A ROBUST FEATURE EXTRACTION FOR AUTOMATIC FAULT DIAGNOSIS OF ROLLING BEARINGS USING VIBRATION SIGNALS

Mehrdad Heydarzadeh, Alireza Mohammadi

1 Department of Electrical and Computer Engineering, 2 Departments of Mechanical Engineering & Bioengineering
The University of Texas at Dallas, Richardson, Texas 75080
Email: {mehhey, alireza.mohammadi}@utdallas.edu

ABSTRACT

Bearing faults are one of the main reasons for rotary machine failure. Monitoring bearing vibration signals is an effective method for diagnosing faults and preventing catastrophic failures in rotary mechanisms. The state-of-the-art vibration monitoring algorithms are mainly based on frequency or time-frequency domain analysis of rotary machines that are operating in steady state. However, the steady state assumption is not valid in applications where the loads and speeds are time-varying. Finding a method for capturing the variability in vibration signals, which are caused by varying loads and speeds, is still an open research problem with potentially many applications in emerging areas such as electric vehicles. In this paper, we address the problem of vibration signal monitoring by applying a feature extraction algorithm to rotary machine signals measured by accelerometers. The proposed method, which is based on the wavelet scattering transform, achieves overall high accuracy while being computationally affordable for real-time implementation purposes. In order to verify the effectiveness of the proposed methodology, we apply our technique to a well-known vibration benchmark dataset with variable load. Our algorithm can diagnose various faults with different intensities with an average accuracy of 99% and thus effectively outperforming all prior reported work on this dataset.

INTRODUCTION

MOTIVATION

A rolling bearing carries load by placing rolling elements such as balls or rollers between two adjacent bearing rings, which are also known as races. The relative motion of the races causes the rolling element to roll with low friction. Due to the low friction, rolling bearings are among the most frequently used components in rotary machines. On the other hand, a fault in bearings can result in catastrophic rotary machine failures. Accordingly, real-time monitoring of bearings for safety critical applications such as transportation and aviation is of crucial importance.

PRIOR WORK

Prior fault diagnosis algorithms fall within two broad categories, namely, model-based and data-driven approaches. Model-based fault diagnosis methods, when there exists a precise knowledge of the underlying governing dynamics of the system, generate reliable results [1–3]. However, accurate physical modeling of rotary machines is a very challenging and demanding task. Furthermore, model-based fault diagnosis algorithms are difficult to generalize to systems with different models [4, 5].

On the other hand, data-driven fault diagnosis methods are considered as pattern recognition problems. In the data-driven approaches, signal processing techniques are applied to the measured vibration signals in order to extract important system features. The health state of the machine is then diagnosed by classifying the extracted features [6, 7].

The key difference between prior data-driven methods is
related to their underlying feature extraction techniques. The purpose of feature extraction is to build a compact representation of the sensory data that is sensitive to fault while insensitive to other factors such as noise and other unwanted variabilities in measured vibration. The vibration data collected from a machine in steady state, constant load and speed, has a semi-random behavior. This behavior becomes more quasi-periodic when a fault happens [8]. Under steady state conditions, using the frequency domain feature extraction methods are suggested since the vibration signal has a sparse representation in the frequency domain, which varies across different faults. The problem of rotary machine fault diagnosis in steady state conditions using frequency domain and time-frequency domain techniques is very well studied. In this context, researchers have proposed numerous methods such as Fourier transforms [9, 10], wavelet transforms for feature extraction [3, 11], and empirical mode decomposition [12, 13], to name a few. However, there are certain safety-critical applications in which the assumption of stationary signal is not valid anymore. Milling machines, compressors, pumps, and electric vehicles are among the applications where this assumption does not hold.

