ARTICLE IN PRESS

Mechanical Systems and Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

# A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis

M. Grasso *, S. Chatterton, P. Pennacchi, B.M. Colosimo

*Dipartimento di Meccanica, Politecnico di Milano, Via La Masa 1, 20156 Milan, Italy*

ARTICLE INFO

*Article history:*
Received 6 May 2015
Received in revised form
30 January 2016
Accepted 27 February 2016

*Keywords:*
Empirical Mode Decomposition
Combined Mode Functions
Vibration
Bearing
Fault detection

ABSTRACT

Health condition analysis and diagnostics of rotating machinery requires the capability of properly characterizing the information content of sensor signals in order to detect and identify possible fault features. Time–frequency analysis plays a fundamental role, as it allows determining both the existence and the causes of a fault. The separation of components belonging to different time–frequency scales, either associated to healthy or faulty conditions, represents a challenge that motivates the development of effective methodologies for multi-scale signal decomposition. In this framework, the Empirical Mode Decomposition (EMD) is a flexible tool, thanks to its data-driven and adaptive nature. However, the EMD usually yields an over-decomposition of the original signals into a large number of intrinsic mode functions (IMFs). The selection of most relevant IMFs is a challenging task, and the reference literature lacks automated methods to achieve a *synthetic* decomposition into few physically meaningful modes by avoiding the generation of spurious or meaningless modes. The paper proposes a novel automated approach aimed at generating a decomposition into a minimal number of relevant modes, called Combined Mode Functions (CMFs), each consisting in a sum of adjacent IMFs that share similar properties. The final number of CMFs is selected in a fully data driven way, leading to an enhanced characterization of the signal content without any information loss. A novel criterion to assess the dissimilarity between adjacent CMFs is proposed, based on probability density functions of frequency spectra. The method is suitable to analyze vibration signals that may be periodically acquired within the operating life of rotating machineries. A rolling element bearing fault analysis based on experimental data is presented to demonstrate the performances of the method and the provided benefits.


## 1. Introduction

Sensor signals involved in health condition analysis of rotating machinery usually exhibit a multi-scale information content, due to the superimposition of features on different time–frequency scales, either stationary or non-stationary. Typical rolling bearing faults are caused by localized defects that generate impact vibrations. Thus, time–frequency analysis is a powerful approach to characterize both the time of impacts and the corresponding frequency ranges. Empirical Mode Decomposition (EMD) gained increasing influence in the technical literature to this aim. This kind of analysis relies on a

* Corresponding author.
*E-mail addresses:* marcoluigi.grasso@polimi.it (M. Grasso), steven.chatterton@polimi.it (S. Chatterton), paolo.pennacchi@polimi.it (P. Pennacchi), biancamaria.colosimo@polimi.it (B.M. Colosimo).

http://dx.doi.org/10.1016/j.ymssp.2016.02.067


Please cite this article as: M. Grasso, et al., A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis, Mech. Syst. Signal Process. (2016), http://dx.doi.org/10.1016/j.ymssp.2016.02.067

