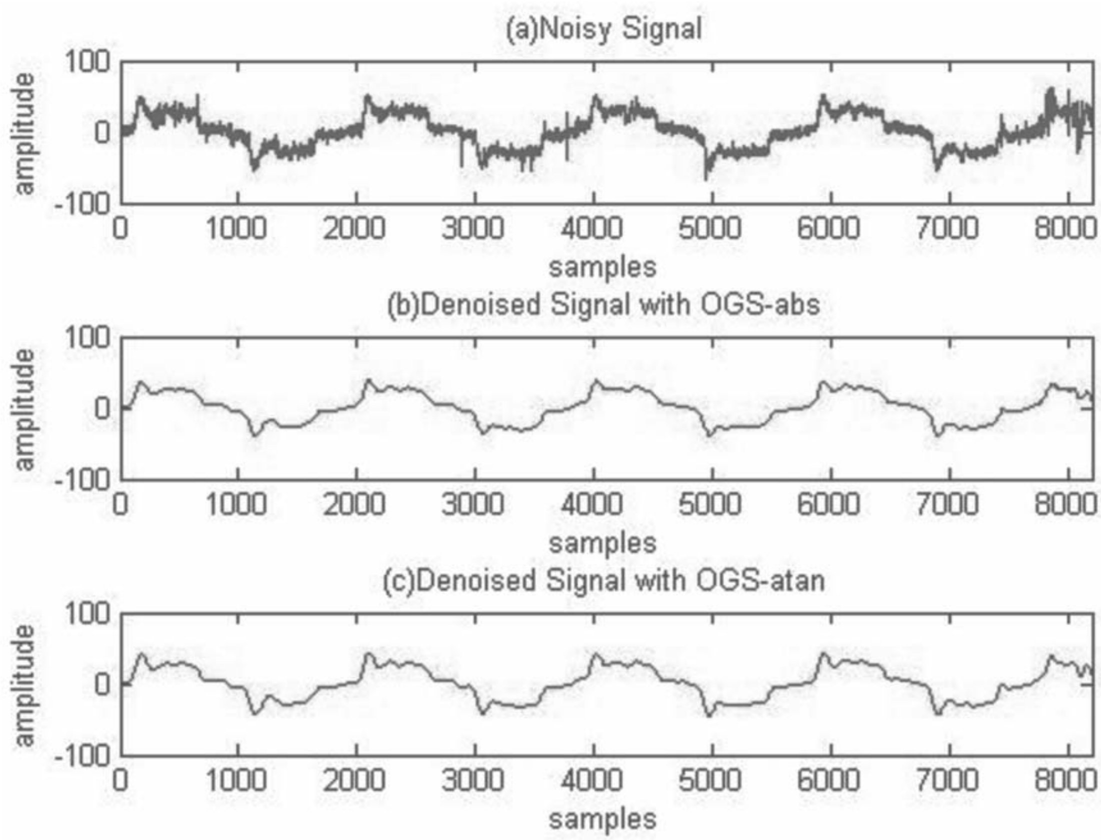

Fig. 5. Type 4 signal(IORF) (*a*) Noisy signal (*b*) Denoised signal with OGS-abs (c) Denoised signal with OGS-atan.

Table 1. Classification accuracy for 5dB denoising.

SVM Type	*Kernel type*	*Noisy signal(%)*	*Denoised signal – OGS-abs(%)*	*Denoised signal – OGS-atan(%)*
C-SVM	linear	71	74	76
C-SVM	polynomial	65	66.5	73
nu-SVM	linear	86.5	89.5	87.5
nu-SVM	polynomial	66	72.5	73.5

Table 2. Classification accuracy for 10dB denoising.

SVM Type	*Kernel type*	*Noisy signal(%)*	*Denoised signal – OGS-abs(%)*	*Denoised signal – OGS-atan(%)*
C-SVM	linear	74.5	86	77.5
C-SVM	polynomial	66.5	67	68.5
nu-SVM	linear	86	87	89
nu-SVM	polynomial	64	65	64.5

A. Denoising on ORF signal

The ORF simulation signal is given by [19]

$$y(t) = \sum_{j=1}^{N} \phi(wa\cos(\psi_w) + \text{we}\,\omega^2\cos(2\pi f_r t + \psi_e)).\, d(t - j\,T_m - \tau_j) \qquad (16)$$