Whenever the steady state assumption is not valid, the vibration signal is distorted by changes in load and speed conditions, which manifest as scattered points in the feature space used for fault diagnosis. When the vibration signal variability is high, the clouds of points that belong to two rather different classes will overlap resulting in decreased classification accuracy and even fault misclassification. Hence, for fault diagnosis under unsteady conditions, new methods are required to capture the vibration signal variabilities. This topic has recently become a mainstream in research due to emerging applications of electric vehicles. Authors in [14] used the Gabor wavelet to model the transient trends in an induction motor application. This work is mostly focused on electric machines and its current measurement. Moreover, Gabor wavelet is used for modeling purposes and a feature extraction method is not introduced for automatic fault diagnosis yet. Complex wavelet analysis is employed in [15] to extract features of vibration signal which are robust against load conditions. Complex wavelet transform has been proven to be useful for capturing signal time shifts. However, it has not been extended to other deformations in the measured signal. A scale invariant method using wavelet transform is proposed in [16]. This work captures non-stationary trends but only considers the start-up transitions and does not address the non-stationarity in measurement caused by variable loads.

KEY CONTRIBUTION
In this paper, we propose an automatic fault diagnosis method for monitoring bearing status using vibration signal measured by accelerometers. We address the problem of variable load and unsteady measurements from a feature extraction perspective. Specifically, we use the wavelet scattering transform for this purpose which is a robust representation against deformations caused by a change in load conditions. The proposed method has an affordable computational cost for real-time implementation and achieves overall high accuracy over many of its well-known fault diagnosis counterparts such as the global spectrum [17], vibration spectrum imaging [18], pool of time and frequency features [19], and statistical wavelet packet features [20], as highlighted in the Experimental Results section.

The rest of the paper is organized as follows. First, we will present the necessary mathematical framework. Next, we present our fault diagnosis methodology. Subsequently, we provide experimental results that are performed on a benchmark dataset. Finally, we conclude the paper with some remarks and future research directions.

MATHEMATICAL FRAMEWORK
DEFORMATION AND NOTION OF STABILITY
In the steady state analysis, the frequency domain is a useful tool to extract features from measured signals. Let \( \hat{x}(\omega) = \int x(t)e^{-j\omega t} dt \) denote the Fourier integral of a signal \( x(t) \in L_2^R \). If \( x_c(t) = x(t - c) \) is a deformed version of \( x(t) \) by translation, then \( |\hat{x}_c(\omega)| = |\hat{x}(\omega)| \). So the modulus of the Fourier transform is a representation of signals which is invariant against translation. In practice, translation is not the only possible deformation which causes a variability. Time-warping is a more general class of deformation which is modeled as \( x_r(t) = x(t - \tau(t)) \) with \( |\tau'(t)| < 1 \). Time warp is a broad model for deformation used in several signal processing applications such as speech processing [21]. Here, the deformation can be quantified using \( |\tau'(t)| \). If the derivative is zero the deformation simplifies to a pure translation. Suppose \( \Phi(x) \) is an arbitrary representation (feature) of \( x(t) \) and \( \Phi(x_c) \) is the same representation for \( x_c(t) \), i.e., the deformed version of \( x(t) \); then, \( \Phi(\cdot) \) is considered as a robust representation if the difference between \( \Phi(x) \) and \( \Phi(x_c) \) caused by the deformation is small. This similarity can be quantified using the Euclidean norm as \( d(x, x_c) = ||\Phi(x) - \Phi(x_c)|| \). Indeed, we are looking for a function \( \Phi(\cdot) \) that is invariant against this signal deformation. This invariance can be formalized using the notion of stability. The representation \( \Phi(\cdot) \) is stable if a small change in \( x(t) \) does not lead to a big change in the distance between the two representations. Mathematically, representation stability is defined as a Lipschitz continuity condition with respect to the norm \( d(\cdot) \). We say that the representation \( \Phi(\cdot) \) is stable, if there exists a constant \( C > 0 \) such that for all \( \tau \) with \( \sup_{t} |\tau'(t)| < 1 \):

\[
\|\Phi(x) - \Phi(x_c)\| \leq C \sup_{t} |\tau'(t)| \|x\| \tag{1}
\]

where \( \sup \) denotes supremum of a set. The constant \( C \) gives a measure of stability. Indeed, if the Lipschitz condition (1) is true,
a change in $x(t)$ caused by time-warp leads to a linear change in its representation. Indeed, time-warping is locally linearized by $\Phi(x(t))$, and thus $\Phi(x) - \Phi(x_t)$ can be approximated by a linear operator if $\sup|\tau'(t)|$ is small. In other words, in the feature space $\Phi(x)$ and $\Phi(x_t)$ lie on the same hyperplane and the corresponding features do not spread all over the feature space.

