Mechanical Systems and Signal Processing 80 (2016) 349–376

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Kurtosis based weighted sparse model with convex optimization technique for bearing fault diagnosis

Han Zhang[a], Xuefeng Chen[a,b,*], Zhaohui Du[a,c], Ruqiang Yan[d]

[a] State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[b] Collaborative Innovation Center of High-End Manufacturing Equipment, Xi'an Jiaotong University, Xi'an 710054, China
[c] Department of Mathematics, University of California, Los Angeles, CA 90095, USA
[d] School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 16 November 2014
Received in revised form
24 February 2016
Accepted 30 April 2016
Available online 12 May 2016

Keywords:
Impulsive feature frequency band splitting phenomenon
Envelope spectrum kurtosis
Weighted sparse regularization
Sparse decomposition theory
Alternating direction method of multipliers
Bearing fault diagnosis

ABSTRACT

The bearing failure, generating harmful vibrations, is one of the most frequent reasons for machine breakdowns. Thus, performing bearing fault diagnosis is an essential procedure to improve the reliability of the mechanical system and reduce its operating expenses. Most of the previous studies focused on rolling bearing fault diagnosis could be categorized into two main families, kurtosis-based filter method and wavelet-based shrinkage method. Although tremendous progresses have been made, their effectiveness suffers from three potential drawbacks: firstly, fault information is often decomposed into proximal frequency bands and results in impulsive feature frequency band splitting (IFFBS) phenomenon, which significantly degrades the performance of capturing the optimal information band; secondly, noise energy spreads throughout all frequency bins and contaminates fault information in the information band, especially under the heavy noisy circumstance; thirdly, wavelet coefficients are shrunk equally to satisfy the sparsity constraints and most of the feature information energy are thus eliminated unreasonably. Therefore, exploiting two pieces of prior information (i.e., one is that the coefficient sequences of fault information in the wavelet basis is sparse, and the other is that the kurtosis of the envelope spectrum could evaluate accurately the information capacity of rolling bearing faults), a novel weighted sparse model and its corresponding framework for bearing fault diagnosis is proposed in this paper, coined KurWSD. KurWSD formulates the prior information into weighted sparse regularization terms and then obtains a nonsmooth convex optimization problem. The alternating direction method of multipliers (ADMM) is sequentially employed to solve this problem and the fault information is extracted through the estimated wavelet coefficients. Compared with state-of-the-art methods, KurWSD overcomes the three drawbacks and utilizes the advantages of both family tools. KurWSD has three main advantages: firstly, all the characteristic information scattered in proximal sub-bands is gathered through synthesizing those impulsive dominant sub-band signals and thus eliminates the dilemma of the IFFBS phenomenon. Secondly, the noises in the focused sub-bands could be alleviated efficiently through shrinking or removing the dense wavelet coefficients of Gaussian noise. Lastly, wavelet

---

[*] Corresponding author at: State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China.
E-mail addresses: zhanghan12@stu.xjtu.edu.cn (H. Zhang), chenxf@mail.xjtu.edu.cn (X. Chen), duzhaohui3@stu.xjtu.edu.cn (Z. Du), ruqiang@seu.edu.cn (R. Yan).

http://dx.doi.org/10.1016/j.ymssp.2016.04.033
0888-3270/

coefficients with faulty information are reliably detected and preserved through manipulating wavelet coefficients discriminatively based on the contribution to the impulsive components. Moreover, the reliability and effectiveness of the KurWSD are demonstrated through simulated and experimental signals.

1. Introduction

Mechanical systems are vital part of our daily lives as we rely on their flawless performance. Thus, the timely and precise fault diagnosis of the machines is vital [1–3]. Among the strategies for maintenance, predictive maintenance has been increasingly applied over the years [4]. This technique aims to detect faults at an early stage, thus greatly reducing unnecessary repair costs. It also improves the reliability of rotational machine systems and reduces its operating expenses. Furthermore, this technique is closely related to condition-based maintenance (CBM), performed online or offline, in which maintenance is scheduled according to the wear of each component, reducing the possibility of an unexpected failure [5].