Table 3. Classification accuracy for 15dB denoising

SVM Type	*Kernel type*	*Noisy signal(%)*	*Denoised signal – OGS-abs(%)*	*Denoised signal – OGS-atan(%)*
C-SVM	linear	77	92	89
C-SVM	polynomial	64	69.5	69
nu-SVM	linear	86	88	89
nu-SVM	polynomial	65.5	71.5	72

Table 4. Classification accuracy for 20dB denoising

SVM Type	*Kernel type*	*Noisy signal(%)*	*Denoised signal – OGS-abs(%)*	*Denoised signal – OGS-atan(%)*
C-SVM	linear	83.5	84.5	84.5
C-SVM	polynomial	62.5	73	77
nu-SVM	linear	88	89.5	89.5
nu-SVM	polynomial	55.5	66.5	68

where T_m denotes impulse period, τ_j is minor fluctuation

occurring randomly around average period , is factor between amplitude and load, is rotational frequency of bearing, is bearing angular velocity, is rotor mass, is gravity acceleration and is eccentricity, is the angle between gravity direction and outer race defect position, is the angle between mass eccentric position and outer race defect position and is an exponentially decaying impulse oscillation.

In fig 6, (*a*) shows the clean ORF signal, is made noisy by adding a Gaussian random noise of 10dB as in (*b*) and then passed onto OGS-abs and OGS-atan denoising algorithms. The denoised signal is in (*c*). The appropriate group size for the ORF signal in STFT domain is K1 = 7 and K2 = 8. The corresponding taken here is 1.9.

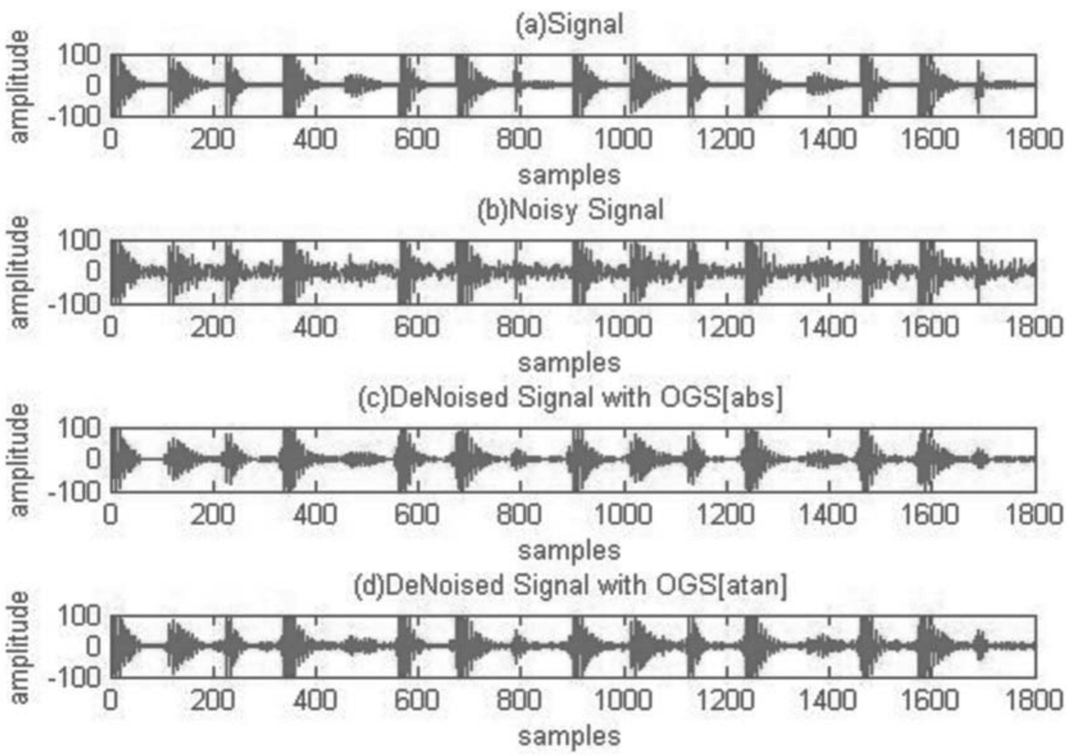

Fig. 6. Denoising on ORF signal.

B. Denoising on IRF signal

The IRF simulation signal is given by [19]

$$y(t) = \sum_{j=1}^{N} \phi(wa\cos(2\pi f_r t + \psi_w)) + we\,\omega^2 \cos(\psi_e)).d(t - j\,T_m - \tau_j) \tag{17}$$

In fig 7, the fault signal on inner race shown in (*a*) is added a Gaussian random noise of 10dB, the noisy signal (*b*) is then passed onto OGS-abs and OGS-atan denoising algorithms. The denoised signal is shown in (*c*). It is observed that the appropriate group size for the signal in STFT domain is K1 = 20 and K2 = 10 . The corresponding taken here is 1.75.