It can be shown that the Fourier modulus is not stable with respect to time-warp deformation. In the Fourier representation, time-warp operator changes each frequency component differently, (i.e., the high frequency signal contents are perturbed more than the low frequency signal contents [21]). The notation of stability helps us to understand why feature extraction methods based on steady state analysis do not work for variable load conditions. Indeed, a deformation caused by a change in load spreads feature points corresponding to the deformed signals in the feature space in a random manner and increases the overlapping of classes. Hence, we need to look for a feature extraction method such that the feature point corresponding to a time-warped signal lies on a hyperplane rather than being spread across the feature space.

**ANALYTIC WAVELET TRANSFORM MODULUS**

The analytic wavelet transform, which is robust against shift deformation, can be calculated using constant Q filterbanks [22]. A wavelet like $\psi(t)$ is a band pass filter where $\tilde{\psi}(0) = 0$. In the analytical wavelet transform $\tilde{\psi}(\omega) \approx 0$ for $\omega < 0$ [23]. A dilated version of $\psi(t)$ with the central frequency of $\lambda > 0$ can be written as $\psi_\lambda(t) = \lambda \psi(\lambda t)$ or in frequency domain $\tilde{\psi}_\lambda(\omega) = \tilde{\psi}\left(\frac{\omega}{\lambda}\right)$, where the central frequency of $\psi(\omega)$ is normalized to 1 and $Q$ is chosen as the number of wavelets per octave, i.e., $\lambda = 2^{\frac{j}{Q}}$ for $k \in \mathbb{Z}$. This guarantees that the bandwidth of $\psi_\lambda$ is in order of $Q^{-1}$ while its central frequency is at $\lambda$. Accordingly, different $\psi_\lambda$’s cover the frequency spectrum except the DC component which is covered using a low-pass filter $\phi$. Let $\Lambda$ denote the set of all values of $\lambda$. Then, the wavelet transform of signal $x(t)$ can be calculated using the convolution of the two filters:

$$Wx = (x(t) \ast \phi(t), x \ast \psi_\lambda(t)) \quad t \in \mathbb{R}, \lambda \in \Lambda$$

(2)

Since $t$ is not critically sampled, this signal representation is redundant with respect to the wavelet basis. The filters $\phi$ and $\psi$ need to be designed in a way that the entire frequency axis is covered. Therefore,

$$A(\omega) = |\tilde{\phi}(\omega)|^2 + \frac{1}{2} \sum_{\lambda \in \Lambda} (|\tilde{\psi}_\lambda(\omega)|^2 + |\tilde{\psi}_\lambda(-\omega)|^2)$$

(3)

must hold for all $\omega \in \mathbb{R}$, where $A(\cdot)$ satisfies [24]:

$$1 - \alpha \leq A(\omega) \leq 1 \quad \text{for} \quad \alpha \leq 1$$

(4)

Multiplying both sides of this inequality by $|\tilde{x}(\omega)|^2$ and apply Plancherel theorem, we obtain [25]:

$$(1 - \alpha)\|x\|^2 \leq \|Wx\|^2 \leq \|x\|^2$$

(5)

where $\|Wx\|^2 = \int |x \ast \tilde{\psi}(\omega)|^2 + \sum_{\lambda \in \Lambda} \int |x \ast \psi_\lambda(\omega)|^2$ is the squared norm of wavelet representation and $\|x\|^2 = \int |x(t)|^2 dt$ is the signal norm. In Eq. 5, the lower bound guarantees a stable inverse while the upper bound ensures that the wavelet is a contractive operator [21]. If $\alpha = 0$, then $W$ becomes a tight frame and $x(t)$ can be reconstructed as $x(t) = (x \ast \tilde{\phi}(\omega)) \ast \phi(-t) + \sum_{\lambda \in \Lambda} \text{Real} \{x \ast \psi_\lambda(\omega)\}$ [24].