One of the most important elements of the rotational machine systems is the rolling bearings and their failure is one of the most frequent reasons for machine breakdowns [6]. Meanwhile, early detection and precise isolation of bearing faults can decrease maintenance costs and increase operating safety. Therefore, the detection and monitoring of bearing faults in incipient stages (low severity) are essential. Once localized faults occur, such as pits, spalls, cracks, the strikes of rollers on the fault surfaces excite the resonances of the structures between the bearing and the transducers, triggering the modulation phenomenon [7]. Moreover, impulsive signals of local faults are generated almost periodically and their characteristics depend on the location of the defect; that is, whether it is on the inner race, outer race or rolling elements. Therefore, a diagnostic scheme might be introduced: (i) detect impulsive nature of the signal, (ii) check if there is periodic nature of impulsive components and try to match it to expected fault frequencies. Unfortunately, in practice, such a cyclic impulse train can be hardly seen. There are two main possible reasons. First, weak impulsive features of incipient faults are always submersed due to the high level noise present during the experiment. Second, periodic information of impulses is corrupted by discrete harmonic interferences coming from blade, shaft, mesh, resonance, ghost, etc. components.

In the last few years, two main families of signal processing tools have gained a leading role in extracting weak periodic feature information for fault diagnosis of rolling bearings.

The first family consists of time and frequency domain indexes derived from the kurtosis index, which is chosen as a direct measure of the impulsiveness of the signal [8–10]. Kurtosis, band kurtosis, spectral kurtosis and kurtogram are the most used techniques belonging to this category. Kurtosis has long been used as a measure of the severity of machine faults, since its proposal by Stewart et al. in the 1970s [11], but there was only a vague suggestion that clearer results might be achieved by using filtering in the frequency bands. Next, Antoni and Randall paid mathematical insights to SK and proposed Fast Kurtogram [12], which consequently represents an efficient tool to select the best band for the filtering step in the process of envelope analysis. Since that time, improvements of both spectral kurtosis and kurtogram have attracted a great deal of attention. Barszcz and JabLonski [13] developed a novel method called the Protrugram based on a higher-resolution kurtosis-based index. This index is obtained by calculating the kurtosis of the analytic signal on a narrow band, which is shifted almost continuously along the frequency axis. Wang and Liang [14] proposed an adaptive spectral kurtosis method that adaptively determines the bandwidth and center frequency of the optimal filter through merging right-expanded windows with its subsequent neighboring windows along the frequency axis. Chen [15] presented a spatial-spectral ensemble kurtosis index to evaluate signal impulsiveness and introduced an improved kurtogram. Tse and Wang [16] substituted the sparsity value for the SK in the enhanced kurtogram to form a sparsogram for bearing fault diagnosis.

The second family of diagnostic tools is represented by wavelet theory. The multi-resolution analysis ability of wavelet analysis makes it suitable for revealing fault-related information from non-stationary signals sampled on rotary machines. The basic idea behind wavelet analysis is that the energy of the signal to be identified will concentrate on a few wavelet coefficients while the energy of noise will spread throughout all wavelet coefficients [17]. Similarity between the basic wavelet and the signal to be identified is very important in order to make the signal concentrate on fewer coefficients. The more similar the signal to the wavelet function, the better the defect-related features will be extracted [18]. Refs. [19,15,1,3] constructed many types of wavelets to matching transient features and Refs. [20–22] quantitatively evaluated the performance of wavelet basis. Based on an optimum wavelet basis, wavelet coefficients could be obtained through fast implementation and two strategies of selecting fault-related coefficients have been studied extensively. One is matched filtering which keeps only those coefficients at a fixed scale corresponding to the optimum impulsive feature frequency band [23–25]. The other is wavelet threshold which manipulates coefficients in detail based on some threshold rules to detect crucial characteristic coefficients and suppress noisy coefficients [26,27]. Moreover, various parameters extracted from wavelet coefficients are often used as indictors to evaluate machine health status. Several studies have attempted to address related issues and there have been tremendous progresses, so that it is a huge work for anyone who attempts to review all the vast achievements. Excellent review articles are recommended as [3,28,1].