**Nomenclature**

| | |
|---|---|
| BSF | Ball Spin Frequency |
| $c_i(t)$ | $i$th IMF extracted from the signal $Y(t)$, $i=1,\ldots,n$, |
| $c_{s_k}(t)$ | $k$th sequential CMF extracted from the signal $Y(t)$, $k=1,\ldots,n$ |
| $c^*_{s_k}(t)$ | $k$th final CMF extracted from the signal $Y(t)$, $k=1,\ldots,K$ |
| CMF | Distance (dissimilarity) between the $k$th and the $(k+1)$th IMFs from $Y(t)$ |
| $D_{k,k+1}$ | Distance (dissimilarity) between the $k$th and the $(k+1)$th IMFs from $Y(t)$ |
| EMD | Empirical Mode Decomposition |
| EEMD | Ensemble Empirical Mode Decomposition |
| $n$ | Number of IMFs extracted from the signal $Y(t)$ |
| $f(x)$ | Probability density function of random process $x(t)$ |
| $\hat{f}(x)$ | Kernel estimator of the probability density function of random process $x(t)$ |
| $F_s$ | Sampling frequency |
| FFT | Fast Fourier Transform |
| FTF | Fundamental train frequency |
| $h$ | Bandwidth of the kernel function |
| $\hat{h}$ | Optimal bandwidth of the kernel function |
| $h_u(t)$ | Difference between the signal $Y(t)$ and $m_u(t)$, at uth step of the sifting algorithm |
| $K^*$ | Number of iteratively generated CMFs |
| $K$ | Final number of CMFs extracted from the signal $Y(t)$ |
| Ker($x$) | Kernel function |
| IMF | Intrinsic Mode Function |
| $M$ | Number of peaks in the $D_{k,k+1}$ function |
| $m_u(t)$ | Mean of envelopes at the uth step of the sifting algorithm |
| $\boldsymbol{p}$ | Vector of “locations” $k$ corresponding to peaks in the $D_{k,k+1}$ function |
| $\tilde{\boldsymbol{p}}$ | Vector $\boldsymbol{p}$ with elements sorted in descending peak amplitude order |
| PDF | Probability density function |
| $q_k$ | Number of IMFs included into the kth CMF, $k=1,\ldots,K$ |
| $r_i$ | Normalized sample correlation coefficient between $Y(t)$ and the ith IMF |
| $r_n(t)$ | Sum-of-squares between the $K^*$ CMFs from $Y(t)$ |
| $rms$ | Sum-of-squares between the $K^*$ CMFs from $Y(t)$ |
| SSB($K^*$) | Sum-of-squares between the $K^*$ CMFs from $Y(t)$ |
| SSW($K^*$) | Sum-of-squares within the $K^*$ CMFs from $Y(t)$ |
| $T$ | Time window length |
| $tol$ | Tolerance threshold |
| UCV | Unbiased Cross Validation |
| $\text{UCV}_k(\hat{h})$ | Unbiased cross-validation statistic for kth CMF from $Y(t)$, with bandwidth $\hat{h}$ |
| $w_k(\omega)$ | Weight function in the PDF for the kth CMF from $Y(t)$ |
| $x_k(\omega)$ | Amplitude of frequency spectrum of the kth CMF from $Y(t)$ |
| $Y(t)$ | Vibration signal |
| $\lambda$ | Threshold used in the index-based approach for IMF selection |
| $\rho(\bullet,\bullet)$ | Sample cross-correlation coefficient |
| ω | Frequency location |

decomposition of vibration signals into their embedded modes. Signal decomposition is a critical step that strongly influences the capability of isolating fault features and determining the health condition of the system. The achievement of a good characterization of the multi-scale content of the signal is of great importance to detect and diagnose faulty conditions in order to reduce plant downtime and to rapidly react to performance worsening caused by degraded states of the machine components. Although several multi-resolution techniques may be applied to this aim [1,2], the achievement of a synthetic decomposition into a minimal number of physically meaningful and interpretable oscillation modes still represents an open issue. In some cases, a number of decomposition levels is imposed *a-priori* and the signal is reconstructed by applying level-dependent thresholding techniques [3]. In many practical applications, the signal is first decomposed into a number of scales (usually larger than the one required to describe the relevant content), and a subset of modes of interest is then selected [4–6]. However, this latter approach yields a potential information loss. Furthermore, mode selection may be a troublesome task in practice, being usually based on the human expert's knowledge and difficult to apply in an automatic way. Among time–frequency analysis techniques, the EMD proposed by Huang et al. [7] has several attractive properties that make it suitable to fault detection and diagnosis problems. The EMD is a nonparametric, data-driven and adaptive method that allows decomposing any signal into a number of Intrinsic Mode Functions (IMFs), without any prior basis selection. Due to its data-driven nature, the number of IMFs may vary over time, when the EMD is applied to periodically acquired signals. As an example, a higher frequency content in vibration signals caused by a defective bearing may lead to a larger number of IMFs than the ones extracted from the signal under healthy conditions, due to increased frequency ranges and energy levels [8]. In addition, the sifting algorithm is known to be affected by the so-called “*mode mixing*” problem [9], which may cause either a splitting of one intrinsic mode into two (or more) adjacent IMFs, or a merging of different scales into a single IMF. The EMD usually yields an over-decomposition of the signal, which can be inflated by the *mode mixing* effect and/or by the specific choice of the stopping criterion, leading to the presence of IMFs with no physical meaning. The introduction of the Ensemble Empirical Mode Decomposition (EEMD) [9] helped to mitigate mode mixing effects, but the EEMD is not able to avoid the over-decomposition imposed by the sifting algorithm. As a matter of fact, the literature devoted to the EMD and other multi-scale analysis methods lacks automated approaches for the achievement of a synthetic decomposition into a

Please cite this article as: M. Grasso, et al., A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis, Mech. Syst. Signal Process. (2016), http://dx.doi.org/10.1016/j.ymssp.2016.02.067