In the scattering transform, the wavelet modulus is employed for feature extraction. In contrast to the Fourier transform, where it is not possible to reconstruct the signal using its Fourier modulus, it is possible to reconstruct the signal using just modulus of complex wavelet [26]. This is due to the redundant representation in Eq. 2. Moreover, since the complex modulus is contractive, $||a| - |b||$ for any $(a, b) \in \mathbb{C}$, the wavelet modulus operator $|W|$ is also contractive. Therefore,

$$\|Wx - |W|x'\| \leq \|Wx - Wx'\| \leq \|x - x'\|$$

(6)

**WAVELET SCATTERING TRANSFORM**

The main idea behind the scattering transform is to analyze the signal using analytical wavelet and then average the wavelet coefficients over time to extract the signal features. The intuition behind the averaging coefficient is to reduce the variability in features and is similar to averaging the Fourier coefficient over Mel frequency intervals to extract Mel-frequency cepstral coefficients (MFCC) in speech processing [27]. However, MFCC loses information by averaging, while scattering transform preserves the information that is required for signal reconstruction [28].

In the scattering transform, a locally translation invariant descriptor is obtained by a time average $\delta_t x = x \ast \phi(t)$ that removes the signal high frequency contents. However, the high frequency content is recovered by a wavelet modulus transform as $|W|_1 x = (x \ast \tilde{\phi}(\omega), x \ast \psi_\lambda(\omega))$. The first order of the scattering coefficients can be obtained by $S_1 x(t, 1) = x \ast \psi_\lambda(\omega) \ast \phi$. These coefficients measure the average signal amplitude in the frequency interval covered by $\psi_\lambda$ with bandwidth corresponding to $Q_1$. In essence, they are calculated by a second wavelet modulus operator as $|W|_2 x \ast \psi_\lambda = (x \ast \psi_\lambda \ast \phi, x \ast \psi_\lambda \ast \psi_\lambda)$. So, the second order scattering coefficients are $S_2 x = (x \ast \psi_\lambda \ast \psi_\lambda)$ which are computed by a $\psi_\lambda$ with a bandwidth corresponding to $Q_2$. Performing this operation iteratively, the scattering coefficients can be defined via any arbitrary order.

For any $m \geq 1$, iterated wavelet modulus convolutions are written as $U_m(x, \lambda_1, \lambda_2, \cdots, \lambda_m) = (x \ast \psi_\lambda \ast \cdots \ast \psi_\lambda) \ast \psi_\lambda$ where the $m$th order wavelet has an octave resolution of $Q_m$ while satisfying Eq. 4. Next, the $m$th order scattering coefficients
are obtained by averaging \( U_{m,x} \) with \( \phi \) via 
\[
S_m x(t, \lambda_1, \cdots, \lambda_m) = U_{m,x}(t, \lambda_1, \cdots, \lambda_m) \ast \phi.
\]
Hence, the scattering decomposition of a signal with the maximum order of \( l \) is an iterative operation obtained from applying \( |W|_{m+1} \) to \( U_{m,x} \) in order to obtain 
\( S_m x \) and \( U_{m+1,x} \) for \( 0 \leq m \leq l \) where \( U_0 x = x \). The scattering transform is the collection of all coefficients given by the set 
\( S x = \{ S_m x | 0 \leq m \leq l \} \). Fig. 1 shows the decomposition graph.