C. Denoising on IORF signal

The IORF signal is given by [19] $\sum_{j=1}^{N} \phi(wa\cos(2\pi f_r t + \psi_w)) + we\,\omega^2 \cos(\psi_e)).d(t - j\,T_m - \tau_j)$ (18)

where or cage frequency is a fundamental frequency for bearing with fixed outer race and is given by

$$f_c = \frac{f_r}{2}(1 - \frac{d}{B}\cos(\beta)) \tag{19}$$

f_r is shaft rotation frequency in hertz, *d* is ball D diameter, is diameter of the pitch and β is angle of contact. In fig 8, the fault signal on inner race (*a*) is added a Gaussian random noise of 10db, the noisy signal (*b*) and then passed onto OGS-abs and OGS-atan denoising algorithms. The denoised signal is in (*c*). The group size K1 = 20 and K2 = 10 is desired for IRF signal. The corresponding λ taken here is 1.75.

5. CONCLUSION

A bearing fault diagnosis system is proposed here. Vibration signals namely good, fault on inner race, on outer race, on both inner and outer race are collected from a bearing fault simulator and corrupted by various amount of noises. These signals are then denoised using a group sparsity inducing algorithm called OGS. The statistical features of the denoised signals are extracted and given to SVM classifier and the performance is evaluated. From the results, we can infer that the SVM accuracy of classification gets improved for denoised signals when compared to the noisy signals. Thus it can be concluded that the proposed bearing fault diagnosis system with OGS preprocessing step is appropriate for machine condition monitoring with good accuracy.

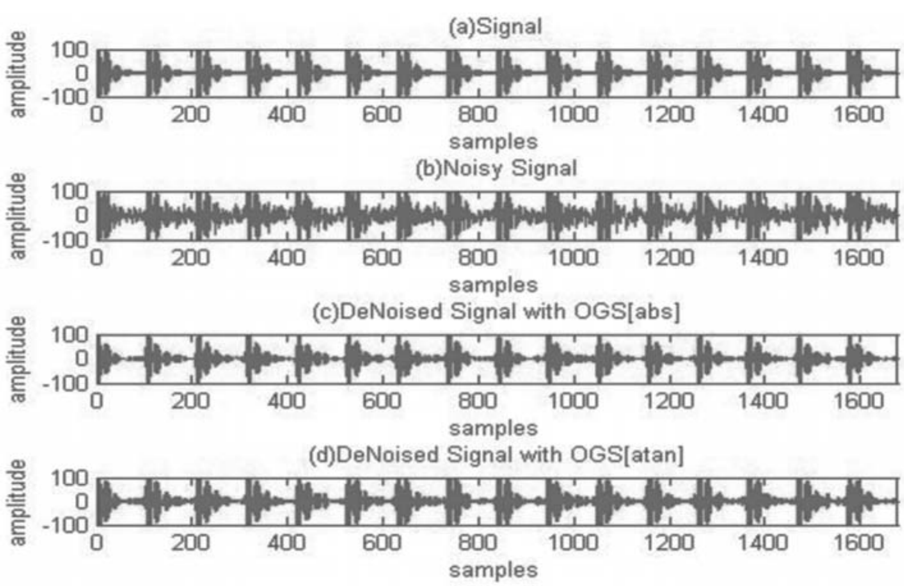

Fig. 7. Denoising on IRF signal.