![Figure 1. Scattering transform of signal x by iterating the \(|W|_m\) operator. The black dots show the output nodes.](image)

It can be shown that the scattering transform enjoys the following properties [25]:

**Time-warp deformation stability:** The scattering transform satisfies the Lipschitz condition, (i.e., there exist a constant \( C \) for any \( x \) such that 
\[
\| S x - S x' \| \leq C sup \| \tau (t) \| \| x - x' \|.
\]

**Contraction:** The scattering transform is contractive, (i.e., 
\[
\| S x - S x' \| \leq \| x - x' \|.
\]
As a result of this property, the scattering transform is robust against the additive noise.

**Energy conservation:** The scattering transform preserves the energy of the signals. In other words, if the chosen wavelet is a tight frame, then the scattering transform preserves the norm, (i.e., 
\[
\| x \|^2 = \| S x \|^2 + \| U_{l+1,x} \|^2.
\]
As a result, 
\[
\| U_{l+1,x} \|^2 \text{ vanishes as } l \to \infty.
\]
In practice, the coefficients become very small after a few iterations.

The aforementioned properties make the scattering transform a suitable candidate for signal classification tasks. In the next section, we propose a robust fault diagnosis algorithm based on this transform.

**FAULT DIAGNOSIS METHODOLOGY**

**SIGNAL PROCESSING PIPELINE**

In this section we present the main result of the paper, i.e., a vibration signal monitoring algorithm for diagnosing bearing faults from the measurements provided by accelerometers. Fig. 2 depicts the signal processing pipeline of our proposed fault diagnosis algorithm.

![Figure 2. The proposed four-stage fault diagnosis algorithm using vibration signal.](image)

As it is depicted in Fig. 2, the vibration signal is pass through the scattering transform, in the first step. Here, we use a two layer (\( l = 2 \)) scattering network with the normalized Morlet wavelet defined by
\[
\psi(t) = e^{-t^2/\sigma^2} \cos(\omega t - \theta(t)) = e^{-t^2/\sigma^2} \cos(\hat{\omega} t - \hat{\theta}(t)),
\]
which is simply a modulated Gaussian function. In the frequency domain, 
\[
\hat{\psi}(\omega) = \hat{\theta}(\omega - 1) - \hat{\theta}(\omega)\hat{\theta}(-1) / \hat{\theta}(0)
\]
which guarantees \( \hat{\psi}(0) = 0 \). The parameter \( \sigma^2 \) determines the bandwidth of the wavelet which is assigned based on the choice of \( Q \) in the scattering network. Fig. 3 demonstrates Morlet wavelets for \( Q_1 = 4 \) and \( Q_2 = 2 \), used in our scattering network.

In the second stage of the pipeline, the scattering coefficients are post-processed in order to obtain a better classification rate. The post-processing stage involves two steps: first, the scattering coefficients are normalized to be invariant with respect to changes in the amplitude of input signal by dividing the coefficients in each layer by the corresponding coefficients in the predecessor layer where the first layer coefficients are normalized by \( S_0 x \). After normalization, the log function is applied to the normalized coefficients as a range compressor. This leads to an improved classification accuracy as well as an improved visualization contrast.

In the third stage, the dimensions of the feature space is
reduced using the principle component analysis (PCA). Indeed, PCA projects high dimensional data onto a space of lower dimension by a linear transformation [29]. The linear transformation is chosen such that the error of data reconstruction is minimized.

In the final stage, a random forest classifier diagnoses the health state of the machine using the extracted features. The random forest classifier is an ensemble of decision trees with fast realization. The classifier is extended to multi-class classifier based on the error correcting output code scheme proposed in [30].

BENCHMARK DATASET

The proposed method is applied to the vibration data provided by the Case Western Reserve University (CWRU) bearing data center [31] as a benchmark. The experimental setup, depicted in Figure 4(a), consists of a 2 hp motor, a torque transducer, in the center, and a dynamo-meter as a load. Bearings of motor shaft are studied in this benchmark. Figure 4(b) shows the setup block diagram. Single point faults were introduced to test bearings using an electro-discharge machine with diameters of 7, 14, and 21 millimeters on different parts of the bearing including the inner race, the outer race and the ball. Vibration data was collected using two single axial accelerometers with magnetic bases. One accelerometer was mounted on the drive end and the other on the fan placed at the end of motor housing. The accelerometer readings were sampled at the rate of 12,000 samples per second in the constant shaft speed.

The experiment was repeated for different fault intensities (7, 14 and 21 mils) and locations (ball, races and drive/fan ends). Moreover, data was collected for four load conditions (0, 1, 2 and 3 hp) for each class of faults. Figure 5 shows examples of vibration data for fault-free ($f_0$) and ball fault ($f_1$), inner and outer race faults ($f_2$ and $f_3$). This Figure compares the effect of load condition on vibration data for two different loads 1 hp and 3 hp. As seen spectrum representation of vibration data from the same class is different for all four classes. This difference is more noticeable for faulty scenarios.

EXPERIMENTAL RESULTS

The benchmark dataset is processed by the aforementioned scattering network. The dimensionality of the feature space is reduced using PCA. The dimension of the new feature space is chosen based on the proportional cumulative variance of the PCA eigenvalues [32]. Fig. 6 depicts the projected features onto the subspace spanned by the first two principal components.

Accuracy of the proposed algorithm is measured twice using 4-fold cross validation. In the first test (Test I in Table 1), the dataset is shuffled and three folds of data are used for tuning the classifier while the system accuracy is tested with the forth fold. In this experiment, the system was exposed to data from all loads in the training phase. The average accuracy of 99.40% is obtained. For investigating the performance of the system to a load which is not exposed to the system in training phase, we
Table 2. Comparing various methods Using CWRU benchmark

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Type of Features</th>
<th>Classifier</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[17]</td>
<td>Global spectrum</td>
<td>Linear</td>
<td>95</td>
</tr>
<tr>
<td>[33]</td>
<td>statistical &amp; frequency domain</td>
<td>ANFIS</td>
<td>91.33</td>
</tr>
<tr>
<td>[18]</td>
<td>vibration Spectrum Imaging of FFT</td>
<td>ANN</td>
<td>96.90</td>
</tr>
<tr>
<td>[20]</td>
<td>statistical &amp; Wavelet packet features</td>
<td>SVM</td>
<td>90</td>
</tr>
<tr>
<td>[19]</td>
<td>pool of time and frequency features</td>
<td>Fuzzy</td>
<td>89.63</td>
</tr>
<tr>
<td>Ours</td>
<td>wavelet scattering (Test I)</td>
<td>SVM</td>
<td>99.08</td>
</tr>
</tbody>
</table>

arranged the second test (Test II in Table 1). Here, we used data from three randomly chosen loads for training and we measured accuracy using data from the forth load. The average accuracy of 98.62% is obtained. The confusion matrix of both tests are tabulated in Table 1. The average accuracy of the system in both tests are close.

We implemented the entire signal processing pipeline depicted in Fig. 2 in MATLAB on a PC running Ubuntu Linux 14.04 at a clock rate of 3.4 GHz. The average runtime (wall-clock) of processing a window of vibration signal with a duration of 30 seconds is almost equal to 5.8 seconds which shows the possibility of real-time implementation of the proposed method on this machine.

Table 2 compares the accuracy obtained in prior works. As seen, the proposed method has a higher accuracy. To the best of our knowledge, there is no report in the literature doing a test like test II on CWRU benchmark dataset. Even in test II, the proposed algorithm outperforms the prior works. However, the training algorithm was not exposed to data from all loads.

CONCLUSION

In this study, a novel method for fault diagnosis of bearings using the vibration signal was proposed. This method was based on a wavelet scattering transform which provides a robust feature extraction against deformations caused by load changes in this signal. The proposed method has an affordable computational cost for real-time implementation and is tested on the CWRU vibration benchmark dataset.

REFERENCES


Figure 7. Scattergram of the vibration signal for different classes. The load effect is captured by the scattering transform.