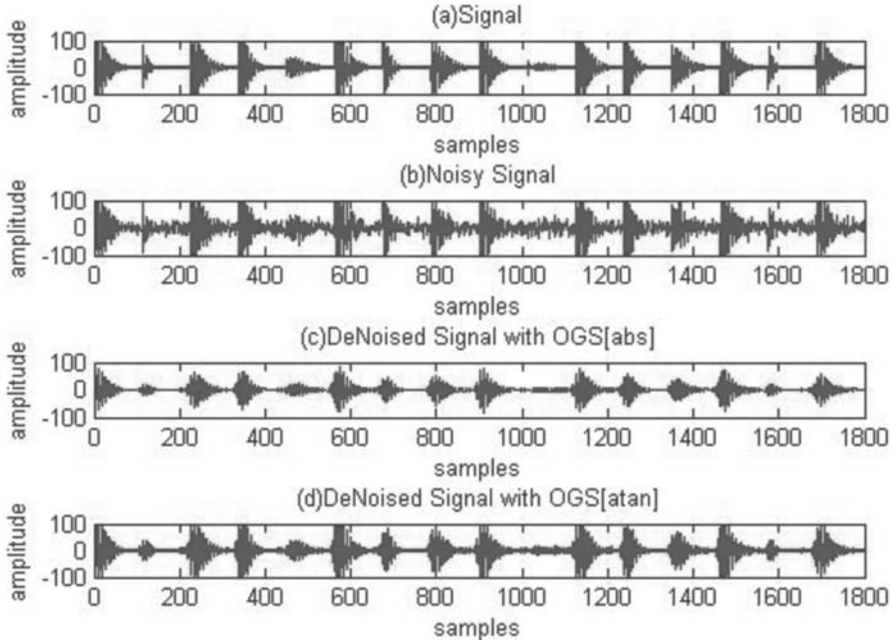

Fig. 8. Denoising on IORF signal.

6. REFERENCES

1. Chen, Po-Yu, and Ivan W. Selesnick. "Translation-Invariant Shrinkage/Thresholding of Group Sparse Signals." arXiv preprint arXiv: 1304.0035 (2013).
2. Chen, Po-Yu, and Ivan W. Selesnick. "Group-sparse signal denoising: Non-convex regularization, convex optimization." Signal Processing, IEEE Transactions on 62.13 (2014): 3464-3478.
3. Li, R. Zhou, Q. Hu, X. Liu ,Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine Mech. Syst. Signal Process., 28 (2012), pp. 608–621.
4. Neethu Mohan, Ambika.P.S, Sachin Kumar.S, "Multicomponent fault diagnosis using statistical features and regularized least squares." IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS) 2015.
5. SachinKumar S, Neethu Mohan, Prabaharan Poornachandran. "Condition Monitoring in Roller Bearings using Cyclostationary Features." Proceedings of the Third International Symposium on Women in Computing and Informatics. ACM, 2015.
6. Sugumaran, V., V. Muralidharan, and K. I. Ramachandran. "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing." Mechanical systems and signal processing 21.2 (2007): 930-942.
7. Cong, Feiyun, et al. "Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis." Journal of Sound and Vibration 332.8 (2013): 2081-2097.[8] Randall, R. B. "Vibration signature analysis-Techniques and instrument systems." Noise Control and Vibration Reduction, vol. 6, Mar. 1975, p. 81-89.. Vol. 6.1975.
8. Sugumaran, V., and K. I. Ramachandran. "Effect of number of features on classification of roller bearing faults using SVM and PSVM." Expert Systems with Applications 38.4 (2011): 4088-4096.
9. Soman, K. P., R. Loganathan, and V. Ajay. Machine learning with SVM and other kernel methods. PHI Learning Pvt. Ltd., 2009.
10. Suykens, J. A. K., Van Gestel, T., Vandewalle, J., & De Moor, B. (2003). A support vector machine formulation to PCA analysis and its Kernel version, ESAT-SCD-SISTA Technical Report.
11. N.E. Huang, Z. Shen, S. R. Long, et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", in: Proceedings of the Royal Society, London, vol. 454, 1998, pp. 903–995.
12. V.K. Rai, A.R. Mohanty , "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform",Mech. Syst. Signal Process., 21 (8) (2007), pp. 3030–3041.

13. C. Sun, Z.S. Zhang, Z.J. He, “Research on bearing life prediction based on support vector machine and its application”, Journal of Physics: Conference Series 305 (2011) 012028.

14. Kumar H S, Dr. Srinivasa Pai P, Dr. Sriram N S, Vijay G S, “ANN based evaluation of performance of wavelet transform for condition monitoring of rolling element bearing”, IConDM.

15. Nikhila Haridas, V.Sowmya and K. P. Soman, “GURLS vs LIBSVM: Performance Comparison of Kernel Methods for Hyperspectral Image Classification”, Indian Journal of Science and Technology, Vol 8(24), September 2015.

16. Chang C-C, Lin C-J. LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology (TIST). 2011; 2(3):27.

17. C. Cortes and V. Vapnik, Support-vector networks. Machine Learning, 20:273-297, November 1995.

18. T. Joachims. A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. In International Conference on Machine Learning (ICML), 1997.

19. F. Cong, J. Chen, G. Dong, et al, Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis, J. Sound Vib. 332 (8) (2013) 2081-2097.